Methods for the Estimation of the NICE Cost Effectiveness Threshold

Revised Report Following Referees Comments

Karl Claxton,[1,2] Steve Martin,[2] Marta Soares,[1] Nigel Rice,[1] Eldon Spackman,[1] Sebastian Hinde,[1] Nancy Devlin,[3] Peter C Smith,[4] Mark Sculpher[1]

1. Centre for Health Economics, University of York, UK
2. Department of Economics and Related Studies, University of York, UK
3. Office of Health Economics, London, UK
4. Imperial College, London, UK

10th June 2013

Acknowledgments

The authors would like to acknowledge funding received from the Medical Research Council/National Institute for Health Research through its Methodology Research Programme (Award G0901498/1).

We acknowledge the contribution of the following individuals who participated in various aspects of this research: Ling-Hsiang Chuang, Craig Currie, Jamie Garside, Simon Gilbody, Charlotte Haylock, Sarah Jenkins-Jones, Tim Kendall David Parkin and John Parkinson.

We are also grateful to those colleagues who attended our research workshop in May 2010 and offered comments on our preliminary methods. We would especially like to thank Peter Littlejohns for his support in advocating this topic to the Methodology Research Programme.

This is an internal report which will be formally reported in *Health Technology Assessment* as a peer-reviewed journal.

All views expressed here, and any errors, are entirely the responsibility of the authors.

Contributions of authors

All of the named authors below contributed to the development of the research questions, study design and implementation (including membership of study management group), analysis and/ or interpretation of data and submission of the final report. Contributions to particular elements of the study are described below.

Karl Claxton (Professor of Economics, University of York) was a co-applicant, and led the overall design and of the study and interpretation of results, and the writing of the report.

Steve Martin (Researcher, University of York) was co-applicant and undertook and wrote-up the econometric analyses detailed in Chapter 3 and Appendix B.

Marta Soares (Research Fellow, University of York) undertook the analyses extending the econometric analysis on mortality changes to broader health effects on QALYs (Chapter 4 and Appendix C.

Nigel Rice (Professor of Health Economics, University of York) was co-applicant and contributed to the design, interpretation and write-up of the econometric analysis.

Eldon Spackman (Research Fellow, University of York) designed and undertook the systematic review (Appendix A), and contributed to analyses linking mortality changes to broader health effects on QALYs.

Sebastian Hinde (Research Fellow, University of York) designed, undertook and wrote-up the systematic review (Appendix A), and contributed to analyses linking mortality changes to broader health effects on QALYs.

Nancy Devlin (Director of Research, Office of Health Economics) was a co-applicant and led the design and write up of the review of local data availability.

Peter C Smith (Professor of Health Policy) was a co-applicant, and contributed to the design, interpretation of all aspects of the analysis.

Mark Sculpher (Professor of Health Economics) was principal applicant, chaired the study management group, and contributed to the design, interpretation and write-up of all aspects of the analysis.

Contents ***Page***

Scientific Summary

1 Introduction

1.1 NICE's comparison of the incremental cost effectiveness ratio (ICER) of a new technology, which is more costly than existing alternatives, with the cost-effectiveness threshold is important in assessing whether the health expected to be gained from its use exceeds the health expected to be forgone elsewhere as other NHS activities are displaced (i.e. whether the new technology is cost effective).

1.2 When NICE issues positive guidance for a new intervention which imposes additional costs on the NHS, the resources required to deliver it must be found by disinvesting from other interventions and services elsewhere. This displacement will inevitably result in health decrements for other types of individual. Thus the threshold represents the additional cost that has to be imposed on the system to forgo 1 QALY of health through displacement.

1.3 Currently NICE uses a threshold range of £20,000 to £30,000 QALY gained, and this has remained the case in NICE's methods guidance since 2004. There has been a number of calls for further research on the value of the threshold.

1.4 This report details a 2-year project, funded by the NIHR and MRC Methodology Research Programme, to develop methods to estimate the NICE cost effectiveness threshold.

1.5 NICE's remit implies a series of characteristics for any empirical research on the threshold:

- Reflect the expected health effects (in terms of length and quality of life) of NICE guidance through the displacement decisions taken across the NHS rather than what specific services are (or could have been) displaced.
- Facilitate regular updates, based on routinely available data, to reflect NHS changes such as real overall expenditure and productivity. This would encourage accountability through scrutiny by stakeholders and provide predictability for technology manufacturers' investment decisions.
- The nature of service displacement and the magnitude of the health forgone will depend on the scale of the budget impact which should, ideally, be reflected in the value of the threshold.
- Methods should recognise the inevitable uncertainty relating to the evidence currently available for the threshold and reflect its implications for policy.

2 Study methods

2.1 The aim was to develop methods to estimate the NICE cost-effectiveness threshold making use of routinely available data. Objectives were:

i. Informed by relevant literature, to provide a conceptual framework to define the threshold and the basis of its estimation.
ii. Using programme budgeting data for the English NHS, to estimate the relationship between changes in overall NHS expenditure and changes in mortality.
iii. Extend the measure of benefit in the threshold to QALYs by estimating the quality of life (QoL) associated with additional years of life and the direct impact of health services on QoL.
iv. Present the best estimate of the cost effectiveness threshold for policy purposes.

2.2 Earlier econometric analysis estimated the relationship between differences in primary care trust (PCT) spending and associated disease-specific mortality. Expenditure came from programme budgeting data which allocates the entire volume of health care expenditure to broad programme budget categories (PBCs) according to primary diagnosis.

2.3 This research extended this in several ways including estimating the impact of marginal increases or decreases in overall NHS expenditure on spending in each of the 23 PBCs. These were linked to changes in mortality outcomes by PBC across 11 PBCs.

2.4 The results of the econometric analysis were translated into broader effects in terms of QALYs. The first stage linked estimated effects on mortality to life years taking into account the 'counterfactual' deaths that would have occurred if the population in a given PBC faced the same mortality risks as the general population. The second stage accounted for the health (QALY) effects of changes in mortality due to changes in expenditure reflecting how QoL differs by age and gender. The third stage incorporated those effects on health not directly associated with mortality and life year effects (i.e., the 'pure' QoL effects) to estimate an overall cost per QALY threshold. The approach uses the estimates of mortality and life year effects as 'surrogate outcomes' for a more complete measure of the health effects of a change in expenditure. This appears more plausible than assuming no effects of NHS expenditure on quality of life outcomes.

2.5 The estimated proportional effect on the mortality and life year burden of disease is applied to measures of QALY burden. Applying a proportionate effect to measures of QALY burden of disease is equivalent to assuming that any estimated effects on life years are lived at quality of life that reflects a proportionate improvement to the quality of life with disease. It also allows quality of life effects of changes in expenditure to be included; also based on proportionate improvement in the quality of life with disease. In those PBCs where mortality effects could not be estimated the proportional effect of changes in expenditure on QALY burden of disease is assumed to be the same as the overall proportional effect on the life year burden of disease across those PBCs where mortality effects could be estimated.

2.6 The methods planned for the study included a consideration of local data, collected routinely by PCTs, on the types of intervention in which local decision-makers were investing and disinvesting. The aim was to inform the link between the effects of expenditure changes on mortality and impacts on broader health in terms of QALYs. These data may have indicated the types of interventions and services, within a given PBC, on which investment and disinvestment were taking place. Using targeted literature reviews, estimates of QoL for those activities may have been identified. However, it was established that there were limited data available at a local level to facilitate this type of analysis, so other data sources were used for this purpose.

3 Central or 'best' estimate of the threshold

3.1 The most relevant threshold is estimated using the latest available data (2008 expenditure, 2008-10 mortality). The central or 'best' threshold is estimated to be £18,317 per QALY.

4 Which PBCs have the greatest influence on the overall threshold?

4.1 Although the 11 PBCs where a mortality effect of changes in expenditure could be estimated only account for 36% of the change in overall expenditure, they account for 80% of the overall health effects. The other 12 PBCs, where mortality effects could not be estimated, account for the greater

part of a change in overall expenditure (64%) but only 20% of the overall health effects, i.e., the cost per QALY estimates associated with a change in expenditure in these PBCs are, in general, much higher.

4.2 Insofar as investment and disinvestment opportunities in these PBCs might have been more valuable (offered greater improvement in QoL) than suggested by the implied PBC thresholds, the overall QALY effects will tend to be underestimated and the overall cost per QALY threshold will be overestimated.

4.3 The overall threshold of £18,317 may be especially conservative (i.e., likely to be overestimated) with respect to health effects in PBC5 (Mental Health Disorders), which accounts for a large proportion of the change in overall expenditure (25%) and contributes most to the overall health effects (9%) compared to these other PBCs. The cost per QALY associated with this PBC is based on an extrapolation rather than observations of the direct impact of changes in expenditure on QoL. Available evidence suggests that the investment and disinvestment opportunities in mental health are likely to have been much more valuable than its implied cost per QALY.

5 How uncertain are the estimates and what are the implications?

5.1 Simulation methods were used to reflect the combined uncertainty in the various estimates from the econometric analysis. This indicated that the probability that the overall threshold is less than £20,000 per QALY is 0.64 and the probability that it is less than £30,000 is 0.92.

5.2 As the consequences of overestimating the threshold are more serious than underestimating it in terms of population health, a *policy threshold* will be lower than the mean of the cost per QALY threshold (i.e., lower than £18,317) to compensate for the more serious consequences of overestimating the 'true' value.

5.3 There were other ('structural') sources of uncertainty associated with the estimated threshold, specifically relating to the choice of econometric models and identification of causal effects. Although all the models passed the relevant tests of validity, there remained some uncertainty about the validity of the instruments. This structural uncertainty constituted a greater part of the overall uncertainty associated with the mortality effects of changes in expenditure, but the central estimate of the cost per QALY threshold was robust to this uncertainty.

5.4 The method of analysis used to link the effects of changes in expenditure on mortality to a fuller measure of health expressed in QALYs was also subject to uncertainty. A preferred analysis (or scenario) was identified as making the best use of available information, with assumptions appearing more reasonable than the available alternatives and providing a more complete picture of the likely health effects of a change in expenditure.

5.5 A critical issue is whether, on balance, the central or best estimate is likely to be an underestimate or overestimate of the cost per QALY threshold. Although other assumptions and judgments are possible that retain some level of plausibility, they do not necessarily favour a higher threshold. Indeed, when considered together, they suggest that, on balance, the central or best estimate of £18,317 is, if anything, likely to be an overestimate.

5.6 There are some reasons why the central estimate of the QALY threshold might be underestimated (e.g., see items 1 to 4 in Box 5.1 in Section 5.4). For example in calculating life year effects it is assumed that those deaths averted by a change in expenditure returns the individuals to the mortality risk of the general population (matched for age and gender). There are a number of other reasons why the central estimate might be overestimated (e.g., see assumptions 5 to 7 in Box 5.1). For example, the health effects of a change in expenditure are restricted to the population at risk during one year. This also means that the health effects of changes in expenditure which reduce incidence (prevention of disease) will not be captured either. A more formal and longer lag structure in the estimation of outcome elasticities would be likely to capture more health effects of a change in expenditure.

5.7 The effect of other assumptions that have been necessary are more ambiguous although some evidence suggests their net effect maybe conservative with respect to health effects of changes in expenditure (e.g., assumptions 8 to 10 in Box 5.1).

6 The impact of investment, disinvestment and non marginal effects

6.1 The central estimate of the cost per QALY threshold is based on estimates of the health effects of changes in expenditure across all 152 PCTs, some of which will be making investments (where expenditure is increasing) and others making disinvestments (where expenditure is reduced or growing more slowly).

6.2 The threshold is, however, likely to differ across these different types of PCT. It would be expected that, other things equal, more expenditure would increase health but at a diminishing rate. Therefore, the amount of health displaced by disinvestment would be expected to be greater, and the associated threshold lower than the central estimate. Conversely, the health gained from investment would be expected to be lower, and the associated threshold higher.

6.3 This was examined by re-estimating the outcome and expenditure effects separately for those PCTs where their actual budget is under the target allocation from the Department of Health resource allocation formula (i.e., those under greater financial pressure and more likely to be disinvesting than investing), and those that are over target (under less financial pressure and more likely to be investing than disinvesting).

6.4 The results confirm these expectations: the health effects of changes in expenditure are greater when PCTs are under more financial pressure and are more likely to be disinvesting then investing. The analysis suggests that budget impact not only displaces more valuable activities within each PBC but that overall expenditure tends to be reallocated to PBCs which can generate more health. Although further research might enable a quantitative assessment of how the relevant threshold should be adjusted for the scale of budget impacts, the qualitative assessment seems clear: the central estimate of the threshold is likely to be an overestimate for all technologies which impose net costs on the NHS (almost all technologies appraised by NICE); and the appropriate threshold to apply should be lower for technologies which have a greater impact on NHS costs.

7 How does the threshold change with overall expenditure?

7.1 The same methods were used to consider how the cost per QALY threshold is likely to have changed from 2007 to 2008 as overall expenditure has increased. This provides some insights into

how the threshold might be expected to change over time as, for example, overall expenditure and NHS productivity changes.

7.2 This has implications for a judgement about the appropriate frequency of periodic reassessment of the cost per QALY threshold. Other things equal, the threshold would be expected to increase following a rise in overall expenditure, although this will depend on whether there is discretion over how additional resources can be spent. However, insofar as the productivity of those activities that are valuable to the NHS also improves through innovation, the threshold will tend to fall. So the net impact of these two countervailing effects on the threshold cannot be determined *a priori*.

7.3 Differences in the estimated thresholds between 2007 and 2008 were assessed. Although overall expenditure increased by 6% between 2007 and 2008 which represented real growth of 2% in 2007 prices, the overall threshold for all 23 PBCs fell by 2% in nominal terms and by 5% in real terms.

7.4 The reasons are complex but reflect changes in productivity, which differ across PBCs, but also a general reallocation of a change in overall expenditure towards those PBCs that appear more valuable in 2008. Given the uncertainty in estimation, subtle differences between 2007 and 2008 should not be over-interpreted. This analysis does suggest, however, that the overall threshold will not necessary increase with growth in the real or even nominal NHS budget. This suggests that the threshold is more likely to fall at a time when real budget growth is flat or falling and PCTs find themselves under increasing financial pressure.

8 What type of health is forgone by approval of a new technology?

8.1 The methods of analysis can identify not only how many QALYs are likely to be forgone across the NHS as a consequence of approving a technology which imposes additional costs on the NHS, but also where those QALYs are likely to be forgone and how they are made up, i.e., the additional deaths, life years lost and the QoL impacts on those with disease.

8.2 As an example, based on the 2008 central estimate of the cost per QALY threshold (£18,317), the approval of Ranibizumab for the treatment of diabetic macular oedema (prior to the patient access scheme agreement) would have imposed additional annual costs of up to £80m on the NHS each year and been likely to displace 4,367 QALYs elsewhere in the NHS. This forgone health is likely to be made up of 295 additional deaths and 1,337 life years forgone, most of which are likely to occur in Circulatory, Respiratory, Gastro-intestinal and Cancer PBCs. However, much of the total health effect of these additional costs (3,509 QALYs) is associated with QoL forgone during disease which is most likely to occur in Respiratory, Neurological and Mental Health PBCs.

9 Conclusions and implications for practice

9.1 The research presented here goes some way to providing an empirically-based and explicit quantification of the scale of opportunity costs the NHS faces when considering whether the health benefits associated with new technologies are expected to offset the health that is likely to be forgone elsewhere in the NHS. As such it provides a basis for determining the appropriate threshold for NICE decisions as well as those made centrally by the NHS and Department of Health more generally.

9.2 The methods presented can be used as a framework for further empirical work as additional and more appropriate data emerge in the NHS. They also offer a basis for threshold estimation in other health care systems with budget constraints that use cost effectiveness analysis to inform resource allocation decisions.

9.3 The study also starts to make the other NHS patients, who ultimately bear the opportunity costs of such decisions, less abstract and more 'known' in social decisions. Since who happens to be known or unknown is only a matter of perspective, time and ignorance, ethical and coherent social decisions require that both should be treated in the same way. These methods contribute to removing some of the 'ignorance' and making the unknown more real.

9.4 This work has implications for the Government's proposals to move to a system of value-based pricing for new prescription pharmaceuticals, which may include some additional weight for health benefits in diseases which impose a large health burden and/or where there are wider social benefits for patients, their carers and the wider economy. The methods developed in this research will allow the same weights to be also attached to the type of health that is lost and estimate the wider social benefits that are likely to be lost when the NHS must accommodate the additional costs of new drugs.

10 Research recommendations

10.1 Update estimates of the threshold with more recent and future waves of expenditure and mortality data.

10.2 If other aspects of social value are applied to health benefits of a new technology they must also be attached to the type of health that is likely to be forgone due to additional NHS costs. The methods developed here can be extended to allow weights to be also attached to the type of health that is forgone and estimate the wider social benefits that are likely to be lost when the NHS must accommodate the additional costs of new drugs.

10.3 We have demonstrated that these methods of analysis can be applied to quality of life data collected as part of PROMs. This type of analysis could be applied to these data in key PBCs as PROMs is rolled out providing some evidence about the quality of life effects of changes in PBC expenditure.

10.4 A key PBC is Mental Health. Currently outcomes data that could be linked to measures of quality of life are routinely collected in primary care. In principle, the same methods of analysis can be applied to these data once they are made available providing some evidence about the quality of life effects of changes in mental health expenditure.

10.5 Improved and more recent estimates of incidence (by age and gender) and duration of disease will soon be available from the recently published updated WHO Global Burden of Disease study. These data could be used when the threshold is re-estimated for later waves of expenditure data. Alternatively, estimates could be based on CPRD data.

10.6 Estimating a more complex lag structure based on the evolving panel data would provide valuable evidence about the duration of the health effects of changes in expenditure. The recent release of census data for 2011 may allow a panel model to be estimated allowing better control for unobserved heterogeneity across PCTs as well as exploiting variation in outcomes, expenditure and other

covariates over time. The formation of Clinical Commissioning Groups (CCG) in 2013 will make the time series problematic for waves of expenditure after 2012 unless it is possible to match CCG and PCT boundaries.

10.7 If PBC expenditure and outcome data are available at CCG level (as well as covariates and suitable instruments), it might become possible to estimate outcome and expenditure equations simultaneously across PBCs. This would enable more of the likely health effects of changes in expenditure to be reflected in the analysis.

Chapter 1: Introduction

1.1 Policy context

A comparison of the incremental cost effectiveness ratio (ICER) of a new technology with a cost-effectiveness threshold is not the only consideration when the National Institute for Health and Clinical Excellence (NICE) and its advisory committees issues guidance. But it is an important one as it allows an assessment of whether the health expected to be gained from the use of a technology exceeds the health expected to be forgone elsewhere as other NHS activities are displaced. For this reason a comparison of the ICER of a technology to a threshold range is a critical part of the reference case in the NICE Guide to Methods of Appraisal[1] and is often taken to be the starting point for deliberations about other considerations including judgements of social value. Therefore, the value of the threshold is critical to the assessment of whether technologies can be regarded as cost-effective. This is also true for other NHS resource allocation decisions which potentially impose additional costs on local NHS commissioners.

From 2014 the Government plans to introduce a new approach to determining the appropriate price of prescription pharmaceuticals. Under value-based pricing, the price the NHS pays for a new product will be directly linked to its cost-effectiveness.[2, 3] Therefore, the value of the threshold will be even more important as it will have a major impact on the prices that the NHS pays for pharmaceuticals, the access that NHS patients will have to new drugs and the return that manufacturers can expect from future research and development.[4, 5]

1.2 Estimating the cost-effectiveness threshold

A key part of NICE's remit is to make decisions which are consistent with the efficient use of NHS resources. In the context of the NHS budget constraint, a consideration of efficiency has to reflect the implications of imposing additional costs on the system which will displace existing services thus leading to health decrements for patients other than those benefiting from the new technology being appraised. The cost-effectiveness threshold is an estimate of health forgone as other NHS activities are displaced to accommodate the additional costs of new technologies. A national decision-making body like NICE needs an estimate of what is likely to be forgone across the NHS as we currently find it.[6] Of course, this will change as circumstances and the NHS change; tending to rise with increases in budget and health care costs but tending to fall with increases in the productivity of health technologies and the efficiency of the NHS in general - including better local commissioning decisions.[7] A body like NICE cannot and does not necessarily need to know what *specific* services and treatments will be displaced in particular localities or who will actually forgo health.

What is required, therefore, is an accountable and empirically-based assessment of the health that is likely to be forgone on average across the NHS. Currently NICE uses a threshold range of £20,000 to £30,000 per quality adjusted life year (QALY) gained, where additional considerations are required towards the upper bound.[1] The empirical basis of this range of values is very limited and there have been calls for further research in this area.[8] Explicit scientific methods are required which will provide accountability so that estimates can be scrutinised by a range of stakeholders. Since estimates of the threshold will need to be periodically revised, methods which make best use of routinely available NHS data are needed. As well as accountability, this will also provide more predictability in likely changes to the threshold for the investment decisions of technology manufacturers.

1.3 Aims and objectives

The aim of this research is to develop and to demonstrate methods to estimate the cost-effectiveness threshold for the NHS which makes best use of routinely available data. Methods are required which can capture the impact of a change in expenditure on length and quality of life (QoL), indicate how estimates of the threshold have changed over time, reflect uncertainty in any estimates and assess its implications, and indicate the impact of increases or decreases in spending. The project also aims to discuss options for developing data sources in the UK to estimate the threshold more precisely over time.
The research has four main objectives:

i. Informed by relevant literature, to provide a conceptual framework to define the threshold and the basis of its estimation.
ii. Using programme budgeting data for the English NHS, to estimate the cost per life year gained on average across the NHS, for marginal changes in budget.
iii. To extend the measure of the health effects of changes in expenditure by estimating the QoL associated with additional years of life and the direct impact of health services on QoL.
iv. To synthesise this work to bring evidence on life-years and QALYs together, to present the best estimate of the cost-effectiveness threshold given existing data, to show the implications of the uncertainty in the current evidence and to provide recommendations for future data collection and analysis.

1.4 Report structure

The main report is set out as a series of chapters, most of which are linked to more detailed analysis in separate appendices. Chapter 2 provides a policy context for the research and a conceptual framework for the subsequent empirical work. Chapter 3 outlines a simple theoretical model and associated econometric analysis of programme budgeting data to estimate the link between changes in overall NHS expenditure and mortality. Chapter 4 considers a range of analyses to extend the measure of health effect from mortality to life-years gained and to QALYs. Chapter 5 draws out the main conclusions and insights from the research.

Chapter 2: Policy Context and Conceptual Framework

2.1 Introduction

The purpose of this chapter is to provide the foundation for the empirical chapters that follow. It addresses a series of questions regarding the nature of the cost-effectiveness threshold that NICE use to guide its decisions, and the principles of how it should be estimated.

The chapter is informed by the results of a systematic literature search relating to these questions. Details of the methods and results of that search, together with a summary of the papers identified, are provided in Appendix A. In brief, the search uses a 'pearl growing' method to identify relevant papers. This identifies a number of initial key articles ('pearls') on the basis of expert advice, and 'grows' these pearls in a series of steps: extraction of citations and references from the initial pearls; identification of further pearls from cited and referenced papers; repetition of citation and reference searches; and manual search of references. This process is repeated until no further papers of relevance are identified. On this basis, 76 relevant papers were identified and are referred to, when relevant, in this chapter.

This chapter is organised as follows. The next section considers, at a conceptual level, what the cost-effectiveness threshold to inform NHS decisions, such as those made by NICE's advisory committees, should represent. Section 2.3 considers alternative routes to generating an empirical estimate of such a threshold. The final section provides a brief overview of the methods used in the study.

2.2 What should the NICE threshold represent?

2.2.1 The threshold as a measure of opportunity cost

NICE uses cost-effectiveness analysis (CEA) to inform the decisions underlying most types of guidance that it publishes. The use of CEA is most prominent in appraisals relating to new medicines,[1] but is also a key input into diagnostics appraisals as well as clinical guidelines and public health guidance.[1, 9] For those interventions and programmes which impose additional costs on the NHS budget, their incremental cost-effectiveness ratios (ICERs) indicate the incremental cost per additional quality-adjusted life-year (QALY) achieved relative to appropriate comparators. Although the ICER is one of a number of evidential inputs into NICE committees' decisions, is has been shown to be the most important, at least for technology appraisals.[10]

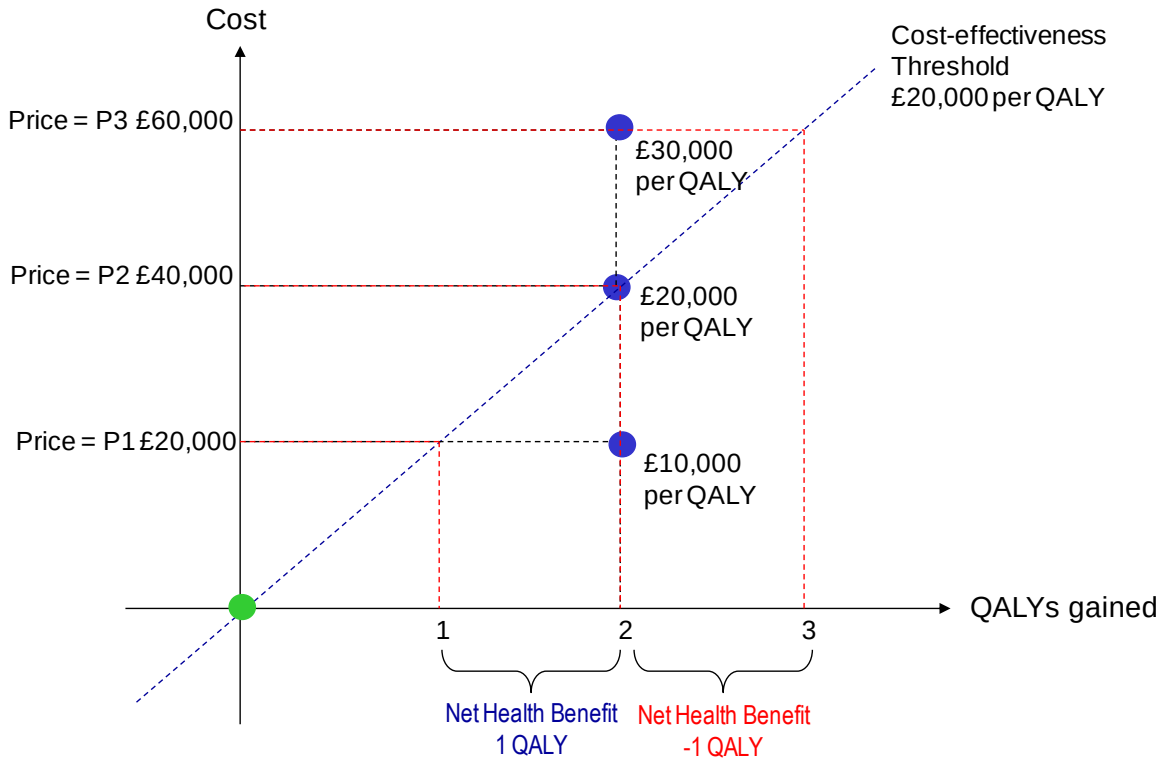
Interpreting whether a given ICER is acceptable requires the use of a cost-effectiveness threshold. Given that NICE has no influence on the level of the NHS budget, its decisions need to consider that budget a fixed constraint.[6] Therefore, the threshold should reflect the opportunity costs, in terms of health forgone, resulting from the imposition of additional costs on the NHS. When NICE issues positive guidance for a new intervention which imposes additional costs on the system, the resources required to deliver it must be found by disinvesting from other interventions and services elsewhere.[11] This displacement of existing services will result in health decrements for other types of individual.[12] Thus the threshold represents the additional cost that has to be imposed on the system to forgo 1 QALY worth of health through displacement.

Resource allocation decisions based on comparing an ICER with a cost effectiveness threshold uses some simplifying assumptions including those of constant returns to scale and perfect divisibility of programmes.[13] Some have suggested that this makes these methods unreliable,[14] although it has also been argued that they provide useful approximations to guide decisions.[15] This report takes NICE's use of these methods as a starting point, and does not review the literature relating to this debate in any depth.

As Figure 2.1 illustrates, CEA effectively becomes an analysis of net health benefits: does the health gain from the new intervention outweigh the health decrements associated with the displacement of existing services necessary to fund it? Figure 2.1 shows the incremental costs and QALYs associated with a new intervention relative to a comparator (the latter being shown at the origin). The new intervention

generates 2 additional QALYs per patient and, at price P1, imposes an additional £20,000 per patient; the ICER is, therefore, £10,000 per QALY gained. At a threshold of £20,000 per QALY, the additional cost of £20,000 per patient translates into a decrement of 1 QALY (the distance between the y-axis and the threshold). This is because the threshold indicates the additional cost that needs to be imposed on the NHS budget in order to displace services that result in 1 QALY being forgone. Therefore, at that price, there is a net health gain of 1 QALY per patient (2 gained from the new intervention and 1 forgone through displacement). At a price of P2, the additional cost per patient of the new intervention is £40,000 and the net health gain is zero: the 2 additional QALYs from the new intervention are the same as the QALYs forgone through displacement. At the highest price of P3, the adoption of the new intervention would actually result in a net health decrement of 1 QALY as it generates fewer QALYs (2) than are forgone (3).

Figure 2.1: graph showing illustration of the NICE threshold as a basis for assessing net health benefit. Adapted from Claxton *et al*[4]



The use of the threshold to facilitate this net health benefit (NHB) analysis can be expressed as in Equation 2.1:

$$NHB = \Delta h - \frac{\Delta C_h}{k} \qquad \text{Equation 2.1}$$

where Δh is the change in health generated by the new intervention, ΔC_h is the additional health care cost imposed on the NHS, and k is the cost effectiveness threshold. The net health gain from adopting the new intervention is, therefore, the health gained, Δh, minus the health forgone, $\frac{\Delta C_h}{k}$.

Understanding the NICE cost-effectiveness threshold as representing opportunity costs in terms of health is explicit in NICE documentation (for example, the methods guide for technology appraisal[1]). It is also clear in reports published by the Department of Health, such as the consultation report on value-based pricing.[8, 16, 17] This conceptualisation of the principles of the NICE threshold is also described in the broader literature.[6, 7] Formally, the threshold can be seen as the shadow price of the budget constraint.[6, 7, 12, 18-20] Although this project focussed on the use and estimation of a cost-

effectiveness threshold for NICE decisions, the methods and estimates relate to any resource allocation decision within the NHS where the opportunity cost could fall anywhere in the system. Hence it could apply, for example, to Department of Health targets or to Commissioning Board directives, as well as NICE guidance.

2.2.2 The threshold as the consumption value of health

Another view of what the threshold used in CEA should represent exists in the literature, however. In general terms, this is based on the rate at which individuals are willing to forgo other forms of consumption to achieve health improvement (sometimes referred to as 'willingness to pay').[21-40] Although this consumption value of health can provide information on the value of health improvement and may guide decisions such as the level of the overall NHS budget, it does not inform decisions regarding how to allocate a fixed budget within the health care system.

The reason for this is that the consumption value of health applies equally to health gained as well as to health forgone. This is shown in Equation 2.2 where the consumption value of health, v, is added to the definition of NHB in Equation 2.1. This simply involves valuing both health gained and health forgone by the same consumption value of a unit of health, *v*. Therefore, the use of the consumption value is irrelevant: a treatment considered cost effective in Equation 2.1 (i.e. to have a positive NHB) will inevitably be considered cost-effective in Equation 2.2, and an intervention with negative NHB (i.e. not cost effective) will remain as such in Equation 2.2.[1] Therefore, the magnitude of the threshold, k, is not a value judgment but an empirical question which can, in principle, be estimated.

$$NHB = v.\Delta h - \frac{v}{k}\Delta C_h$$

Equation 2.

2.3 Estimating the threshold

2.3.1 NICE's threshold range

NICE has been reluctant to specify a single cost effectiveness threshold used in its decision making.[10] It has also consistently emphasised that factors other than CEA are taken into consideration by the various advisory committees.[1, 9, 10, 41-43] Therefore, it has preferred to indicate the range within which its threshold value lies - £20,000 to £30,000 per QALY gained.[1, 9] Alongside this, it has provided an indication of the role other factors play in determining which point of threshold range is relevant. The latest guide[1] suggests that an ICER below £20,000 is likely to lead to recommendation unless the evidence is considered highly uncertain; an ICER between £20,000 and £30,000 will lead to recommendation if the committee is also happy with the levels of uncertainty in the evidence and/or the QALY does not capture all aspects of benefit; and an ICER above £30,000 would only be recommended if issues related to levels of evidential uncertainty and a failure to capture all benefits in the QALY are particularly compelling.

---

[1] This is the case so long as all incremental costs are health care system costs or, as currently, the perspective adopted by NICE is commonly restricted to the health care system. If a broader perspective was to be adopted and, insofar as there are some incremental costs (or benefits) of adopting a technology that fall on private consumption, then v does become relevant to decision making because it represents the value of these consumption effects in terms of health. In these circumstances it would be inappropriate either to compare an ICER which included consumption effects to k (because consumption costs do not displace health in the NHS), or to compare it to v (because some of the costs do not displace private consumption but displace health at rate k). The ratio of k/v represents the value of NHS resources relative to private consumption. Observing k < v would suggest a positive shadow price on NHS resources and public expenditure more generally, i.e., it would indicate that a public sector £ is scarce relative to a private £. See Claxton K., Walker S., Sculpher MJ. and Palmer S. Appropriate perspectives for heath care decisions. Centre for Health Economics, University of York. CHE Research Paper 54; 2010 for a more extended treatment of perspective, the implications for decision rules and the centrality of an estimate of the threshold, k.

In the following year, NICE issued further supplementary guidance relating to the appraisal of interventions for patients with short life expectancy, although this can be considered to relate more to the measure of benefit than factors to be considered outside of cost effectiveness.[44] In 2012 NICE issued a draft update of its methods guide which added that, if a new technology has an ICER above £20,000 per QALY, the committee's deliberations would also consider 'aspects that relate to non-health objectives of the NHS' (e.g. wider social considerations and/or costs that fall outside of the NHS budget).[45]

Although NICE has carefully argued the case for why its decisions are not driven entirely by a comparison of the ICER with its threshold range, it has not provided any empirical evidence for why the threshold range takes the value it does. Indeed it has been widely argued than an empirical basis for these values should be generated.[8, 46-50] For example, the House of Commons Health Select Committee in 2008 argued:

> *"The affordability of NICE guidance and the threshold it uses to decide whether a treatment is cost-effective is of serious concern. The threshold is not based on empirical research and is not directly related to the budget, it seems to be higher than the threshold used by PCTs for treatments not assessed by NICE. Some witnesses, including patient organisations and pharmaceutical companies, thought that NICE should be more generous in the cost per QALY threshold it uses, and should approve more products. On the other hand, some PCTs struggle to implement NICE guidance at the current threshold and other witnesses argued that a lower threshold should be used. We recommend that the threshold used by NICE in its full assessments be reviewed; further research comparing thresholds used by PCTs and those used by NICE should be undertaken...." ([8], page 6).*

2.3.2 The basis for empirical work

Although there is acceptance of the need for empirical work on the NICE cost-effectiveness threshold, a set of issues exists regarding the starting point for such analysis. One aspect of this is the view that the nature of the services that are displaced in response to additional costs being imposed by NICE guidance, and hence the magnitude of the health forgone for other patients, will depend on the productivity of the NHS and its overall (inflation adjusted) budget, both of which have increased since NICE initially defined its threshold range.[51, 52] In principle an increase in the (real) NHS budget would allow it to introduce interventions which were previously not cost effective which might be expected to increase the threshold if these interventions were the marginal ones displaced in response to the budget impacts of NICE recommendations. However, any increase in the NHS budget may be allocated to non-discretionary expenditure. This would include, for example, expenditure relating to national initiatives such as new contracts for consultants and activities to meet waiting list targets as well as, of course, the implementation of NICE guidance. The non-discretionary nature of such expenditure means that these types of activities cannot easily be disinvested from given a need to release resources to fund NICE guidance. Therefore, if an increase in the NHS budget is largely devoted to these types of non-discretionary expenditure, there will be a limited impact on the threshold.

Gains in productivity may come through doing worthwhile activities more cost effectively, including for those marginal interventions displaced by NICE recommendations, suggesting a reduction in the threshold. Alternatively, productivity gains might come through discontinuing activities which are not worth doing (i.e. that produce no health improvement), freeing resources for additional cost effective interventions which may be the marginal services displaced by NICE guidance - this can have the result of increasing the threshold.

The net effect of these changes on the threshold could not be determined *a priori* and would depend how any additional (real) budget were allocated and how the gains in productivity where achieved. This does emphasise the fact that the threshold may change over time in response to these and other broader developments, and this would have to be considered as part of any regular updating of the empirical analysis of the threshold.

A second issue to be considered relates to how decisions are taken locally about any displacement following NICE guidance. The principles of CEA suggest that such displacement should relate to interventions which are the least cost effective of those currently covered by the budget.[18] The basis

for how local commissioners and providers make their disinvestment decisions is not clear, however, and there have been calls for greater transparency and guidance in this area.[51] It would be entirely unrealistic to assume that displacement only takes place in those existing services which are the least cost effective. The reality is that numerous criteria are likely to be used by commissioners in implementing disinvestment, and that significant variation will exist between local decision makers.[12] Such criteria might include, for example, equity concerns about a particular disadvantaged group locally or capacity constraints regarding particular services. Therefore, NICE needs to know what is likely to happen on average across the NHS given the reality of local decisions. If local decision making changes over time - for example, if local commissioners become more focussed on displacing services which are the least cost effective, in terms of population health - this may affect the estimate of the threshold.

2.3.3 Studying displacement locally

A reasonable conclusion from a consideration of these issues is, therefore, that local decisions about disinvestment are likely to be an important determinant of the NICE threshold.[53-58] Appleby *et al* sought to assess whether it was possible to study local decisions about service investment and disinvestment to infer the cost effectiveness thresholds being used (implicitly) locally and to draw conclusions about the appropriate level of the NICE threshold.[59] They identified six primary care trusts (PCTs) and undertook structured interviews with each of the directors of public health. They also administrated questionnaires to an opportunistic sample of finance directors from NHS trusts. On this basis they developed a list of new services as well as those that had been deferred or discontinued. An attempt was made to estimate the implicit local ICER relating to these decisions by using any cost effectiveness evidence used to inform the decisions together with relevant evidence on cost effectiveness from the published literature.

The study found it quite straightforward to identify specific services that had been introduced, discontinued or deferred, but concluded that these decisions were typically based on clinical and other non-economic factors. A number of 'decisions at the margin' were identified but none of these was based on cost effectiveness analysis. Instead, the basis for changes in services was a 'business case', or overall cost impact. It was possible to impute cost effectiveness for most of the services affected, but the study concluded that, even with a larger sample of commissioners and providers, it would be very difficult to estimate an implied cost effectiveness threshold locally. This would be because, firstly, most PCT decisions were service reconfigurations including demand management and waiting list initiatives. By their nature, teasing out the incremental cost and health effects, potentially across numerous types of patients, would be an enormous challenge. Secondly, there would be difficulty in identifying all local decisions as many options for investment, deferment or discontinuation are rejected before they are made more explicit in documentation. A third problem would be the finding that a range of criteria is used to make local decisions, with relatively little concern for cost effectiveness , making a local threshold estimated in this way hard to interpret. A final challenge would be that it would be very difficult to establish a causal link between a change in local NHS budget and specific local investment and disinvestment decisions. The Appleby *et al* study highlights the problems that exist in deriving a cost effectiveness threshold from a bespoke study of specific local resource allocation decisions.

2.3.4 What evidence is needed?

Given the challenges of studying local decisions as a means of establishing the NICE threshold, and keeping in mind NICE's remit, it is possible to suggest a series of important characteristics that estimation methods should have from the perspective of principle and practice:

- They should reflect the effect of NICE guidance on the *average* of the displacement decisions taken across the NHS, with less consideration on which types of patients and interventions are affected and why the decision are taken. NICE cannot be expected to reflect what is likely to be marked variation between local commissioners and providers in how they react to an effective reduction in their budget as a result of positive guidance. Given NICE's remit, it is the expected health effects (in terms of length and quality of life) of the average displacement

within the current NHS (given existing budgets, productivity and the quality of local decisions) that is relevant to the estimate of the threshold.

- The methods used should not be a 'once and for all' effort but should facilitate regular updates to reflect changes in the broader NHS context such as changes in the overall real budget and productivity. This requires the use of data sources that are currently routinely available, are expected to become so in the future or could be made available at reasonable cost. It may be possible to glean some idea of how the threshold may change in the future by studying how it has changed in the past, which would require routine data sources to extend back over a period of time. Periodic updating using explicit scientific methods would encourage accountability through scrutiny of estimates by relevant stakeholders. It would also provide more predictability in likely changes to the threshold for the investment decisions of technology manufacturers.

- The nature of the displacement of existing services (and hence the magnitude of the health forgone) will depend on the scale of the budget impact coming through NICE guidance. Therefore, the methods used to estimate the threshold should ideally be able to reflect this budget impact.

- The methods should recognise the inevitable uncertainty relating to the evidence currently available for threshold estimation and translate this into an expression of the uncertainty in the estimate of the threshold. As well as providing information with which NICE can determine the appropriate implications for its choice of a threshold value, this consideration of uncertainty can help to prioritise further research or the collection of routine data.

2.4 An introduction to study methods

The current study has sought to develop methods consistent with these desired characteristics. This section provides a summary of the methods used. Further details are provided in each of the later chapters relating to the various components of work, and in the associated appendices. The general approach taken is to use routinely available data to look at the relationship between overall NHS expenditure and patients' health outcomes. By exploiting differences between PCTs in expenditure and outcomes, it is possible to infer the costs of generating health improvement from NHS services at the margin. In principle, this is what is needed as the basis of the NICE cost effectiveness threshold as it provides an indication of the health forgone through the services displaced by the additional budget effect of the Institute's guidance.

2.4.1 Past work

The study was able to build on some key existing research relating to the relationship between NHS expenditure and mortality.[60-62] Since 2003 data on expenditure on health care across 23 programme budgeting categories (PBCs) of care have been available for each PCT in the NHS in England. These programme budgeting data seek to allocate, to broad areas of illness according to the primary diagnosis (using ICD10 codes) all items of NHS expenditure, including expenditure on inpatient care, outpatient care, community care, primary care and pharmaceuticals and devices.

For the purposes of this study, the merit of these data is that they open up the possibility of examining the relationship between differences in local spending and associated disease-specific mortality outcomes routinely available from the National Centre for Health Outcomes Development. In each programme, the elasticity of outcome with respect to changes in expenditure was estimated controlling for differences between PCTs in need. Changes in mortality were then transformed into life-years gained using assumptions regarding life expectancy without the change in expenditure. This provides estimates of the marginal cost per life-year gained on average across the NHS by PBC.

This work focused largely on spending and outcomes in two of the largest programmes: circulatory disease and cancer,[63] but has also informed the link across other programme categories.[61, 64]

Estimates of the cost per life year gained for 2006/07 were £15,387 for cancer, £9,974 for circulation problems, £5,425 for respiratory problems, £21,538 for gastro-intestinal problems and £26,428 for diabetes. These estimates were based on a straightforward, though carefully constructed, theoretical model of health production which informs the specification and estimation of a set of equations. These dealt with the challenge of there being alternative plausible directions of causation - for example, between expenditure and health outcomes within a programme. This problem of endogeneity was addressed by identifying and testing suitable instrumental variables. In doing so, they accounted for variation in the clinical needs of the local population relevant to each programme together with broader local environmental factors relevant to the costs of care and outcomes.

This earlier work provides a strong foundation for the current study through its consideration of the average marginal elasticity of outcome with respect to programme expenditure. However, to estimate the threshold suitable for NICE decision making, a number of further elements of research are necessary, and these are described below.

2.4.2 Further econometric analysis

This further econometric research is covered in Chapter 3, with full details in Appendix B. The earlier work estimated the cost per life-year gained for the major programme areas. The NICE threshold needs to relate to the whole NHS and will, therefore, depend on all the programmes of care where disinvestment takes place. Given that each programme of care has been estimated separately, it is not clear how expenditure on particular programmes changes with the overall budget. For example, does disinvestment tend to fall on respiratory care or diabetes following a budget impact from NICE guidance? Therefore, the current study has further developed the econometric analysis to reflect the need for PCTs to operate within a fixed overall budget. This provides an estimate of the 'budget elasticity of expenditure' in each PBC, and facilities estimates of the impact of marginal increases (or decreases) in overall PCT budgets on spending in each PBC.

As well as indicating budgetary influences on programme spending these have then been linked to changes in mortality outcomes by programme. These changes are used to estimate years of life lost taking account of the fact that some of the observed deaths would have occurred anyway (had the same population not been at risk in the particular PBC); that is, taking account of unobserved counterfactual deaths. This takes into account how such budgetary changes (such as those imposed by NICE guidance) translate through local decisions into changes in expenditure on programmes of care and then to health outcomes.

Changes in budgets are in practice incremental rather than marginal, and it may be the case that the outcome elasticities of programme expenditure in times of budgetary increase (when new initiatives are introduced) are not the same as in times of budgetary decrease (when the focus is on disinvestment). The possible effects of non-marginal changes have, therefore, been explored. The project has also sought to explore how both expenditure and outcome elasticities, and hence the threshold, vary over time, and this has been assessed by generating relevant estimates for three sets of data.

A development from earlier work has been to relate expenditure in period *t* to mortality in periods *t, t+1* and *t+2*. Whilst the data used are largely cross-sectional, mortality data are linked so as to follow expenditures. Given the inevitable uncertainty relating to assumptions in the analysis, extensive sensitivity analysis is undertaken to consider the implications for the estimates.

2.4.3 Moving from life-years to quality-adjusted life-years gained

A key element of the research has been to take the results of the econometric work linking NHS spending and mortality, and to translate this into effects on life years and quality adjusted life years (QALYs). The methods planned for the study included a consideration of local data, collected routinely by PCTs, on the types of interventions in which local decision-makers were investing and disinvesting. The aim was to inform the link between the effects of expenditure changes on mortality and impacts on broader health in

terms of QALYs. These data may have indicated the types of interventions and services, within a given PBC, on which investment and disinvestment were taking place. Using targeted literature reviews, estimates of QoL for those activities may have been identified. However, it was established that there were limited data available at a local level to facilitate this type of analysis, so other data sources were used for this purpose (see Addendum C2).

It has, therefore, been necessary to consider alternative data and approaches. This is tackled using three sequential steps:

i. Translate the estimated effects on mortality from the econometrics work into life years by exploring the limitations of the mortality data available at PCT level and the published years of life lost (YLL) figures used in the econometric analysis, and by considering how to improve the estimates using additional data and analysis.
ii. Consider how estimates of life year effects can be adjusted for the quality of life in which they are lived, taking account of the gender and the age at which life years are gained or lost as well as the quality of life implications of particular diseases.
iii. Explore ways to take account of those effects on health not directly associated with mortality and life year affects (i.e., the 'pure' quality of life effects) to estimate an overall cost per QALY threshold.

This aspect of the analysis is described in Chapter 4 with further details provided in Appendix C.

The central or 'best' estimate is based on two assumptions relating to the health effects associated with expenditure, one conservative and the other more optimistic. The first assumption is that the health effects of changes in one year of expenditure are restricted to one year. This is implicit in the estimates of outcome elasticities estimated in the econometric analysis. This is likely to underestimate effects on mortality since expenditure that reduces mortality risk for an individual in one year may well also reduce their risk over subsequent years, and expenditure may also prevent disease in future patient populations. Therefore, total health effects will be underestimated and the cost per life year or QALY threshold will be overestimated. Although undoubtedly conservative, it may be offset to some extent by the more optimistic assumption. It is assumed that any death averted by expenditure in one year will return the individual to the mortality risk of the general population, i.e., the years of life gained associated with each death averted are based on what would have been their life expectancy taking account of their of age and gender (using life tables for the general population).

The extreme upper and lower bounds for cost per life year and cost per QALY thresholds are based on making both of these assumptions either optimistic (providing the lower bound for the threshold) or conservative (an upper bound for the threshold). The lower bound is based on assuming that health effects are not restricted to one year but apply to the remaining disease duration for the population at risk during the expenditure year. The upper bound is based on the combination of assuming that health effects are restricted to one year and that any death averted is only averted for the minimum duration consistent with the mortality data used to estimate the outcome. It is very important to note that the lower and upper bounds are very much extreme values with limited plausibility.

2.5 Conclusions

A cost effectiveness threshold is needed to inform decisions by NICE, the NHS more generally or the Department of Health which reflects the fact that opportunity costs fall on services and population health at a local level. Given that it is (and will continue to be) unfeasible to know precisely which services are displaced across all localities within the NHS, the threshold should reflect the average implications on health of actual local decisions about marginal changes in local service caused by changes in expenditure. The absence of an empirical estimate of the threshold which reflects these principles lies behind the project. Using data routinely collected in the NHS or available data that could be routinely updated, the study is organised into two major parts. The first updates earlier analysis to estimate the relationship between NHS expenditure and mortality, and the second seeks to translate these mortality effects into the more general measure of health - the QALY.

Chapter 3: The link between NHS spending, mortality and the cost of a life year

3.1 Introduction

This section presents an overview of the econometric work undertaken to estimate the link between NHS spending and mortality and how this is used to calculate the cost of a life year. As well as providing the analytical foundations for estimates of cost per QALY threshold presented in Chapter 4 and 5 this work also contributes to the on-going debate about the extent to which additional health care expenditure yields improved patient health outcomes.

The work presented in this report takes advantage of the availability of two new datasets to examine the relationship between National Health Service (NHS) expenditure and mortality rates for various disease categories. One dataset contains mortality rates for various disease categories at the level of geographically defined local health authorities, Primary Care Trusts (PCTs). The other dataset presents NHS expenditure by PCT on 23 broad programmes of care (these programmes are listed in Table 3.1). This dataset embraces most items of publicly funded expenditure, including inpatient, outpatient and community care, and pharmaceutical prescriptions. NHS revenue derives almost entirely from national taxation, and access to the system is generally free to the patient. The system is organized geographically, with responsibility for the local administration of the NHS devolved to PCTs.[2] PCTs are allocated fixed annual budgets by the Department of Health, within which they are expected to manage the health care in the locality.

We employ a model that assumes that each PCT receives an annual financial lump sum budget and allocates its resources across the 23 programmes of care to maximize the health benefits associated with that expenditure. Estimation of this model using the expenditure and mortality data facilitates two related studies: first, a study of how changes in the NHS budget impact on expenditure in each care programme; and second, a study of the link between expenditure in a programme and the health outcomes achieved, notably in the form of disease specific mortality rates. The latter also permits the calculation of the cost of an additional life year for individual programmes of expenditure.

The work presented here draws heavily upon on previous studies using these data[60, 62, 63, 65, 66] and innovates in four major ways: (1) we relate expenditure in time period *t* to outcomes in periods *t*, *t+1*, and *t+2* combined[3]; (2) we present plausible outcome models for a large number of budgeting categories - previous studies have tended to focus on the four largest care programmes; (3) we present estimates of the cost of a life year for the enlarged number of programmes and, importantly, with the aid of assumptions about the productivity of programmes without a meaningful mortality-based outcome indicator, we extend our individual programme estimates to incorporate expenditure across all programmes of care; and (4) while the models we present appear well specified according to appropriate statistical tests, we subject our results to a substantial sensitivity analysis.

The next section presents a brief review of the relevant literature upon which the study builds. This is followed by a summary overview of our approach to estimating the cost per life year across the various programmes of care and the results obtained using Programme Budgeting data provided by the Department of Health. Further details of all aspects of the modelling approach, description of the data, the results we derive and calculation of costs per life year are set out in Appendix B. This section is intended to be supported by the information contained within Appendix B.

---

[2] Strictly speaking, these local health authorities are Primary Care Organisations (PCOs) but the vast majority of these are 'Trusts' and we retain this terminology throughout.

[3] Due to data limitations the cited studies were only able to relate expenditure in period *t* to mortality in periods *t*, *t-1*, and *t-2* combined. Such studies assumed that PCTs had reached some sort of equilibrium in the expenditure choices they make and the outcomes they secure.

3.2 Previous studies

One of the most fundamental yet unresolved issues in health policy is the extent to which additional health care expenditure yields patient benefits, in the form of improved health outcomes. The work of health technology agencies such as the English National Institute for Health and Clinical Excellence (NICE) has greatly improved our understanding at the micro-level of the costs and benefits of individual technologies. However, there remains a dearth of reliable evidence at the macro-level on the benefits of increased health system expenditure.

The empirical problems of estimating the link between spending and health outcomes are manifest. If one relies on a time series of health outcome data for an individual health system it is difficult to disentangle the impact of expenditure from a wide range of other temporal influences on health, such as technological advances, epidemiological changes, and variations in broader economic circumstances. Similar methodological difficulties arise if one attempts a cross-sectional comparison of different health systems. In particular, when seeking to draw inferences from international comparisons, researchers might have failed to adjust for all the potential external influences on health outcomes and this might account in part for their findings. For example, in an early cross-sectional study of 18 developed countries, Cochrane et al.[67] use regression analysis to examine the statistical relationship between mortality rates on the one hand and per capita GNP and per capita consumption of inputs such as health care provision on the other. They found that the indicators of health care provision were generally not associated with outcomes in the form of mortality rates. Thereafter, the failure to identify strong and consistent relationships between health care expenditure and health outcomes (after controlling for other factors) has become a consistent theme in the literature, whilst, in contrast, socioeconomic factors are often found to be good determinants of health outcomes.[68-70]

There is furthermore the possibility that indicators of health system inputs, such as expenditure, are endogenous, in the sense that they have to some extent been influenced by the levels of health outcome achieved. And the difficulty of satisfactorily estimating the impact of health system inputs on outcomes is compounded by the great heterogeneity of health care, the multiple influences on outcomes, and the rather general nature of the outcome mortality measure traditionally used. Consequently, the failure to detect a significant positive relationship between expenditure and health outcome might reflect the difficulties associated with any such study rather than the absence of such a relationship. For example, Gravelle and Backhouse[71] examine some of the methodological difficulties associated with empirical investigation of the determinants of mortality rates. These include simultaneous equation bias and the associated endogeneity problem (that the level of health care input might reflect the level of health outcome achieved in the past), and that a lag may occur between expenditure and outcomes (studies typically assume that expenditure has an immediate effect on mortality).

To avoid the difficulties imposed by data heterogeneity inherent in international analyses, the study by Cremieux et al.[72] examines the relationship between expenditure and outcomes across ten Canadian provinces over the fifteen-year period 1978-1992. They find that lower healthcare spending is associated with a significant increase in infant mortality and a decrease in life expectancy. Although challenging the received empirical wisdom, one difficulty with the Cremieux et al.[72] study is that the estimated regression equation consists of a mixture of potentially endogenous variables (such as the number of physicians, health spending, alcohol and tobacco consumption, expenditure on meat and fat) and exogenous variables (such as income and population density). The authors' chosen estimation technique (GLS) does not allow for this endogeneity and consequently the coefficients on the endogenous variables may be biased.[71] Similarly, Nixon and Ulmann's[73] study, which uses three health outcome measures and various explanatory variables (such as per capita health expenditure) for 15 EU countries over the period 1980-1995, does not allow for the possibility that some of the explanatory variables may be endogenous.

More recently, studies have started to address the endogeneity issue.[74, 75] Bokhari, Gai and Gottret[74] estimate a cross-section model for 127 countries using data for 2000. They employ two health outcome indicators (the under-five mortality rate and the maternal mortality rate). Bokhari et al. allow for the endogeneity of health expenditure via the use of instrumental variable techniques, and they

estimate the elasticity of these indicators with respect to total government health expenditure conditional on the level of education and basic infrastructure (such as road transport and sanitation). They find that health expenditure has a statistically significant negative impact on both under-five mortality and maternal mortality. The authors do note, however, that their focus on child and maternal mortality implicitly assumes that these outcome indicators are in some way representative of outcomes across all activities financed by government health care expenditure. Data permitting, it would be preferable to relate health care expenditure on those aged under five years to under-five mortality, and expenditure on maternal care to maternal mortality.

In this study we relate expenditure in a specific disease area to mortality associated with those diseases. We also address the endogeneity issue through the use of instrumental variables and, unlike previous studies; we examine the sensitivity of our results to questions of instrument validity. Moreover, although previous empirical work has been loosely based on the notion of a health production function, it has rarely been informed by an explicit theoretical model. This is understandable, as the processes giving rise to the observed health outcome are likely to be very complex, and any theoretical model might become rather unwieldy. However, this absence of a theoretical model has sometimes led to an atheoretical search for measures of health inputs demonstrating a statistically 'significant' association with health outcomes. In contrast, in this study we inform our empirical modelling with a theoretical framework. We believe that this may lead to a more convincing and better specified model of health outcomes than that used in many previous studies, and this model is outlined in the next section.

3.3 Modelling framework

In the literature on the relationship between health expenditure and health outcomes, the statistical model estimated often contains a mixture of exogenous variables (such as income and population density) and endogenous variables (such as health spending, the number of doctors, and spending on cigarettes and alcohol). In such circumstances, the application of ordinary least squares will lead to biased coefficients on the endogenous variables. To avoid this problem, Gravelle and Backhouse [71] recommend that analysts model, even if only informally, the decision making process which generates the observed data set.

To avoid the problem of simultaneous equation bias we have constructed a very basic model of the budgeting and outcomes data generation processes. In places, the model makes some heroic assumptions (which we hope to relax in future work) but the framework reveals some of the more salient features of the data generation processes.

We assume – quite realistically – that each PCT, *i*, receives an annual financial lump sum allocation, y_i, from the Department of Health and that total within year expenditure for each PCT cannot exceed this amount. We also assume – less realistically – that this lump sum is allocated across the *J* programmes of care (J=23) by a single decision maker (although we know that in practice the programme budget data will in part reflect the myriad of individual clinical decisions that health care professionals take every day and that these are decisions over which PCTs exercise little control).

We assume that each PCT adheres to a social welfare function, $W(.)$, that incorporates the health outcome (h) across all 23 programmes of care so that for each PCT

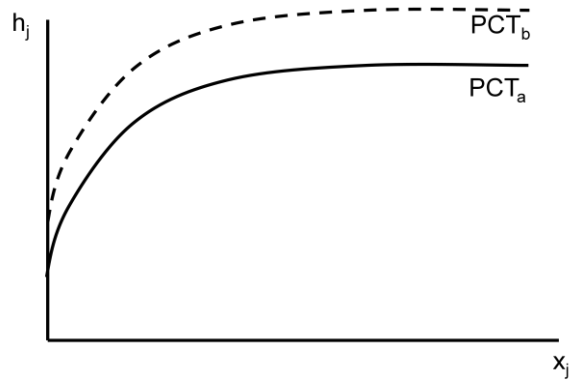
$$W = W(h_1, h_2, \ldots, h_J) \qquad \text{Equation 3.1}$$

Health outcomes might be measured in a variety of ways, but the most obvious is to consider some measure of improvement in life expectancy, possibly adjusted for quality of life, in the form of a quality adjusted life year.

We assume that, for each PCT and for each programme of care, there is a 'health production function' that indicates the link between local spending on programme j (x_j) and health outcomes in the same programme (h_j). Two such production functions are illustrated in Figure 3.1. We assume that increased

expenditure yields improvements in health outcomes, as expressed, for example, in local mortality rates, but at a diminishing rate. Clearly the shape of the curve might depend on the health needs of the local population (such as epidemiological conditions) and other local circumstances, such as socio-economic conditions and local service input prices. Note that in Figure 3.1 the cost of securing a given level of health outcome is – for whatever reason – higher in PCT_a than PCT_b.

Figure 3.1: The health production function for programme j in two PCTs



In algebraic form, each PCT seeks to maximise total welfare across all *J* programmes of care (*J*=23) subject to the health production function for each programme of care of the form

$$h_j = f_j(x_j, n_j, z_j) \qquad \text{Equation 3.2}$$

where n_j is the need for health care in programme j, x_j is PCT expenditure on programme j, and z_j represents environmental variables affecting the production of health outcomes in programme j (which might include private (non-PCT) health care expenditure in the disease area). Each PCT's problem is to select an expenditure level for each programme (x_j^*), so as to maximise the utility function in equation (3.1) subject to the health production functions in Equation 3.2 and the budget constraint that total expenditure on all programmes should not exceed PCT income (y).

Algebraically, the budget constraint is:

$$x_1 + x_2 + \ldots + x_{23} \leq y \qquad \text{Equation 3.3}$$

Solving this maximisation problem yields the result that the optimal level of PCT expenditure in each category, x_j^*, is a function of the need for health care in each category $(n_1, n_2, \ldots, n_{23})$,, environmental variables affecting the production of health outcomes in each category $(z_1, z_2, \ldots, z_{23})$, and PCT income (y). Thus:

$$x_1^* = x_1(n_1, n_2 \ldots, n_{23}, z_1, z_2 \ldots, z_{23}, y)$$

$$x_2^* = x_2(n_1, n_2 \ldots, n_{23}, z_1, z_2 \ldots, z_{23}, y) \qquad \text{Equation 3.4}$$

...

$$x_{23}^* = x_{23}(n_1, n_2 \ldots, n_{23}, z_1, z_2 \ldots, z_{23}, y)$$

These results imply that each PCT will allocate expenditure across the 23 programmes of care so that the

marginal utility of the last pound spent in each programme of care is the same. Of course, this does not mean that each programme receives the same amount of cash; financial allocations will depend on both the relationship between utility and outcomes, and on the relationship between outcomes and expenditure for each programme of care. If we assume that one extra unit of health outcome improves managerial utility by the same amount irrespective of the programme of care, then the decision maker simply allocates expenditure across all programmes to maximise total health outcomes. This is achieved by ensuring that the marginal health outcome benefit (measured perhaps in QALYs) is the same for the last pound spent across all programmes of care.

Thus, for each programme of care, there exists an expenditure equation (equation 3.4) explaining the expenditure choice of PCTs and a health outcome equation (equation 3.2) which models the associated health outcomes achieved. As presented, our basic model is static in the sense that the health production function (equation 3.2) assumes that all health benefits occur contemporaneously with expenditure. We acknowledge that for some programmes of care benefits might occur one or more years after expenditure has occurred. This is particularly likely to be the case for those programmes aimed at encouraging healthy lifestyles, where some benefits may occur decades after the actual programme expenditure. For other programmes, such as maternity/reproductive conditions and neonate conditions, benefits may be largely contemporaneous with expenditure. However, while our data are largely cross-sectional in nature, we are able to link mortality data in such a way that this follows expenditures. Accordingly, for our empirical modelling we estimate models using expenditure for period *t* with mortality data for periods *t*, *t+1*, and *t+2* combined. Appendix B presents a number of sensitivity checks on these assumptions including models where mortality data precedes expenditure data[4] and shows that these results are fairly consistent with the results presented here.

3.4 Data

3.4.1 Programme budgeting in England

Prior to October 2006, there were 303 PCTs in England with an average population of about 160,000 people. In October 2006 the 303 PCTs became 152 PCTs. Some PCT boundaries remained unchanged while other PCTs were merged with one or more neighbours to form a new, larger, PCT. In a few cases the geographic area covered by an existing PCT was split between two or more new PCTs. These 152 PCTs have an average population of about 330,000 people. PCTs are allocated fixed annual budgets within which they are expected to meet expenditure on most aspects of health care, including inpatient, outpatient and community care, primary care and pharmaceutical prescriptions.

Programme budgeting data collection was initiated by the Department of Health in April 2003 when each PCT was required to prepare expenditure data disaggregated according to 23 programmes of health care. These programmes are defined by reference to the International Classification of Diseases (ICD) Version 10 codes at the four digit level, and most programme budget categories reflect ICD 10 chapter headings (e.g., cancer and tumours, circulation problems, renal problems, neonates, problems associated with the skin, problems associated with vision, problems associated with hearing, etc). In some cases, the 23 categories are broken down into further sub-areas to achieve a closer match with the various National Service Frameworks (NSFs): for example, the large mental health category is broken down into 'substance abuse', 'dementia', and 'other'.
Programme budgeting seeks to allocate all types of PCT expenditure to the various programme budget categories, including secondary care, community care and prescribing. However, the system acknowledges that a medical model of care may not always be appropriate, and two specific non-clinical groups -- 'Healthy Individuals' and 'Social Care Needs' -- have been created. These are intended to capture the costs of disease prevention programmes and the costs of services that support individuals with social

[4] Due to data availability constraints previous studies had to relate expenditure in period t to mortality data in periods t, t-1, and t-2 combined. Implicitly this assumes that data represent a quasi long-run equilibrium position, and that relative expenditure levels and health outcomes within each PCT have been reasonably stable over a period of time.

rather than health care needs. In addition, in some cases it is not possible to assign activity by medical condition, preventative activity, or social care need and, in these cases, expenditure is assigned to a residual category (PBC 23) entitled 'Other'. The most important element of this residual programme is expenditure on general practitioner services (PBC 23a). In principle, it should be possible to allocate each GP consultation to a particular care programme. However, at the moment the available data information systems do not permit such an allocation and so all primary care expenditure is allocated to this residual programme. The use of this residual category ensures that all expenditure is assigned to a programme of care.[76]

The aim of the programme budget classifications is to identify the entire volume of health care resources assigned to broad areas of illness according to the primary diagnosis associated with an intervention. It serves a number of purposes, most notably to assist in the local planning of health care. But for this study its crucial merit is that it opens up the possibility of examining the statistical relationship between local programme spending and the associated disease-specific outcome. Various forms of data collection and analysis are required to map PCT expenditure on acute, community and other services to the 23 programme budget categories. From the PCT perspective, however, the construction of each PCT's return largely involves collating information provided by other bodies and drawing on other information already in the PCT's own annual accounts. Details of how expenditure is assigned to programmes of care can be found in Section B4.2 of Appendix B.

Table 3.1 shows the expenditure per head and the growth in this expenditure for each programme budget category for 2003/04 to 2008/09.5 Year on year comparisons of expenditure in each group are complicated by the fact that the algorithms used to allocate activity to PBCs are regularly revised.6 However, by 2008/9 total PCT expenditure per person had increased to £1,531 (up 28% from 2004/5). The residual 'other' category (programme budget category 23) still accounted for the largest share of expenditure (14.9%) with per capita expenditure of almost £228, of which £145 was accounted for by primary care expenditure. Mental health (budget category 5) accounted for just over 12% of expenditure, but the expenditure share recorded by circulation problems (budget category 10) had fallen from 10.2% to 8.5%. Other categories recording a fall in budget share of more than one half of one percentage point included: the gastro-intestinal system (down from 6.1% to 5.1%), the musculo-skeletal system (down from 6.0% to 5.2%), trauma and injuries (down from 6.0% to 4.2%), and maternity (down from 4.6% to 3.9%). Categories recording an increase in budget share of more than one half of one percentage point included neurological problems (up from 2.9% to 4.4%) and dental problems (up from 1.1% to 4.1%).

Some of these changes will partly reflect revisions to the algorithms used to allocate expenditure to particular PBCs. For example in 2006/7 expenditure per person on musculo-skeletal problems fell by 11% and expenditure on trauma and injuries fell by 25%. In the same year, expenditure on neurological problems increased by 35%. This suggests that some types of activity, which were previously allocated to musculo-skeletal problems and/or trauma and injuries, were re-allocated to neurological problems.

Similarly, up to and including 2006/7 expenditure that was not directly attributable to a particular programme category was apportioned using admitted patient care percentages.[7] In other words, if x% of total admitted patient care expenditure was allocated to PBC 1, then x% of all expenditure that was not directly attributable to a particular programme category was also allocated to PBC 1. With effect from 2007/8, however, NHS organisations were asked to select an appropriate basis for the apportionment of this non-programme specific expenditure and that, where no reasonable basis existed, such expenditure was to be allocated to the 'Other – Miscellaneous' (PBC 23X) category. These two changes to the algorithm used to allocate expenditure to particular PBCs illustrate that year-on-year comparisons of expenditure need to be interpreted with care.

---

[5] Comparable data for each programme budget sub-category is shown in Table BA.1 in Appendix B.

[6] These revisions are documented in Appendix B, Section B4.3.

[7] Expenditure on, for example, community care, A&E, ambulance services, and outpatients can be difficult to attribute a particular PBC. Critical care, rehabilitation, and specialised commissioning across care settings will also be difficult to attribute to a particular programme.

Expenditure per head on any given programme varies from one PCT to another and Table 3.2 presents some statistics that indicate the degree of variation in expenditure levels across PCTs by programme budget category. The first four columns of Table 3.2 present descriptive statistics for PCT expenditure per person. These reveal that, for example, PCT per capita expenditure in the cancer programme averaged £96.30 across all PCTs, with the minimum spend being £62.90 and the maximum being £155.70.

Some PCTs will be spending more than other PCTs simply because they face higher input costs. The second set of four columns in Table 3.2 present descriptive statistics for PCT per capita expenditure that has been adjusted for the unavoidable geographical variation in costs (input prices) faced by PCTs.[8] However, if anything this adjustment appears to increase the variation in expenditure across PCTs; for example, the range of per capita expenditure on cancer increases from between £62.90 and £155.70 (unadjusted) to between £59.10 and £163.10 (adjusted for local health care input prices).

Another cause of the variation in expenditure levels is the fact that the need for health care varies from one PCT to another. For example, areas with a relatively large proportion of elderly residents, or PCTs operating in relatively deprived locations, can be expected to experience relatively high levels of spending. The Department of Health has a well-developed methodology for estimating the relative health care needs, which it uses as the basis for allocating health care funds to.[77]

The final set of four columns in Table 3.2 present descriptive statistics for PCT per capita expenditure that has been adjusted for both the unavoidable geographical variation in costs and the local need for health care faced by PCTs.[9] For virtually every PBC, this adjustment reduces the variation in expenditure across PCTs; for example, the standard deviation of PCT per capita expenditure falls from £19.70 to £15.30 for the cancer programme. Although this adjustment reduces the variation in expenditure levels across PCTs, this decline is quite modest and there are still substantial differences in expenditure even after allowing for differences in local cost and need. For example, expenditure per head in the circulation problems category varies between £78 and £328 using cost adjusted expenditure data, but falls between £76 and £327 using cost and need adjusted population data.

The variation in expenditure across PCTs has led some commentators to question the reliability of the programme budgeting data. The National Audit Office[78] undertook a survey of Trusts, PCTs and SHAs to assess the quality of the data. They concluded that while the processes for collecting the budgeting data were well defined in most areas, there remained scope for improvements to the robustness of some of the data (e.g. non-admitted patient care). Appleby et al.[79] also considered the issue of data reliability in variations in spending on cancer services and noted some large year-on-year changes. However, the authors point out that it is difficult to define what might be either an implausible level of expenditure or an implausibly large change in expenditure. This is complicated by the fact that the Department of Health makes regular improvements to the way in which activity is matched to programme categories.

As with most datasets, there are likely to be recording and other errors associated with the programme budgeting data. However, while we note that the allocation of programme budgeting data might not be perfect there is no systematic evidence of this. Accordingly, for each disease category, we observe that PCT expenditure per person varies considerably and this variation – holding constant input prices and the need for health care – offers the opportunity to examine whether PCTs that spend more on health care achieve a better outcome and, if so, at what cost. Empirical estimates of the strength of this relationship for several programmes of care are presented in this report.

---

[8] This cost adjustment reflects the fact that health economy input prices vary considerably across the country and, for some inputs, are up to 40% higher in London and the south east of England than elsewhere. We have used a weighted average of the three Market Forces Factor Indices (MFFs) for HCHS, for prescribing, and for GMS/PMS to adjust the raw expenditure figures in Table 2 for local input prices (see Department of Health, 2009)

[9] This needs adjustment incorporates the AREA resource allocation formula for HCHS (see Department of Health, 2005).

3.4.2 Health outcome data

Most studies of the relationship between expenditure and outcome have used some measure of mortality as an indicator of the latter. We also employ mortality as an outcome measure. First, it is a relevant (albeit not comprehensive) measure of the outcome of health care expenditure; and second, it is available for more disease areas than any other outcome measure at PCT level.

Although mortality is available (by PCT) for several disease areas, it is not available for just over one-half of all programmes not least because it is simply not relevant for these programmes (e.g., for learning disabilities, vision problems, hearing problems, dental problems, and skin problems). Moreover, even where a mortality measure is available, the ICD10 coverage of the mortality data often falls short of the coverage of the expenditure data. For some programmes, therefore, we have combined the published mortality rates for two or more disease areas in an attempt to match the ICD10 coverage of the mortality data with that of the expenditure data.

Table B5.1 (Appendix B) shows how we have attempted to marry the mortality data (column c) and the expenditure data (column a). ICD10 coverage of the component mortality rates for some PBCs falls short of the expenditure data and the extent of this shortfall is illustrated by the ratio reported in the final column of Table 3.3. For example, the cancers and tumours programme covers all expenditure associated with ICD10 codes C00-C97 and D00-D49 but the PCT-based mortality data only relates to ICD10 codes C00-C97. At the national (all England) level, figures are available which show that, in 2008, there were 62,072 deaths of those aged under 75 years from codes C00-C97 and that there were 63,076 deaths from codes C00-C97 and D00-D49 combined. In other words, the PCT level mortality data reflects 98.4% of all deaths associated with the expenditure codes. We adjust our cost of life (year) estimates for this mismatch.

Table 3.1: National (all PCT) expenditure per head (£) and growth in expenditure (%) by PBC group, 2003/4 - 2008/9

PBC #	PBC description	Spend (£) per head 2003/4	Spend (£)per head 2004/5	Spend (£)per head 2005/6	Spend (£)per head 2006/7	Spend (£)per head 2007/8	Spend (£)per head 2008/9	Growth (%) 2004/5	Growth (%) 2005/6	Growth (%) 2006/7	Growth (%) 2007/8	Growth (%) 2008/9	Share of total spend (%) 2004/5	Share of total spend (%) 2008/9
1	Infectious diseases	17.95	20.22	23.61	20.88	22.08	23.46	13	17	-12	6	6	1.7%	1.5%
2	Cancers and tumours	64.95	75.54	83.24	81.67	90.21	94.55	16	10	-2	10	5	6.3%	6.2%
3	Blood disorders	14.08	17.00	17.48	16.58	19.44	19.50	21	3	-5	17	0	1.4%	1.3%
4	Endocrine, nutritional	28.96	31.86	37.26	36.70	39.39	43.38	10	17	-1	7	10	2.7%	2.8%
5	Mental health	133.31	146.83	158.95	166.53	180.90	191.21	10	8	5	9	6	12.2%	12.5%
6	Learning disability	37.93	43.37	46.54	48.36	54.20	56.11	14	7	4	12	4	3.6%	3.7%
7	Neurological	29.83	35.09	41.06	55.27	62.43	67.64	18	17	35	13	8	2.9%	4.4%
8	Vision problems	24.61	27.65	28.24	26.97	30.69	32.95	12	2	-4	14	7	2.3%	2.2%
9	Hearing problems	5.73	6.32	6.27	6.21	8.07	8.16	10	-1	-1	30	1	0.5%	0.5%
10	Circulatory disease	110.12	122.37	124.28	122.06	124.77	129.94	11	2	-2	2	4	10.2%	8.5%
11	Respiratory system	54.60	62.71	69.56	65.07	67.68	77.97	15	11	-6	4	15	5.2%	5.1%
12	Dental problems	10.78	13.55	24.91	51.93	59.45	62.44	26	84	108	14	5	1.1%	4.1%
13	Gastro intestinal system	63.56	73.22	81.30	73.30	75.05	77.89	15	11	-10	2	4	6.1%	5.1%
14	Skin problems	20.98	24.90	26.84	28.31	30.41	32.34	19	8	5	7	6	2.1%	2.1%
15	Musculo Skeletal system	61.36	71.72	74.74	66.75	75.91	79.68	17	4	-11	14	5	6.0%	5.2%
16	Trauma and Injuries	62.31	72.13	76.41	57.29	57.56	63.54	16	6	-25	0	10	6.0%	4.2%
17	Genito Urinary system	55.32	62.38	67.38	68.98	67.83	73.78	13	8	2	-2	9	5.2%	4.8%
18	Maternity	52.28	55.04	60.42	57.64	57.09	60.44	5	10	-5	-1	6	4.6%	3.9%
19	Neonate conditions	11.72	13.93	13.42	13.17	15.15	17.23	19	-4	-2	15	14	1.2%	1.1%
20	Poisoning	9.68	12.32	14.25	14.59	15.84	18.31	27	16	2	9	16	1.0%	1.2%
21	Healthy individuals	20.29	22.77	26.18	26.85	31.44	35.74	12	15	3	17	14	1.9%	2.3%
22	Social care needs	24.81	30.93	33.59	30.29	35.29	36.58	25	9	-10	17	4	2.6%	2.4%
23	Other (includes GMS/PMS)	136.94	157.75	171.82	209.70	232.02	227.71	15	9	22	11	-2	13.2%	14.9%
1 to 23	All PBCs	1052.12	1199.60	1307.76	1345.10	1452.91	1530.59	14	9	3	8	5		

Notes: (i) The population figures for 2003/4, 2004/5 and 2005/6 are identical (the total for England is 49,175,998).
(ii) The corresponding figure for 2006/7 is 50,476,231, for 2007/8 it is 50,695,989, and for 2008/9 it is 51,220,531.
(iii) The spend per head figures are calculated by summing expenditure across all PCTs and dividing by the national population.
(iv) All figures are at current prices.

Table 3.2: PCT expenditure per head by PBC, 2008/9: (a) unadjusted; (b) adjusted for local costs; and (c) adjusted for local costs and local need.

		Spend per head (unadjusted), £				Spend per head (cost adjusted), £				Spend per head (cost and need adjusted), £			
	Programme budget category	Mean	StdDev	Min	Max	Mean	StdDev	Min	Max	Mean	StdDev	Min	Max
1	Infectious diseases	26.5	24.6	8.6	151.8	25.7	21.7	8.6	136.7	25.0	21.4	9.5	139.5
2	Cancers and tumours	96.3	16.9	62.9	155.7	96.7	19.7	59.1	163.1	94.2	15.3	55.2	154.0
3	Blood disorders	20.3	7.0	7.7	49.4	20.2	6.5	8.0	49.1	19.7	6.0	8.2	44.2
4	Endocrine, nutritional	44.6	8.8	28.9	74.8	44.7	9.5	27.4	77.0	43.3	6.1	29.9	61.5
5	Mental health	201.4	60.0	118.9	474.1	200.3	54.0	122.8	422.8	194.0	41.9	132.3	362.0
6	Learning disability	56.8	18.8	7.7	125.9	57.0	19.4	6.8	123.6	55.7	18.8	6.7	136.6
7	Neurological	68.5	13.8	41.1	133.8	68.8	15.6	38.4	137.5	66.9	12.1	41.5	125.2
8	Vision problems	33.2	6.7	16.7	57.7	33.4	7.5	14.8	59.2	32.5	6.1	15.6	48.3
9	Hearing problems	8.6	3.7	0.9	24.0	8.7	3.9	0.9	25.5	8.3	3.3	0.8	22.0
10	Circulatory disease	131.6	26.7	88.0	317.3	132.2	30.5	78.2	327.6	128.5	24.4	75.7	326.9
11	Respiratory system	80.5	17.4	48.0	141.2	80.9	19.8	42.7	145.3	78.1	12.4	48.2	126.0
12	Dental problems	64.8	13.4	28.0	111.9	64.9	14.1	24.9	115.8	63.0	10.7	28.1	97.1
13	Gastro intestinal system	80.0	14.5	46.7	119.6	80.4	16.8	41.5	124.6	78.0	11.3	41.6	114.4
14	Skin problems	33.1	8.0	18.1	66.4	33.3	8.6	16.5	69.1	32.2	6.3	16.0	57.7
15	Musculo Skeletal system	79.9	17.6	43.3	127.3	80.4	19.9	39.6	132.5	78.2	16.6	41.0	116.4
16	Trauma and Injuries	63.2	16.7	12.5	139.3	63.4	17.4	11.5	125.0	61.8	15.6	10.4	103.6
17	Genito Urinary system	75.7	13.7	49.9	112.3	75.6	13.6	48.4	108.9	73.7	10.1	50.6	105.5
18	Maternity	63.3	16.7	24.6	124.4	63.1	15.8	21.9	117.9	61.4	12.8	24.4	96.5
19	Neonate conditions	18.4	7.3	6.4	46.4	18.2	6.8	6.6	43.7	17.8	6.6	5.8	47.8
20	Poisoning	18.6	4.2	10.8	31.2	18.7	4.7	9.6	32.3	18.2	3.9	10.1	33.1
21	Healthy individuals	38.4	18.1	9.7	125.0	38.4	17.8	8.9	115.6	36.7	14.5	9.4	104.5
22	Social care needs	40.8	56.6	0.1	415.2	41.2	59.2	0.1	432.9	39.7	55.0	0.0	411.5
23	Other (includes GMS/PMS)	230.8	44.5	138.2	396.1	230.2	42.4	140.7	356.5	226.8	45.8	134.1	346.0
All	All PBCs	1,575.6	196.7	1,225.7	2,079.9	1,576.3	217.3	1,183.0	2,173.1	1,534.0	86.2	1,390.1	1,987.0

Note: the above statistics relate to 152 PCTs and the mean expenditure figures will differ slightly from the national ones in Table 3.1 because the statistics across PCTs are not weighted for the size of each PCT's population.

We acknowledge that mortality is a more relevant outcome indicator for some programmes (e.g., for circulatory problems) than for others (e.g., for epilepsy) and, for this reason, we would expect better results in some programmes than others. We also acknowledge that this focus on mortality ignores the impact of expenditure aimed at chronic care and at palliative care. Nevertheless, our focus on mortality is purely practical: it is both a widely available measure and it is clearly a relevant outcome indicator.[10]

The mortality data provide us with a number of possible outcome indicators including the under 75 years of age standardised mortality rate (SMR) and the (under 75 years) standardised years of life lost rate (SYLLR). The SMR gives equal weight to all deaths irrespective of the age at which they occur but the SYLLR gives greater weight to deaths that occur at earlier ages. For our purposes we focus on a measure of the avoidable years of life lost (YLL).[11] This is calculated by summing over ages 1 to 74 years the number of deaths at each age multiplied by the number of years of life remaining up to age 75 years. The crude YLL rate is simply the number of years of life lost divided by the resident population aged under 75 years. Like conventional mortality rates, the crude YLL rate can be age standardised to eliminate the effects of differences in population age structures between areas, and this (age) standardised YLL rate is the health outcome variable generally employed in this study.[80]

3.4.3 Other variables

We employ an instrumental variables (IV) estimation technique to our empirical models of the outcome and expenditure equations as described in the next section. This is due to (i) own programme expenditure is likely to be endogenous in the outcome equation and (ii) other programme need is likely to be endogenous in the own programme expenditure equation. Endogeneity of programme expenditure results from expenditure levels being responsive to levels of outcomes and/or unobserved need rendering expenditure correlated with the residuals in an OLS regression of outcomes on expenditure. Due to limitations of data other programme need in the expenditure equation is proxied by death rates (minus that due to the programme under investigation). This will be influenced by expenditure decisions, including expenditure in other programmes and is treated as endogenous in the expenditure model.

IV estimation basically involves replacing the endogenous variable in the equation of interest with its predicted value from an OLS regression which regresses the endogenous variable on a set of instrumental variables. These instruments should be good predictors of the endogenous variable (i.e., they should be relevant and strong predictors) but should be appropriately excluded from the equation of interest (i.e., they should be valid instruments).

We have a number of potential instruments available, mostly derived from 2001 Population Census. In our earlier studies we found that a small sub-set of these instruments proved sufficient to generate plausible results. These included: the proportion of the population providing unpaid care; the proportion of households that are one pensioner households; index of multiple deprivation; proportion of the population in the white ethnic group.

We also had available a further set of potential instruments and, where our more limited set of instruments failed to generate plausible results, we extended our instrument search to include this wider set of variables. This extended set of instruments is shown in Table 3.3.[12]

---

[10] The approach adopted here is extendable in principle to other non-mortality based outcome indicators. We illustrate such an application in Section B8.8 of Appendix B where we use EQ-5D utility scores pre- and post- an operative procedure from the PROMs programme to generate a non-mortality-based outcome indicator, and we use this indicator to estimate our outcome model.

[11] One exception to this is the mortality rate for the trauma and injuries programme where initially only SMRs were available.

[12] Details of the construction of all instruments are shown in Table BA.2 of Appendix B.

Our instruments reflect factors, such as socio-economic deprivation and the availability of informal care in the community, which might indirectly impact upon mortality rates and/or health care expenditure levels. As we shall see, although our instruments 'pass' the appropriate statistical tests, some commentators claim that such tests may have 'low power' to detect the presence of invalid instruments. Consequently in section B9 of Appendix B we examine how sensitive our results are to the presence of invalid instruments.

Table 3.3 reports descriptive statistics for the socio-economic and needs variables used in the study (these statistics are for the variables in absolute form). For example, on average, lone pensioner households comprise 14% of all households, the 'white ethnic' group accounts for 89% of the population and 10% of the population provide unpaid care.

In addition to the instrumental variables, Table 3.3 also reports descriptive statistics for the Department of Health's 'need for health care' index,[13] its need for HIV services index, and its need for maternity services index. The latter two indices are used to either supplement or replace the all service measure of need when estimating our models. The 'need for health care' index averages about 1 but varies substantially, with some PCTs having a needs index more than 25% below the national average and others facing a need for health care more than 30% above the national average. The Table also reports descriptive statistics for some disease prevalence rates (e.g., for diabetes and for epilepsy) and, again, these are used to either supplement or replace the all service measure of need when estimating our models. Finally, the MFF index shows that input prices in the most expensive PCT are almost 20% above those in the least expensive PCT.

3.5 Approach to model estimation

The theoretical framework suggests the specification and estimation of a system of equations, with an expenditure and health outcome equation for each of the 23 programmes of care. However, this approach makes infeasible data demands, requiring variables to identify expenditure, need, environmental factors and health outcomes in each of the 23 programmes of care. Moreover, mortality rates are available for less than half of the 23 programmes. Rather than estimate a system of equations, we proceed on a programme-by-programme basis, estimating health outcome and expenditure equations for those programmes for which mortality data are available.

In line with the theoretical framework presented above, we specify the following expenditure (*equation 3.5*) and health outcome (*equation 3.6*) models for each of the 23 programmes of care. Accordingly, for the *j-th* programme of care we have:

$$x_i = \propto + \beta n_i + \gamma m_i + \theta y_i + \varepsilon_i, \qquad i = 1, \dots, 152 \qquad \text{Equation 3.5}$$

$$h_i = \rho + \delta n_i + \pi x_i + \epsilon_i, \qquad i = 1, \dots, 152 \qquad \text{Equation 3.6}$$

where x_i is expenditure; n_i is the own programme need for care; m_i is the need for care in other programmes; y_i is the total budget and h_i is the health gain in PCT $_i$.

Ideally we should employ a programme specific indicator of the level of need for each care programme (n_{ij}) but these are not readily available. When estimating both the outcome and expenditure models we therefore proxy the own programme health care need using the 'needs' component of the Department of Health's resource allocation formula.[14] This needs element is specifically designed to adjust PCT allocations for local health care needs and accordingly, ceteris paribus, we would expect a positive

[13] This incorporates the CARAN formula for HCHS and reflects need across all health care services.
[14] However, we do experiment replacing and supplementing this all service measure of need with more programme specific measures where these are available (e.g., the diabetes and epilepsy prevalence rates).

relationship between expenditure and need for each programme of care. We would also expect a positive relationship between need and adverse health outcomes.[15]

The expenditure model includes both the own programme health care need (which is proxied using the 'needs' component of the Department of Health's resource allocation formula) and the need for health care in all other programmes. In the absence of programme-specific measures of need, we use the 'all cause mortality rate excluding the mortality rate in the programme of interest', m_i, as the proxy for need in other programmes of care.

All variables have been log transformed so that parameter estimates can be interpreted as elasticities. In other words, a regression coefficient of 0.5 implies that a 1% increase in the regressor is associated with a 0.5% increase in the dependent variable.

3.5.1 IV estimation

Other programme need, m_i, in the expenditure equation 3.5 and expenditure, x_i, in the outcome equation 3.6 are both likely to be endogenous rendering OLS both biased and inconsistent. Endogeneity of programme expenditure results from expenditure levels being responsive to levels of outcomes and/or unobserved need. Other programme need in the expenditure equation is proxied by death rates which is influenced by expenditure decisions and hence is treated as endogenous. To deal with this endogeneity we employ instrumental variables (IV) estimation and implement two-stage least squares (2SLS). Unlike OLS, IV is a consistent estimator in the presence of an endogenous regressor and, although in finite samples the IV estimator will be biased, with the bias (providing certain assumptions are met) being less than that associated with OLS.

For the health outcome equation, IV estimation can be viewed as finding variables (instruments) that are good predictors of programme expenditure but which are appropriately excluded from the outcome equation of interest (that is, from equation 3.6) because they are not predictive of outcome. The assumption is that these instruments impact upon health outcome through their impact on expenditure only, and that they do not have a direct effect on the outcome.[16]

Similarly, for the expenditure equation, IV estimation can be viewed as finding variables (instruments) that are good predictors of the proxy for other programme need (m_i) but which do not belong in the expenditure equation of interest (that is, equation 3.5). The assumption is that these predictors impact upon own programme expenditure only through their impact on other programme need and that they do not have a direct effect on own programme expenditure.

The outcome and expenditure equations for any given programme may contain different instrumental variables because these instruments are trying to predict different variables (own programme expenditure and other programme mortality respectively). In addition, the instrument set for, say, the expenditure equation may vary across programmes because the other programme need variable will reflect need in a different basket of programmes for each expenditure equation.

---

[15] Whilst need is a function of mortality/morbidity in the resource allocation formula, the relationship is not sufficiently strong enough for us to be concerned about the endogeneity of the need in any individual care programme.

[16] IV estimation of say, equation 3.6, involves a first-stage regression of the endogenous expenditure variable, $\mathbf{x}$, on the instrument, $\mathbf{z}$, and the set of exogenous regressors in equation 3.6, $\mathbf{n}$. Predictions, $\hat{\mathbf{x}}$, from this model can then be included in a second-stage regression of equation 3.6 as a replacement for the endogenous regressor, $\mathbf{x}$.

We have a number of potential instruments available, mostly derived from 2001 Population Census. In previous studies, we have often found that a small sub-set (four) of these instruments often proved sufficient to generate plausible results. However, if plausible results were not obtainable with some combination of these four instruments, we employed an extended instrument set. Further details of the identification of suitable instruments for each model can be found in Section B7.3 of Appendix B.

Table 3.3: Descriptive statistics for the instrumental and other variables

Description	Obs	Mean	Std. Dev.	Min	Max
Proportion of residents born outside the European Union	151	0.0794	0.0876	0.0088	0.3817
Proportion of population in white ethnic group	151	0.8927	0.1299	0.3942	0.9926
Proportion of population of working age (16-74) with LLT illness	151	0.1182	0.0250	0.0709	0.1798
Proportion of population providing unpaid care	151	0.0990	0.0118	0.0662	0.1221
Proportion of population providing unpaid care (<20 hrs week)	151	0.0667	0.0079	0.0461	0.0817
Proportion of population providing unpaid care (20-49 hrs week)	151	0.0113	0.0025	0.0065	0.0195
Proportion of population providing unpaid care (>50 hrs week)	151	0.0210	0.0051	0.0093	0.0353
Proportion of population aged 16-74 with no qualifications	151	0.2960	0.0642	0.1301	0.4555
Proportion of population aged 16-74 that are full-time students	151	0.0720	0.0270	0.0425	0.1626
Proportion of households without a car	151	0.2932	0.1046	0.1325	0.5761
Proportion of owner occupied households	151	0.6692	0.1128	0.2891	0.8205
Proportion of households in rented social (LA/HA) housing	151	0.2071	0.0918	0.0817	0.5356
Proportion of households in rented private housing	151	0.0924	0.0449	0.0349	0.2961
Proportion of lone pensioner households	151	0.1434	0.0184	0.0979	0.1942
Proportion of one parent households	151	0.0684	0.0180	0.0401	0.1207
Proportion of population aged 16-74 that are permanently sick	151	0.0574	0.0213	0.0242	0.1215
Proportion of population aged 16-74 are long-term unemployed	151	0.0113	0.0052	0.0036	0.0287
Proportion of 16-74 in employment that are in agriculture	151	0.0117	0.0119	0.0016	0.0668
Proportion of those aged 16-74 that are in professional occupations	151	0.2672	0.0688	0.1470	0.4958
Index of Multiple Deprivation 2007	151	23.8098	9.1168	8.0857	48.2627
Need index (incorporates CARAN formula)	151	1.0253	0.1334	0.7311	1.3479
MFF index for HCHS and prescribing	151	1.0021	0.0559	0.9410	1.1243
Diabetes prevalence rate 2007/8 (%, over 17 years)	151	5.4872	0.7982	3.22	8.51
Epilepsy prevalence rate 2007/8 (%, over 18 years)	151	0.7884	0.1489	0.41	1.09
HIV need index	151	1.1848	1.4984	0.1648	8.3332
Chronic kidney disease 2007/8 (%, over 18 years)	151	4.1687	1.2711	1.35	8.41
Maternity need index	151	1.0345	0.2106	0.6845	1.8129
Raw (unadjusted) population 2007/8	151	335,735	196,501	90,142	1,264,298

Note: these statistics are unweighted across PCTs and reflect the values for these variables as available for the regression analysis of PB expenditure data for 2007/8 and for 2008/9.
Sources: Population Census 2001, Department of Health (2009), NHS Information Centre website.

The available instruments reflect factors, such as socio-economic deprivation and the availability of informal care in the community, which might indirectly impact upon mortality rates and/or health care expenditure levels. The set of instruments associated with each estimated equation was selected on both technical and pragmatic grounds. From a pragmatic point of view, we require a parsimonious set of instruments that satisfy the necessary technical criteria. These are, firstly, that they have face validity, that is,

that they are plausible determinants of the endogenous variable being instrumented, and secondly, that the instruments are both relevant and valid. The relevance of an instrument set refers to its ability to predict the endogenous variable of concern, whereas validity refers to the requirement that instruments should be uncorrelated with the error term in the equation of interest.

Should the instrument set be strong, relevant and valid, 2SLS will produce consistent estimates of the parameters of the reduced form models. We subject the instrument sets to tests for validity using the Sargan-Hansen test of over identifying restrictions.[81] The joint null hypothesis is that the instruments are valid instruments, i.e., they are uncorrelated with the error term, and that the instruments are correctly excluded from the outcome equation of interest. A rejection of the null hypothesis casts doubt on the validity of the instruments. We test for instrument relevance using Shea's [82] partial R-squared measure; this reflects the correlation between the excluded instruments and the endogenous regressor. However, even where valid and relevant, a non-zero but small correlation between the set of instruments and the endogenous regressors can lead to the problem of weak instruments, again rendering IV estimation biased. We test for the presence of weak instruments using the procedures set out in Stock and Yogo[83] and the Kleibergen-Paap LM statistic. A general test of model specification is provided through the use of Ramsey's[84] reset test for OLS and an adapted version of the test for instrumental variables[85].

Finally, we check that the presumed endogenous variable is in fact endogenous using the test proposed by Durbin.[86] If the null hypothesis of exogeneity cannot be rejected, then we revert to using OLS. While, in general, our instruments 'pass' the appropriate statistical tests, some commentators claim that such tests may have 'low power' to and hence may fail to reject the validity of the instruments when this is false in small samples. Consequently in Section B9 of Appendix B we examine how sensitive our results are to the relaxation of the assumption that the instruments are valid.

Further details of our approach to IV estimation are set out in Appendix B.

3.6 Results

The work presented here builds on previous studies of the link between expenditure and health outcomes. Martin, Rice and Smith[63] reported outcome elasticities for two programmes (cancer and circulatory disease) using expenditure data for 2004/5 and pooled mortality data for 2002, 2003, and 2004.[17] This work was extended in a subsequent study[66] to include several other programme and updated expenditure data (2005/6). However, the authors struggled to obtain sensible outcome models for some programmes of care. Attempts to improve model estimates by considering alternating measures of the population need for health care[18] and an extended set of potential instrumental variables are presented in Section B7 of Appendix B. This work forms the basis for the set of key results from the empirical modelling of health care expenditures and outcomes using more contemporaneous data presented in the following sections. Details of all results presented are set out in Appendix B.

3.6.1 2006/7 expenditure data and mortality data for 2006/2008

This section presents results that relate expenditure in 2006 to mortality in the same year and in the two following years (i.e., in 2006, 2007 and 2008). Throughout our measure of the need for health care is derived from the Department of Health's resource allocation model based on the CARAN needs

[17] Note that the mortality data precedes expenditure in these models. This was due to data limitations at the time of the study.

[18] Initial modelling work employed the Department of Health's resource allocation model of the need for health care based on the AREA report (Department of Health, 2005c). Subsequent refinements and updates to this model employed the implementation of the CARAN model (Department of Health, 2009) and the initial findings of a Person Based Resource Allocation study (Dixon et al, 2011). The use of these alternative models for the need for health care was explored.

formula.[87] This represents a more up-to-date needs adjustment than the AREA based model[88] that has been applied in previous studies[63, 66] and is directly applicable to the 152 PCTs in existence in the 2006/7 expenditure year. Expenditure data has been adjusted for differences in input prices using the market forces factors (MFFs for HCHS and prescribing).[19] The outcome and expenditure results for the big four programmes are shown in Table 3.4 with the relevant outcome and expenditure elasticities highlighted.

In all four outcome models expenditure has a significant negative effect on mortality and the all service measure of need has a significant positive effect. The squared value of the measure of need is also positive and significant in the cancer outcome equation. In the respiratory outcome model, there is an additional indicator of need – the proportion of the population that are permanently sick – and this is both positive and statistically significant. The diagnostic statistics suggest that, in all four cases, own programme expenditure is endogenous and that the instruments are valid. They also suggest that the instruments are relevant. There is no evidence that the instruments are weak in three of the four outcome results. The Pesaran-Taylor test suggests that there is no evidence of model mis-specification.

However, the Kleibergen-Paap F statistic for the respiratory disease outcome model is 7.022 and this is less than the 'critical' target of 10.0. This indicates that the instruments may be weak and not good predictors of the programme expenditure. However, if we re-estimate this model having dropped the least significant instrument, the coefficient on own programme expenditure becomes -2.622 and is significant at the 1% level. Moreover, there is now no evidence of weak instruments (the Kleibergen-Paap F statistic is 11.025) and it is this coefficient that we use for the respiratory outcome model in the cost of a life year calculations below.

In three of the four expenditure models both the need and budget variables have a positive and significant effect on own programme expenditure. In addition, the proxy for need in other programmes is negative and significant in all four cases. The diagnostic statistics suggest that, for all four expenditure models, expenditure is endogenous and the instruments are valid. They also suggest that the instruments are relevant and there is no evidence that the instruments are weak. The Pesaran-Taylor test suggests that there is no evidence of model misspecification.

[19] An exception to this is expenditure on GMS/PMS (PBC23a) which is adjusted using the GMS/PMS market forces factor.

Table 3.4: Outcome and expenditure models for the big four programmes using spend data for 2006/7 (two MFFs) and mortality data for 2006/7/8

	(1) PBC 2 cancer outcome model	(2) PBC 2 cancer spend model	(3) PBC 10 circulation outcome model	(4) PBC 10 circulation spend model	(5) PBC 11 respiratory outcome model	(6) PBC 11 respiratory spend model	(7) PBC 13 gastro outcome model	(8) PBC 13 gastro spend model
own programme spend per head	**-0.342*** **		**-1.434*** **		**-2.029*** **		**-1.536*** **	
	[0.099]		**[0.218]**		**[0.636]**		**[0.468]**	
need CARAN per head	0.995***	1.626***	2.860***	2.306***	2.696***	1.449***	4.160***	2.040***
	[0.106]	[0.343]	[0.252]	[0.372]	[1.044]	[0.331]	[0.577]	[0.378]
need CARAN per head squared	1.163***				2.451			
	[0.348]				[1.561]			
SYLLR all deaths exclude cancer		-0.855***						
		[0.191]						
PCT budget per head		**0.465**		**0.540***		**0.679*** **		**0.446***
		[0.300]		**[0.299]**		**[0.251]**		**[0.263]**
SYLLR all deaths exc circulatory				-1.666***				
				[0.295]				
permanently sick					0.759**			
					[0.367]			
SYLLR all deaths exc respiratory						-0.672**		
						[0.305]		
SYLLR all deaths exclude gastro								-1.206***
								[0.314]
lone pensioner households								
Constant	6.501***	5.913***	11.413***	10.696***	13.756***	3.346	9.719***	8.370***
	[0.436]	[2.815]	[1.046]	[2.379]	[3.279]	[2.075]	[2.009]	[2.299]
Endogeneity test statistic	13.695	19.421	42.548	24.461	17.687	8.439	16.373	15.211
Endogeneity p-value	0.000215	1.05e-05	6.90e-11	7.58e-07	2.60e-05	0.00367	5.20e-05	9.61e-05
Hansen-Sargan test statistic	0.685	0.021	0.949	1.262	1.462	0.302	2.761	0.0164
Hansen-Sargan p-value	0.408	0.084	0.814	0.261	0.227	0.583	0.0966	0.0898
Shea's partial R-squared	0.164	0.445	0.300	0.296	0.0785	0.327	0.140	0.356
Kleibergen-Paap LM test statistic	17.85	41.88	32.37	32.02	10.02	34.98	14.86	35.72
Kleibergen-Paap p-value	0.000133	8.04e-10	1.61e-06	1.11e-07	0.00666	2.54e-08	0.000592	1.75e-08
Kleibergen-Paap F statistic	13.28	56.69	17.14	31.84	7.022	20.94	11.63	22.40
Pesaran-Taylor reset statistic	0.00537	0.18	0.136	0.00349	0.0120	1.497	1.669	0.007
Pesaran-Taylor p-value	0.942	0.668	0.712	0.953	0.913	0.221	0.196	0.935

Note: robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1.

3.6.1.1 Cost of a life year

The outcome and expenditure elasticities presented in Table 3.4 can be used to calculate the cost of a life year in each programme. These calculations -- for both the big four programmes as well as for the other six programmes with mortality based outcome indicator -- are shown in Table 3.5. The cost of a life (year) estimates presented in Table 3.5 assume a 1% increase in each PCT's budget and are calculated as:

the cost of an additional life in a particular programme
= the change in expenditure in that programme / the change in mortality in that programme
= (annual spend * expenditure elasticity) / (annual mortality * outcome elasticity * expenditure elasticity)

and

the cost of an additional life year in a particular programme
= the change in expenditure in that programme / the change in life years lost in that programme
= (annual spend * expenditure elasticity) / (annual life years lost * outcome elasticity * expenditure elasticity)

To illustrate this calculation let us calculate the cost of a life year for, say, the cancer programme. The annual spend on cancer in 2006/7 is £4,122m and the expenditure elasticity for the programme is 0.465 so that the change in expenditure associated with a 1% increase in each PCT's budget is £19.1673m (=1%*£4,122m*0.465). The total number of life years lost to cancer for 2006, 2007, and 2008 totals 2,207,021 life years and so the average annual loss is 735,674 life years. The outcome elasticity for the cancer programme is 0.342 and the expenditure elasticity is 0.465 so the reduction in the number of life years lost associated with a 1% increase in each PCT's budget is 1,170 (=1%*735,674 life years*0.342*0.465). The cost of an additional life year is therefore £19.1673m (the change in expenditure in the programme) divided by 1,170 (the reduction in the number of life years lost), and this equals £16,383.

An integral part of the calculation of the cost of a life year is the annual mortality (life years lost) figure associated with a particular programme. Ideally, the ICD10 coverage of the expenditure data should coincide with that of the mortality data. However, as shown in Table B5.1 of Appendix B, the ICD10 coverage of the mortality data typically falls short of that for the expenditure data. Unless we adjust the annual mortality figure so that its ICD10 coverage approximates that of the expenditure data, our cost of life (year) estimates will be too large because they will underestimate the mortality gain.

Table 3.5 reports cost of a life year estimates both with and without this adjustment for ICD 10 coverage. Having incorporated this adjustment, the results show that the cost of a life year for the big four PBCs is estimated as £10,604 and, for all ten programmes with a mortality outcome measure, the estimate is £19,965. For all programmes, assuming a zero gain for the 13 PBCs without an outcome indicator, the corresponding estimate is £73,457.

If we assume that PBC23 (largely primary care) generates a zero health gain (because the gains from primary care are already reflected in the mortality rates for disease specific programmes) and that the gain attributable to the remaining 12 programmes is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes is £22,565.[20]

[20] Refer to Appendix B, Table B8.23.

3.6.1.2 Non-PCT Department of Health funded expenditure

PCT expenditure accounts for a large proportion of Department of Health expenditure but PCTs do not account for the Department's entire budget. In 2006/7 the Department of Health's gross expenditure totalled £83.5bn. Charges raised £3.4bn so net expenditure totalled £80.1bn. Of this net expenditure, PCTs accounted for £67.3bn (that is, 84%) and various other bodies accounted for the remaining £12.8bn. A breakdown of this gross and net expenditure by major body is shown in Table B8.24 of Appendix B. The Department of Health has allocated net non-PCT expenditure across the 23 PBCs. Of the additional £12bn of net expenditure, £11.2bn (93%) has been allocated to PBC23. This largely reflects: (a) the allocation of almost all Strategic Health Authority expenditure to either PBC23B ('other: SHAs including workforce development committees') or PBC23X ('other: miscellaneous'), and (b) the allocation of almost two-thirds of Department of Health expenditure to PBC23X ('other: miscellaneous'). The remaining £0.8bn of additional net expenditure is spread across all PBCs according to various allocation rules and although this approach avoids allocating expenditure to the 'Other: Miscellaneous' category, this allocation of expenditure does not necessarily reflect actual expenditure.

The cost of a life (year) estimates presented above are based on the impact of a 1% exogenous change in total net PCT spend. All of our outcome and expenditure models have been estimated using net PCT expenditure, and all of our elasticities relate to this expenditure. Implicitly we assume that any budgetary shock only affects PCT funding and that it leaves non-PCT funding unchanged. Suppose instead we assume a 1% exogenous change in the Departmental budget. We have no information on how this Departmental budgetary shock is likely to be split between PCT and non-PCTs budgets. One might assume that the non-PCT budget is as responsive to a Departmental budgetary shock as is the PCT budget. If this was the case then it would add 17.7% to our cost of a life year estimate for 2006/7. However, in the absence of any information about the responsiveness of the non-PCT budget, it is difficult to come to any firm conclusion about the impact of non-PCT expenditure on our cost of a life year estimates.

3.6.2 2007/8 expenditure data and mortality data for 2007/2009

Outcome and expenditure models were estimated using updated data for expenditure (from 2006/7 to 2007/8) and updated mortality data (from 2006/2007/2008 to 2007/2008 /2009). Appendix B, Section B10 presents detailed discussion of the findings including tables of results.

3.6.2.1 Outcome models

As before we model outcome as a function of own programme expenditure and a measure of health care need, where the latter is proxied by the measure of need as employed by the Department of Health for resource allocation purposes.[21] There are, however, a few exceptions. For the respiratory programme we further included the square of the measure of need to improve model fit. In some other PBCs we found that the all service measure of need performed poorly and we replaced or supplemented this measure with either a more programme specific measure (e.g., the epilepsy prevalence rate for neurological mortality) or with a better performing proxy for need (e.g., the percentage of residents born outside the EU for maternity/neonate mortality). These amendments improved model specification[22]. Full results for all programmes are presented in Table B10.1 Appendix B; below is a summary of the findings.

Two sets of models were estimated for three (for cancer, respiratory problems and gastro-intestinal problems) of the big four programmes. One of the two models used two instruments and so we report

[21] Using the CARAN model (Department of Health (2009).

[22] In addition to respiratory and neurological programmes the other programmes where the all service measure of need was replaced are: endocrine: IMD07 and diabetes prevalence rate; genitor-urinary: lone parent households; infectious diseases: IMD07 and HIV need per head and its square; maternity and neonates: proportion born outside EU and proportion of population with no qualification aged 16 to 74. For trauma and injuries, the all service measure of need was supplemented with the proportion of households without a car and proportion of full time students.

the instrument validity test statistic. In all three cases we failed to reject the null hypothesis of instrument validity. However, there is some evidence of weak instruments (at least in the respiratory and gastro-intestinal programmes) and if we dropped one instrument and re-estimated the model, evidence of instrument weakness disappeared. The removal of one instrument has little impact on the coefficient on expenditure and it is this coefficient that we use below in our cost of a life year calculations reported in Table 3.5.

For the big four programmes the need variable has a positive and significant effect on mortality, and expenditure has the anticipated negative effect. The diagnostic statistics reveal that, in all four PBCs, own programme expenditure is endogenous and that the instruments are valid. They also suggest that the instruments are relevant and there is no evidence that they are weak in the models with one excluded instrument. The Pesaran-Taylor test reveals no evidence of model miss-specification.

The outcome results for the other programmes are similar to but more diverse than those for the big four programmes. This is to be anticipated because mortality is a much rarer outcome in these programmes than it is in the big four programmes. Own programme expenditure is not endogenous in four of these programmes but we retain the IV estimator for three of these four because this yields more plausible results than the OLS estimator (the results are more plausible in the sense that the signs on the coefficients are more in line with our prior expectations)[23].

Expenditure has the anticipated negative effect on mortality in the endocrine problems programme but this is not statistically significant. The all service measure of need is not relevant for this PBC; instead, we find that the diabetes prevalence rate is positively associated with mortality, as is a measure of deprivation (the IMD2007).
Mortality from epilepsy is negatively and significantly associated with expenditure in the neurological programme. Both the all service need for health care and the epilepsy prevalence rate are positively and significantly associated with mortality in this programme.

Expenditure has a negative and statistically significant effect on mortality (from renal problems) in the genitor-urinary problems programme. The prevalence of lone parent households is positively associated with mortality.

Expenditure has the anticipated negative effect on mortality in the infectious disease programme and this is statistically significant. The all service measure of need is not relevant for this PBC; instead, we find that a measure of need associated with HIV is positively associated with mortality, as is a measure of deprivation (the IMD2007).

Expenditure has the anticipated negative effect on mortality in the maternity & neonates programme but the estimated coefficient is not statistically significant. In this PBC the generic all service measure of need has been replaced with two other indicators of deprivation – the proportion of residents born outside the EU and the proportion of those aged 16-74 without any qualifications – both of these are positively associated with mortality.

Finally, expenditure and need have the anticipated effects on mortality in the trauma and injuries programme. In addition, the proportion of households without access to a car is negatively associated with mortality from fractures (perhaps access to a car facilitates involvement in serious road traffic accidents), and the proportion of residents that are students is positively associated with mortality from fractures.

The relevant statistical test suggests that expenditure is endogenous in six of the ten programmes but we have retained the IV estimates for three of the other four programmes because they provide plausible results. The Hansen-Sargen test suggests that the selected instruments are valid, and the Kleibergen-Paap LM statistic suggests that they are relevant (i.e., correlated with the endogenous regressor). With the

[23] The four programmes are: endocrine, infectious diseases, maternity/neonates and trauma/injuries.

possible exception of the trauma and injuries programme, the Kleibergen-Paap F statistic suggests that we do not have a problem with weak instruments.[24] Finally, the Pesaran-Taylor/Ramsey reset test statistics reveal no evidence of misspecification.

3.6.2.2 Expenditure models

The majority of the expenditure models contain the three variables: the PCT budget, a proxy for the own programme need for health care, and a proxy for the need for health care in other programmes. The budget term is positive in all eleven models and it is statistically significant in eight of these eleven models.

The usual proxy for the own programme need for health care (i.e., the all service measure of need) is present in six of the models and it is significant in five of them. Its presence is supplemented with the addition of its squared value to improve model fit in the respiratory problems programme. In some programmes (e.g., the endocrine, metabolic & nutritional programme and the neurological programme)[25], we have replaced and/or supplemented the all service measure of need with a more programme specific measure (e.g., the diabetes prevalence rate and the epilepsy prevalence rate) and these measures of need have the anticipated positive impact on expenditure.

In addition, in a couple of other programmes we have used alternative proxies for the own programme need (e.g., with the use of the Department of Health's measure of maternity need in the maternity/neonates expenditure equation). Full results for all programmes are presented in Table B10.2 Appendix B; below is a summary of the findings.

For eight of the eleven programmes we have used the all cause mortality rate less own programme mortality rate as the proxy for the need for health care in other programmes, and the coefficient on this term is negative in seven programmes and statistically significant in six of the seven. In three programmes -- maternity/neonates, GMS/PMS and trauma & injuries programmes -- we have used the all cause mortality rate as the proxy for the need for health care in other programmes due to difficulties associated with the measurement of the own programme mortality rate. The coefficient on this term is not significant in any of the three models.

The relevant statistical test suggests that expenditure is endogenous in six of the eleven programmes but we have retained the IV estimates for two other programmes (GMS/PMS and trauma & injuries) because the IV estimator provides more plausible results. In the other three programmes we report OLS results.

The Hansen-Sargen test suggests that the selected instruments are valid, and the Kleibergen-Paap LM statistic suggests that they are relevant (i.e., correlated with the endogenous regressor). The Kleibergen-Paap F statistic suggests that we do not have a problem with weak instruments. Finally, the Pesaran-Taylor reset test statistics and the Ramsey reset F statistics reveal no evidence of model misspecification.

3.6.2.3 Calculation of the cost of a life and life year

Expenditure and outcome elasticities for preferred models are used to calculate the cost of a life year, both for individual programmes and for all programmes collectively. The relevant figures are summarised in Table 3.5.[26] The cost per life year gained is £13,830 for the big four programmes and £28,983 for all ten programmes with a mortality-based outcome indicator. These represent 30% and 45% increases on the respective costs for the previous year (i.e., using expenditure data for 2006/7 and mortality data for 2006/2007/2008).

---

[24] The Kleibergen-Paap F statistic is very close to the target value of ten for both the genitor-urinary and infectious diseases outcome models.

[25] These are endocrine: all service measure of need and diabetes prevalence rate; neurological: epilepsy prevalence; GMS/PMS: proportion of lone pensioner households; trauma/injuries: proportion of population working in agriculture.

[26] Full details of these calculations can be found in Tables B10.3 and B10.4 of Appendix B.

If we assume that the other 13 programmes (all without a mortality based outcome indicator) offer no health gain, then the cost per life year across all PCT expenditure is £82,765. This is up from £73,457 using data for the previous year (an increase of 13%).

Table 3.5: Cost of life year estimates by PBC for PCT expenditure in 2006/7, 2007/8 and 2008/9

	Expenditure 2006/07 Outcome 2006/08				Expenditure 2007/08 Outcome 2007/09				Expenditure 2008/09 Outcome 2008/10			
PBC description	Spend (£m) 2006/7	Total life years lost, <75years, 2006/08	Cost per life year gained (£)	Cost per life year gained adj for YLL coverage (£)	Spend (£m) 2007/8	Total life years lost, <75years, 2007/09	Cost per life year gained (£)	Cost per life year gained adj for YLL coverage (£)	Spend (£m) 2008/9	Total life years lost, <75years, 2008/10	Cost per life year gained (£)	Cost per life year gained adj for YLL coverage (£)
Cancer	£4,122	2,207,021	£16,383	£16,121	£4,573	2,189,685	£17,165	£16,891	£4,843	2,170,660	£21,802	£21,454
Circulatory problems	£6,161	1,361,634	£9,466	£9,390	£6,325	1,313,223	£11,315	£11,224	£6,655	1,285,026	£11,779	£11,685
Respiratory problems	£3,285	324,223	£11,593	£8,961	£3,431	315,457	£14,798	£11,439	£3,994	311,034	£21,307	£16,470
Gastro-intestinal problems	£3,700	345,908	£20,892	£11,929	£3,805	343,355	£25,034	£14,295	£3,989	341,884	£25,662	£14,653
Big four programmes summary:	£17,268	4,238,786	£12,333	£10,604	£18,134	4,161,720	£16,345	£13,830	£19,481	4,108,604	£16,688	£14,650
Infectious diseases	£1,053	106,552	£630,798	£630,798	£1,119	106,092	£57,742	£57,742	£1,201	100,078	£71,432	£71,432
Endocrine problems	£1,852	57,672	£114,416	£72,539	£1,997	55,492	£190,745	£120,932	£2,222	54,779	£104,008	£65,941
Neurological problems	£2,790	66,137	£1,129,960	£153,675	£3,165	64,873	£431,749	£58,718	£3,466	64,222	£388,267	£52,804
Genito-urinary problems	£3,482	10,030	£20,421,090	£3,512,427	£3,439	8,529	£652,096	£112,160	£3,779	8,004	£877,038	£150,851
Trauma & injuries*	£2,892	30,000	n/a	n/a	£2,918	21,273	£1,115,197	£195,159	£3,255	6,881	#DIV/0!	#DIV/0!
Maternity & neonates*	£3,574	492,600	£45,158	£30,662	£3,662	489,170	£204,168	£138,630	£3,978	479,905	£198,939	£135,080
Other six programmes summary:	£15,643	762,991	£258,046	£146,108	£16,300	745,429	£274,309	£99,428	£17,901	£713,869	£254,794	£112,674
All ten programmes summary:	£32,911	5,001,777	£23,780	£19,965	£34,434	4,907,149	£38,110	£28,983	£37,382	4,822,473	£38,328	£30,883
Other 13 programmes summary:	£34,985				£39,223				£41,016			
All 23 programmes	£67,896		£87,494	£73,457	£73,657		£108,829	£82,765	£78,398		£105,460	£84,974

In addition, if we assume that PBC23 generates a zero health gain and that the gain attributable to the remaining 12 programmes is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes is £31,846 (it was £22,565 using data for the previous year).

The next section presents outcome and expenditure models using PB data for 2008/9 and mortality data for 2008/9/10, and it explores the reasons for the increase in the cost of an additional life year identified in this section.

3.6.3 2008/9 expenditure data and mortality data for 2008/2010

Outcome and expenditure models were estimated using updated data for expenditure (from 2007/8 to 2008/9) and updated mortality data (from 2007/2008/2009 to 2008/2009/2010). Detailed results for the outcome model and expenditure model are shown in Tables B11.1 and B11.2, Appendix B respectively. First stage regressions for these IV models can be found in Tables BA.9 and BA.10 in the annex to Appendix B.

3.6.3.1 Outcome models

The majority of the outcome models contain the two variables: own programme expenditure and a measure of the need for health care (the measure of need as employed by the Department of Health for resource allocation purposes[27]). For the respiratory disease programme we have added the square of the need measure to improve the model fit. In other PBCs (e.g., for the endocrine, metabolic and nutritional programme), we found that the all service measure of need performed poorly and we have replaced it with a more programme specific measure (e.g., the diabetes prevalence rate) or with a better performing proxy for need (e.g., the percentage of residents born outside the EU for maternity/neonate mortality).[28]

The relevant statistical test suggests that expenditure is endogenous in six of the ten programmes but we have retained the IV estimates for the other four because they provide plausible results. The Hansen-Sargen test suggests that the selected instruments are valid, and the Kleibergen-Paap LM statistic suggests that they are relevant (i.e., correlated with the endogenous regressor). The Kleibergen-Paap F statistic suggests that we do not have a problem with weak instruments. Finally, the Pesaran-Taylor reset test statistics reveal no evidence of misspecification.

In all of the big four programmes the need for health care variable has a positive and significant effect on mortality, and expenditure has the anticipated negative effect. As we have noted before, the outcome results for the other programmes are similar to but more diverse than those for the big four programmes. This is to be anticipated because mortality is a much rarer outcome in these programmes than it is in the big four programmes.

Expenditure has the anticipated negative effect on mortality in the endocrine problems programme and this is statistically significant. The all service measure of need is not relevant for this PBC; instead, we find that the diabetes prevalence rate is positively associated with mortality, as is a measure of deprivation (the IMD2007).

Expenditure has a negative but statistically insignificant impact on mortality from epilepsy in the neurological programme, and the all service indicator of the need for health care is positively and significantly associated with mortality in this programme.

---

[27] The CARAN measure of service need.

[28] The amendments are: respiratory diseases: all service need and all service need squared; endocrine: IMD07 and diabetes prevalence rate; genitor-urinary: lone parent households; infectious diseases: IMD07 and HIV need per head and its square; maternity and neonates: all service need and proportion born outside EU and proportion of population with no qualification aged 16 to 74.

Expenditure also has a negative but not statistically significant effect on mortality (from renal problems) in the genitor-urinary problems programme. The prevalence of lone parent households is positively associated with mortality.

Expenditure has the anticipated negative effect on mortality in the infectious disease programme and this is statistically significant. The all service measure of need is not relevant for this PBC; instead, we find that a measure of need associated with HIV is positively associated with mortality, as is a measure of deprivation (the IMD2007).

Expenditure has the anticipated negative effect on mortality in the maternity & neonates programme. In this PBC the coefficient on the generic all service measure of need is positive but not significant. It has been supplemented with two other indicators of deprivation – the proportion of residents born outside the EU and the proportion of those aged 16-74 without any qualifications – and both of these are positively associated with mortality.

Finally, we were unable to develop a plausible outcome model for the trauma and injuries programme.

3.6.3.2 Expenditure models

The majority of expenditure models contain the three variables: the PCT budget, a proxy for the own programme need for health care, and a proxy for the need for health care in other programmes.

The budget term is positive and statistically significant in ten of the eleven models.

The usual proxy for the own programme need for health care (i.e., the all service measure of need) is positive and significant in five of the eleven results. In a couple of programmes (respiratory disease and endocrine problems) we have added the squared value of need to improve the model fit and in both cases this term is positive and significant. In some programmes (e.g., the endocrine PBC and the neurological PBC), we have replaced and/or supplemented the all service measure of need with a more programme specific measure (e.g., the diabetes and the epilepsy prevalence rates) and these usually have a positive and significant impact on expenditure. In addition, in a couple of programmes we have used alternative proxies for own programme need (e.g., with the use of the Department of Health's measure of maternity need in the maternity/neonates expenditure equation and the use of HIV need in the infectious diseases programme).[29]

For eight of the eleven programmes we have used the all cause mortality rate less the own programme mortality rate as the proxy for the need for health care in other programmes, and the coefficient on this term is negative in seven programmes and statistically significant in six of the seven. In three programmes -- maternity/neonates, GMS/PMS and trauma & injuries programmes -- we have used the all cause mortality rate as the proxy for the need for health care in other programmes due to difficulties associated with the measurement of the own programme mortality rate. The coefficient on this term is negative but not significant in these three models.

The relevant statistical test suggests that expenditure is endogenous in five of the eleven programmes but we have retained the IV estimates for two further programmes (endocrine problems and maternity/neonates) because the IV estimator provides more plausible results than the OLS estimator. In the other four programmes we report OLS results.

---

[29] These are infectious diseases: HIV need and its square; endocrine: all service measure of need, its square and diabetes prevalence rate; genitor-urinary: all service measure of need and proportion of residence born outside EU; maternity/neonates: maternity measure of need; GMS/PMS: all service measure of need, proportion of residents reporting permanent sickness (16yrs – 74yrs), proportion of lone pensioner households and proportion in professional occupations; trauma/injuries: proportion of population working in agriculture.

The Hansen-Sargen test suggests that the selected instruments are valid, and the Kleibergen-Paap LM statistic suggests that they are relevant (i.e., correlated with the endogenous regressor). The Kleibergen-Paap F statistic suggests that we do not have a problem with weak instruments. Finally, the Pesaran-Taylor reset test statistics and the Ramsey reset F statistics reveal no evidence of model misspecification.

3.6.3.3 Calculation of the cost of a life and life year

Expenditure and outcome elasticities for our preferred models are used to calculate the cost of a life year, both for individual programmes and for all programmes collectively. This results in the cost per life year gained having increased slightly compared with that using the previous expenditure and mortality data set (i.e., for 2007 and 2007/8/9 respectively): increasing from £13,830 to £14,650 for the big four programmes and from £28,983 to £30,883 for all ten programmes with a mortality-based outcome indicator. If we assume that the other 13 programmes offer no health gain, then the cost per life year across all PCT expenditure has increased from £82,765 in 2007/8 to £84,974 in 2008/9.

In addition, if we assume that PBC23 generates a zero health gain and that the gain attributable to the remaining 12 programmes is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes in 2008/9 is £33,333. This is a 5% increase on the figure (£31,846) for the previous year.

3.6.4 Comparing the cost of life year estimates associated with different data sets

Table 3.6 presents expenditure and outcome elasticities for the five combinations of expenditure and outcome data that have been used to estimate our model. It also reports the corresponding unadjusted cost of life year estimates (i.e., estimates that are unadjusted for the mismatch in the ICD10 coverage of the expenditure and mortality data). It is clear from this Table (see row 13) that the (unadjusted) cost of a life year for the ten programmes with a mortality based outcome indicator fluctuated around £22,000 for the first three sets of estimations (see columns M-O). However, using the two most recent sets of expenditure data (i.e., for 2007/8 and then for 2008/9), the figures in the table suggest that this cost has increased to about £38,000.

What are the proximate causes of this increase? Recall that the cost of a life year is calculated as

$$\frac{\text{the change in expenditure associated with a 1\% budget increase}}{\text{the change in the number of life years lost associated with this increase}}$$

For 2006/7 (using mortality data for 2006/7/8) and for the ten programmes with a mortality based outcome indicator, the change in expenditure associated with a 1% budget increase is £184.53m and the change in the number of life years lost associated with this increase is 7,760 (see Table B8.21 in the appendix for the calculation of these figures). Thus the cost of a life year is £23,780 (=£184.53m/7,760).

For 2007/8 (using mortality data for 2007/8/9) and for the ten programmes with a mortality based outcome indicator, the change in expenditure associated with a 1% budget increase is £257.94m and the change in the number of life years lost associated with this increase is 6,768 (see Table B10.3 in the appendix for the calculation of these figures). Thus the cost of a life year is £38,110 (=£257.94m/6,768).

It is clear that the 60% increase in the cost of a life year between 2006/7 and 2007/8 is largely attributable (a) to the 40% increase in the additional expenditure (up from £184.53m to £257.94m) directed towards these ten programmes following a 1% budget increase and (b) to the 12% decline in the number of life years saved by this increase in expenditure (down from 7,760 to 6,768 life years).

The rise in the share of the budget increase directed towards these programmes can be attributed to the increase in the implied expenditure elasticity associated with these ten programmes (up from 0.561 to 0.749). The decrease in the number of years of life saved appears to be due (a) to an overall reduction in the (absolute) size of the outcome elasticities and (b) to a shift in the additional expenditure towards those

programmes with a relatively high cost of a life year. For example, the cost of a life year for the 'small six' programmes is much larger than for the 'big four' programmes. However, in 2007/8 the spend elasticity for the small six increases from 0.561 to 0.961 (71%) while the expenditure elasticity for the big four rises from 0.528 to 0.559 (6%). A similar pattern – of additional expenditure shifting away from the low cost PBCs – can be seen within the big four programmes. However, it is not clear why such rather dramatic changes should have taken place.

If we correct the cost of life year estimates adjusting for the mismatch in the ICD10 coverage of the expenditure and mortality data, these reveal similar increases in the cost of a life year between 2006/7 on the one hand and 2007/8 and 2008/9 on the other. The cost of a life year increased from £19,965 in 2006/7 to £28,983 in 2007/8 for the ten programmes with mortality rate, an increase of 45%; and it increased from £22,565 to £31,846 for all programmes if we assume a zero health gain in PBC23 and the same gain in the other 12 programmes as in the ten with a mortality rate (an increase of 41%).

A potential reason for this apparent step change in the cost of a life year is the adjustment that was made to the methodology for the collection of the 2007/8 programme budgeting data. In previous years expenditure that was not directly attributable to a particular programme category was apportioned using admitted patient care percentages.[30] In other words, if x% of total admitted patient care expenditure was allocated to PBC 1, then x% of all expenditure that was not directly attributable to a particular programme category was also allocated to PBC 1. With effect from 2007/8, however, NHS organisations were asked to select an appropriate basis for the apportionment of this non-programme specific expenditure and that, where no reasonable basis existed, such expenditure was to be allocated to the 'Other – Miscellaneous' (PBC 23X) category.

The Department of Health estimates that this allocation rule change increased the amount of expenditure attributed to PBC 23X by £700 million. It will also, of course, have reduced expenditure across other programmes by the same amount in total. However, not all programmes will have been equally affected; PBCs that are more heavily inpatient based would have 'lost' expenditure while others, such as learning disabilities, social care, and mental health, will have 'lost' considerably less. In addition, not all PCTs will have been equally affected because each will have employed different apportionment rules for the non-programme specific expenditure.[89]

Although this allocation rule change has considerably increased the estimated cost of a life year, we believe that this rule change has led to a more accurate allocation of expenditure across PBCs, and that the more recent estimates of the cost of a life year (for 2007/8 and 2008/9) are more accurate than those for the earlier years (for 2005/6 and 2006/7).

3.6.5 Adjusting the cost of a life year estimates to constant prices

The estimates of the cost of a life year presented above are all at current prices. To put them on a constant price basis, we need an index of pay and price inflation for the labour and goods/services purchased by the NHS. Curtis[90] reports a pay and prices index for Hospital and Community Health Services and this implies an inflation rate of 3.7% in 2006/7, 2.9% in 2007/8, and 3.9% in 2008/9.[31] If we assume that similar inflation rates also apply to the purchase of pharmaceuticals and the provision of primary care (items that are excluded from the HCHS index), then we can use these figures to put the estimates of the cost of a life year on a constant price basis.

---

[30]Expenditure on, for example, community care, A&E, ambulance services, and outpatients can be difficult to attribute to a particular PBC. Critical care, rehabilitation, and specialised commissioning across care settings will also be difficult to attribute to a particular programme.

[31]With the index for 1987/8 set equal to 100, then 2005/6=240.9, 2006/7=249.8, 2007/8=257.0, and 2008/9=267.0 (Curtis, 2011, p209).

For example, if we assume that PBC23 generates a zero health gain and that the gain attributable to the 12 programmes without a mortality indicator is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes in 2008/9 is £33,333 at current (2008/9) prices. The cost for 2007/8 is £31,846 at current (2007/8) prices or £33,088 at constant (2008/9) prices, and the figure for 2006/7 is £22,565 at current (2006/7) prices or £24,125 at constant (2008/9) prices. The conversion of the costs from a current to constant price basis has relatively little impact because the inflation rate over the relevant period is quite small.

Table 3.6: Expenditure and outcome elasticities for five combinations of expenditure and outcome data, and corresponding (unadjusted) cost of life year estimates

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
		spend elasticities					outcome elasticities					cost of an additional life year (unadjusted for YLL coverage)				
	PBC description	(a) using spend for 2005 and mortality for 2002/4	(b) using spend for 2006 and mortality for 2004/6	(c) using spend for 2006 and mortality for 2006/8	(d) using spend for 2007 and mortality for 2007/9	(e) using spend for 2008 and mortality for 2008/10	(a) using spend for 2005 and mortality for 2002/4	(b) using spend for 2006 and mortality for 2004/6	(c) using spend for 2006 and mortality for 2006/8	(d) using spend for 2007 and mortality for 2007/9	(e) using spend for 2008 and mortality for 2008/10	(a) using spend for 2005 and mortality for 2002/4	(b) using spend for 2006 and mortality for 2004/6	(c) using spend for 2006 and mortality for 2006/8	(d) using spend for 2007 and mortality for 2007/9	(e) using spend for 2008 and mortality for 2008/10
1	Cancer	0.968	0.548	0.465	0.890	0.525	-0.394	-0.337	-0.342	-0.365	-0.307	£13,741	£16,518	£16,383	£17,165	£21,802
2	Circulatory problems	0.682	0.701	0.540	0.293	0.648	-1.370	-1.447	-1.434	-1.277	-1.319	£8,328	£8,725	£9,466	£11,315	£11,779
3	Respiratory problems	0.849	0.718	0.679	0.536	0.652	-1.574	-3.507	-2.622	-2.205	-1.808	£20,601	£8,747	£11,593	£14,798	£21,307
4	Gastro-intestinal problems	0.772	0.667	0.446	0.622	0.456	-2.018	-2.137	-1.536	-1.328	-1.364	£18,303	£15,795	£20,892	£25,034	£25,662
5	All big four PBCs	0.801	0.660	0.528	0.559	0.579	-0.941	-1.083	-0.965	-0.872	-0.825	£12,855	£10,783	£12,333	£16,345	£16,688
6	Infectious diseases	0.742	0.731	0.792	1.436	1.545	-0.152	-0.030	-0.047	-0.548	-0.504	£215,054	£1,036,377	£630,798	£57,742	£71,432
7	Endocrine problems	0.425	0.966	0.953	0.264	0.484	-0.244	-0.812	-0.842	-0.566	-1.170	£371,601	£112,882	£114,416	£190,745	£104,008
8	Neurological problems	1.111	0.648	0.616	1.035	0.98	-0.182	-0.098	-0.112	-0.339	-0.417	£503,201	£1,241,253	£1,129,960	£431,749	£388,267
9	Genito-urinary problems	1.041	0.837	0.912	1.004	0.697	-0.034	-0.073	-0.051	-1.855	-1.615	£29,144,918	£12,384,965	£20,421,090	£652,096	£877,038
10	Trauma & injuries*	0.627	0.617	0.358	1.686	1.344	-1.332	-0.527	0	-0.369	0	£282,132	£548,767	n/a	£1,115,197	n/a
11	Maternity & neonates*	0.388	0.601	0.224	0.514	0.975	-0.237	-0.035	-0.482	-0.110	-0.125	£17,490	£631,700	£45,158	£204,168	£198,939
12	All small six PBCs	0.780	0.717	0.596	0.961	0.962	-0.262	-0.122	-0.392	-0.254	-0.300	£295,074	£449,706	£258,046	£274,309	£254,794
13	All 10 PBCs with mortality	0.792	0.687	0.561	0.749	0.762	-0.844	-0.940	-0.877	-0.778	-0.747	£21,256	£20,893	£23,780	£38,110	£38,328
14	All 23 PBCs assuming zero gain in PBCs without mortality indicator											£56,799	£62,718	£87,494	£108,829	£105,460
15	GMS/PMS	0.926	0.759	0.739	0.563	0.494	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
16	All 23 PBCs assuming zero gain in PBC 23 but average gain in other PBCs without a mortality indicator											£24,200	£23,697	£26,876	£41,875	£41,369

Notes:

(i) that the spend and outcome elasticities reported for groups of programmes are the implied elasticites calculated from the totals for the relevant individual programmes (i.e., group spend elasticity=∑(PBC spend*PBC spend elasticity)/∑ PBC spend, and group outcome elasticity=∑(PBC mortality*PBC outcome elasticity)/∑ PBC mortality). For the purpose of the calculation of the group outcome elasticity, we have used the years of life lost as the mortality indicator. The implied group elasticities cannot be used to calculate directly the cost of a life (year) for a group of PBCs. Instead, the latter should be calculated by summing across the change in spend and the change in mortality for the individual PBCs within the group. For further details see, for example, Table B8.21 in appendix B.

(ii) for each individual programme: the cost of an additional life year = expenditure elasticity*annual spend/(expenditure elasticity *outcome elasticity*annual life years lost)

(iii) for a group of programmes: the overall cost of an additional life year = ∑ (annual spend*spend elasticity) / ∑ (spend elasticity*outcome elasticity*annual life years lost)

3.7 Summary and concluding remarks

The findings presented in this report build on four previous studies. These studies and the results presented here draw on the availability of two new data sets to obtain empirical estimates of the relationship between mortality and expenditure across all English local health authorities.

In this research we have extended the previous studies in several ways. First, we have derived plausible outcome and expenditure models for a larger number of programmes (ten) than previous studies.

Second, we relate expenditure in time period *t* to mortality in that period (*t*) and in the next two periods (*t+1* and *t+2*). In other words, we assume that the health benefits associated with expenditure occur either in the same period as the expenditure or in the next two periods. This is an improvement on past practice where data constraints forced researchers to relate expenditure to the current and two previous periods.[32] When we re-estimated our models using expenditure data for 2006/7 and mortality data for 2006/7/8, we found that the cost of a life year across the ten programmes with a mortality based outcome indicator is £23,780 (up from £20,893 when expenditure data for 2006/7 is combined with mortality data for 2004/5/6; an increase of 14%).

Third, we have noted the mismatch in the ICD10 coverage of the expenditure and mortality data. If we adjust the calculation of the cost of a life year for 2006/7 for this mismatch then the cost of a life year across the ten programmes with a mortality based outcome indicator declines from £23,780 to £19,965 (a decrease of 16%).

Fourth, previous estimates of the cost of a life year have been for individual programmes of care. In this report we have presented estimates of the cost of a life year for an enlarged number of programmes and, with the aid of assumptions about the productivity (health gain) of programmes without a meaningful mortality-based outcome indicator, we have extended our individual programme estimates to incorporate expenditure across all programmes of care. Thus for 2006/7, the cost of a life year for those PBCs with a mortality based outcome indicator is £19,965. If we assume that (a) that the health gain associated with PBC23, which includes primary care and workforce training expenditure, are reflected in the mortality rates for disease specific programmes and (b) that the average health gain across the other programmes without a mortality based outcome indicator is the same as that for those PBCs with a mortality based outcome indicator, then the cost of life year across all programmes is £22,565.

Fifth, we have extended our cost of life year estimates beyond 2006/7. Re-estimation of our model using budgeting expenditure for 2007/8 generates an all programme cost of a life year estimate of £31,846, and re-estimation of our model using budgeting expenditure for 2008/9 generates a similar cost of a life year estimate (£33,333). Together, the last two estimates suggest that there has been step change in the cost of a life year, and that this appears to have occurred between 2006/7 and 2007/8. The cost of a life year estimates are very similar up to and including 2006/7, and they are very similar for 2007/8 and 2008/9. However, there is a substantial difference between the figures for 2004/5, 2005/6 and 2006/7 on the one hand (at about £22k), and for 2007/8 and 2008/9 on the other (at about £33k). The reason for this step change is not obvious but it might be due to changes in the algorithm used by the Department of Health to allocate non-admitted patient care activity to budget categories. Although this allocation rule change has considerably increased the estimated cost of a life year, we believe that this rule change has led to a more accurate allocation of expenditure across PBCs, and that the more recent estimates of the cost of a life year (for 2007/8 and 2008/9) are more accurate than those for the earlier years (for 2005/6 and 2006/7). A summary of the estimates of the cost of a life year adjusted for the mismatch between ICD10 chapters for expenditure and mortality are provided in Table 3.7.

Virtually all of the cost of a life year estimates presented in this report are calculated at current prices. However, it is possible to put them on a constant price basis using the Hospital and Community Health

[32] Such studies assumed that PCTs had reached some sort of equilibrium in the expenditure choices they make and the outcomes they secure.

Services pay and prices index.[90] For 2006/7, 2007/8 and 2008/9 this index recorded an annual rate of inflation of about 3.5% and so the impact of this constant price adjustment is fairly minimal. For example, if we assume that PBC23 generates a zero health gain and that the gain attributable to the 12 programmes without a mortality indicator is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes at constant 2008/9 prices is £33,333 for 2008/9, £33,088 for 2007/8, and £24,125 for 2006/7.

Finally, although previous results and our current models 'pass' the appropriate statistical tests and, in particular, the Hansen-Sargen test for valid instruments, we are aware that this test might be unable to detect the presence of invalid instruments in some circumstances and that the validity of instrumental variables is often open to question. Responding to this, several studies [91, 92] have suggested that researchers using IV techniques should subject the estimated coefficient on the endogenous variable to a sensitivity analysis. We undertake a comprehensive sensitivity analysis for the outcome equation for each of the big four models. This sensitivity analysis reveals that uncertainty associated with instrument validity has little effect on our estimate of the cost of a life year but it does increase the degree of uncertainty associated with this estimate.

We recognize that this study has a number of limitations. The estimates of the cost of an additional life year for programmes with a mortality-based outcome indicator are unadjusted for the quality of life during the additional year. Accordingly, the quoted costs will be an under-estimate of the QALY-adjusted cost of a life year to the extent that additional life years are not in perfect health. In previous studies we have noted that a rudimentary adjustment for this issue using HODaR data increased the cost of a life year by about 50% to 60%.[63, 66]

At the same time, however, the estimated costs will exaggerate the cost of an additional QALY-adjusted year for those programmes with a mortality-based outcome indicator because they ignore any health benefits that are not associated with a reduction in mortality. In other words, expenditure that improves the quality of life (e.g., cancer palliative care) but which does not extend the length of life is implicitly given a zero health gain value.

In addition, the expenditure data relates to expenditure on all patients whereas the mortality data is based on a life expectancy of 75 years. Thus implicitly our calculations attribute a zero health gain to all expenditure on those aged over 75. To illustrate the magnitude of the potential health gain ignored by this restriction, note that in a recent study of costs associated with all inpatient and outpatient activity (excluding mental health), those aged over 75 years accounted for 25% of all costs in 2007/8[93] for details of this study).

The results presented in this study are all from the estimation of the relationship between expenditure and mortality using data for a single time period. With the availability of several years of data for both expenditure and mortality, we wanted to estimate a panel data model because a panel can offer advantages over a one period model (e.g., it is better able to handle any unobserved heterogeneity across PCTs). However, most of the instruments employed here are based on the 2001 Census and thus time invariant rendering them of little use in panel data modelling.

Table 3.7: Adjusted cost of life year estimates for various combinations of programmes

A	B	C	D	E
	Programme budgeting category	Cost per life year (adjusted for ICD10 coverage of spend and mortality data)		
		2006/7	2007/8	2008/9
1	Cancer	£16,121	£16,891	£21,454
2	Circulatory disease	£9,390	£11,224	£11,685
3	Respiratory problems	£8,961	£11,439	£16,470
4	Gastro-intestinal problems	£11,929	£14,295	£14,653
5	All big four programmes	£10,604	£13,830	£14,650
6	Other six programme with a mortality rate	£146,108	£99,428	£112,674
7	All ten PBCs with a mortality rate	£19,965	£28,983	£30,883
	(a) If we assume a zero health gain in those PBCs without a mortality rate…			
8	All 23 programmes	£73,457	£82,765	£84,974
	…or (b) if we assume a zero gain in PBC23 and that the average gain from the the 10 PBCs with a mortality rate is applied to the remaining programmes			
9	All 23 programmes	£22,565	£31,846	£33,333

Note that the figures for 2006/7 relate to the use of mortality for 2006/2007/2008 combined.

Chapter 4: Translating mortality effects into life years and quality adjusted life years

4.1 Introduction

This chapter presents an overview of how the results of the econometric work undertaken to estimate the link between NHS spending and mortality, which was summarised in the previous chapter and detailed in Appendix B, can be translated in to effects on life years and quality adjusted life years (QALYs).

In this chapter we present three sequential steps of analysis which lead to estimates of the overall cost per QALY threshold for the NHS:

i. In section 4.2 we reconsider how the estimated effects on mortality from the econometrics work conducted in Chapter 3 might better translate in to life years by exploring the limitations of mortality data available at PCT level and the published years of life lost (YLL) figures presented. We explore how these estimates might be improved using additional data and analysis.
ii. In section 4.3 we consider how these estimates of life year effects might be adjusted for the quality of life in which they are lived, taking account of the gender and the age at which life years are gained or lost as well as the disutility associated with particular diseases.
iii. In section 4.4 we explore ways to also take account of those effects on health not directly associated with mortality and life year affects (i.e., the 'pure' quality of life effects) to estimate an overall cost per QALY threshold.

This sequence of analysis is set out and explained based on the analysis of 2006 expenditure and mortality data from 2006 to 2008. In section 5.2 we present estimates for 2008 expenditure and 2008 to 2010 mortality data using the same methods and discuss the uncertainties associated with these estimates. As in the previous chapter much of the detail of data and analysis that supports this overview is presented in an appendix (see Appendix C). At the end of each section we present a summary which includes a central 'best' estimate as well as extreme lower and upper bounds for the cost per life year and cost per QALY threshold.

The core assumptions which underpin these three values are common across sections 4.2 to 4.4. The central or 'best' estimate is based on two assumptions; one conservative and the other more optimistic with respect to the health effects associated with expenditure. The first is that the health effects of changes in one year of expenditure are restricted to one year. This is implicit in the estimates of outcome elasticities presented in the previous chapter.[33] This is likely to underestimate effects on mortality since expenditure that reduces mortality risk for an individual in one year may well also reduce their risk over subsequent years; possibly over the whole of their remaining disease duration. Expenditure may also prevent disease in future patient populations. Therefore, total health effects will be underestimated and the cost per life year or QALY threshold will be overestimated. Although undoubtedly conservative, it may be offset to some extent by the more optimistic assumption used to translate mortality effects into life years. Any death averted by expenditure in one year is assumed to return the individual to the mortality risk of the general population, i.e., the years of life gained associated with each death averted are based on what would have been their life expectancy taking account of their of age and gender (using life tables for the general population).

The extreme upper and lower bounds for cost per life year and cost per QALY thresholds are based on making both assumptions either optimistic (providing the lower bound for the threshold) or both

[33] Although 3 years of mortality data are used in the analysis of each year of expenditure, these are averaged to an annual value prior to estimating outcome elasticities. Therefore, the estimated outcome elasticities represent the proportionate effect on mortality in one year due to a proportionate change in expenditure.

conservative (an upper bound for the threshold). The lower bound is based on assuming that health effects are not restricted to one year but apply to the remaining disease duration for the population at risk during the expenditure year (although this still does not account for the effects of expenditure on preventing disease). The upper bound is based on the combination of assuming that health effects are restricted to one year and that any death averted is only averted for the minimum duration consistent with the mortality data used to estimate the outcome elasticities in Chapter 3 (see Section 4.2.5 for a more detailed discussion). It is very important to note that the lower and upper bounds represent extreme values rather than alternative but plausible views that could reasonably be taken. We discuss this in more detail in Section 5.8 and explain why establishing narrower bounds, which might retain some plausibility, has not been possible given the data available and therefore the analysis that has been feasible.

4.2 From mortality to life years

In this section we summarise our examination of a number of issues associated with available PCT-based mortality data and the associated published estimates of YLL. We then examine how, given the limited information available about the population at risk in each PBC, we might take proper account of the fact that some of the observed deaths would have occurred anyway (had the same population not been at risk in the particular PBC) when estimating YLL, i.e., taking account of unobserved counterfactual deaths. This allows us to estimate the YLL that better reflects the effect of expenditure on the mortality observed in each PBC, and infer the excess deaths associated with each PBC. Finally we present cost per death averted and cost per life year which accounts for the issues raised in this section.

4.2.1 Mortality and YLL coverage

The mortality data that is available at PCT level does not offer full coverage of all deaths across all the ICDs that make up each PBC (see Table B5.1 in Appendix B for how three-digit ICD-10 are mapped to PBCs). However, national (English) data is available that covers all deaths associated with all the ICDs that make up each PBC. Therefore, it is possible to adjust the incomplete reporting of mortality at PCT level (see section 3.2 in Chapter 3) before applying the estimated outcome elasticities to calculate the deaths averted due to expenditure.[34] Applying published estimates of YLL per death to all the deaths averted provides the estimate of the cost per life year reported in Chapter 3.

The published estimates of YLL (NHS IC) used in Chapter 3 only include deaths below 75 years (but exclude deaths below 1 year) and are based on the difference between age 75 and the age of each death below 75. These estimates have the same limited coverage as PCT level mortality data so are not available for all the ICDs that make up each PBC. Therefore, applying the available estimates of YLL per death to the estimated number of deaths averted requires an assumption that the YLL per death is similar for those groups of ICDs covered and not covered by the published YLL figures.

This can be examined by using national ONS data to calculate YLL in the same way as NHS IC, but with full coverage of all the ICDs that make up each PBC.[35] Although ONS data provides complete coverage and reports gender; age at death is only reported in 5 year ranges (these data are not available at PCT level so could not be used when estimating outcome elasticities in Chapter 3). Therefore, using ONS data to estimate YLL requires taking the midpoint of each range as the age of death, i.e., assuming reported

---

[34] This does assume that the proportionate effects on mortality due to changes in expenditure are similar for mortality that is and is not recorded at PCT level. This seems more reasonable than assuming no effect of expenditure on mortality that happens not to be recorded at PCT level.

[35] The estimated outcome elasticity for PBC 16 (Trauma and injuries) was zero for 2006 and could not be estimated for 2008 expenditure. Therefore, this PBC does not contribute any changes in health outcomes, although the changes in this expenditure are included in subsequent estimates of cost per life year and QALY thresholds. However, there was a very limited coverage of mortality data recorded at PCT level and the expenditure data for this PBC. In addition, the mortality data that was available (ICDs S72, S02, S06 and T90) was less likely to be associated with changes expenditure in this PBC and more likely to be associated with changes in expenditure in others. Consequently the health effects of changes in expenditure in PBC 16 may be underestimated.

deaths are equally likely over the range in which they are reported. For this reason it is not possible to precisely recover the published YLL figures using ONS data for those ICD groupings that can be precisely matched to the NHS IC coverage. However, the differences are small (see Table C2, Appendix C), suggesting that taking the midpoint of each range as the age of death may be a reasonable approximation.

The differences between estimates of YLL based on ONS and NHS IC data are, however, much more significant and are reported in Table 4.1. These reflect differences in the distribution of ages at death between those groups of ICDs covered and not covered in the NHS IC figures. For example, NHS IC figures available at PCT level for PBC7 (neurological problems) have low coverage of all deaths in this PBC (0.136 in column 1). The deaths that are reported in NHS IC are associated with epilepsy and the YLL (22,046 in column 2) reflects the generally younger age at death in this group. When adjusted for full coverage (22,046/0.136 = 162,100 in column 3) the estimated YLL is much greater than the YLL based directly on all deaths by age group reported in ONS. This difference in YLL reflects the fact that the deaths in PBC7 which are not covered by NHS IC figures tend to be in older age groups so generate fewer YLL.

Table 4.1: Estimates of YLL for NHS IC and ONS

PBC		Coverage of mortality data relative to spend data [1]	$YLL_{<75}$ (NHS IC) [2]	$YLL_{<75}$ adjusted (NHS IC) [3]	$YLL_{<75}$ no adjustment needed (ONS) [4]	Difference from adjusted NHS IC to ONS [5]
1	Infectious diseases	1.00	35,517	35,517	40,928	15%
2	Cancer	0.98	735,674	747,636	758,804	1%
4	Endocrine problems	0.63	19,224	30,322	41,548	37%
7	Neurological problems	0.14	22,046	162,100	93,755	-42%
10	Circulatory	0.99	453,878	457,538	481,246	5%
11	Respiratory	0.77	108,074	139,812	147,465	6%
13	Gastro-intestinal	0.57	115,303	201,931	177,532	-12%
17	Genito-urinary	0.17	3,343	19,438	17,380	-11%
18+19	Maternity & neonates	0.68	164,200	241,826	15,409	-94%

Using ONS data also allows deaths under the age of one year to be appropriately assigned to PBCs via the ICD in which they occurred (NHS IC YLL figures exclude deaths under one year), rather than assigning them all to PBC18 & 19 as in the previous Chapter.[36] This explains the large reduction in YLL for PBC18 & 19 (Maternity and neonates) as much of the mortality is re-assigned to ICDs which contribute to other PBCs. Since most of the deaths that are re-assigned are allocated to PBC1 (infectious diseases) the YLL for this PBC increases despite complete reporting of deaths at PCT level and full coverage by NHS IC figures (see also Table C4 in Appendix C).

4.2.2 Life expectancy and YLL

As noted above the NHS IC estimates of YLL only include deaths below 75 years and are based on the difference between age 75 and the age of each death below 75. Implicitly this treats 75 as the appropriate normal life expectancy for males and females for the population at risk in each PBC. However, with the exception of maternity and neonates most deaths in PBCs occur above the age of 75 and life expectancies are significantly greater than 75. For example, based on 2006 to 2008 data, life expectancy for the general population is 80.7 for males and 84.4 for females (considering age distribution) and even life expectancy at birth is greater than 75 (77.74 for males and 81.88 for females).[37]

---

[36] The YLL available from NHS IC represented all deaths from maternity and all deaths under 28 days across PBCs. The coverage factor (0.68 in column 1 of Table 4.1) adjusts this YLL to represent maternity and all deaths < 1year across PBCs. The calculation is described in Appendix B, footnote (v) of Table B5.1.

[37] Figures for England, from http://www.ons.gov.uk/ons/rel/subnational-health4/life-expec-at-birth-age-65/2004-06-to-2008-10/statistical-bulletin.html#tab-National-life-expectancy

Based on ONS data YLLs can be re-recalculated using gender specific life expectancy for the general population.[38] When increasing life expectancy (LE) two effects occur, both of which tend to increase estimates of YLL. Firstly, more deaths are included in the YLL calculation (those that occur between age 75 and LE) and secondly, each death previously counted below 75 will generate 5.7 or 9.4 more YLL for males and females respectively. The effect on the number of deaths and the YLL for each PBC of using the life expectancy of the general population is reported in Table 4.2 (see columns 1, 2 and 3).

Table 4.2: The difference in YLL by life expectancy

PBC		$Deaths_{<75}$ (ONS) [1]	$Deaths_{<LE}$ (ONS) [2]	Difference in deaths due to increased LE [3]	$YLL_{<75}$ (ONS) [4]	$YLL_{<LE}$ (ONS) [5]	Difference in YLL due to increased LE [6]
1	Infectious diseases	2,050	3,710	81%	40,928	62,051	52%
2	Cancer	62,944	95,212	51%	758,804	1,345,013	77%
4	Endocrine	2,367	4,000	69%	41,548	65,015	56%
7	Neurological	5,095	8,975	76%	93,755	145,526	55%
10	Circulatory	41,487	82,098	98%	481,246	916,170	90%
11	Respiratory	14,000	30,500	118%	147,465	310,326	110%
13	Gastro-intestinal	10,611	15,827	49%	177,532	273,303	54%
17	Genito-urinary	1,588	4,197	164%	17,380	39,098	125%
18+19	Maternity & neonates	226	226	0%	15,409	17,167	11%

*LE male=80.7, female=84.4

The number of deaths counted below LE increases for every PBC except for maternity & neonates because, as expected, all deaths are below age 75 in PBC18 & 19. However, YLL increases for all PBCs reflecting the additional years otherwise expected to be lived to an older LE. Of course including more of the deaths observed in each PBC and the greater YLL associated with them will generate more deaths averted and more life years gained when applying the same proportionate effects from the outcome elasticities estimated in Chapter 3. Therefore, the cost per death averted and cost per life year threshold are lower using these figures than those reported in Chapter 3 (see Table 4.6 below and Table C7 in Appendix C for a summary of the effects on the thresholds). However, there are good reasons why YLL figures calculated as the difference between age of death and LE are likely to be overestimated. This is dealt with in the next section (Section 4.2.3). In Section 4.2.4 we take account of the fact that some of the deaths observed in a PBC would have occurred anyway in a similar 'normal' population (i.e., the counterfactual population not at risk through membership of the PBC) so not all observed deaths are 'excess' and generate YLL.

4.2.3 YLL and accounting for counterfactual deaths

The estimates of YLL based on ONS data overcome many of the limitations of the published NHS IC figures. However, the YLLs reported in Tables 4.1 and 4.2, are calculated in the same way as the NHS IC figures, by taking the difference between a fixed LE and the age at death of deaths observed below that LE. This will tend to overestimate the YLL for two reasons: i) it does not account for the fact that not all deaths observed below LE are 'excess' deaths in the sense that some deaths would have occurred (at the same age) in a similar population not at risk in the PBC and ii) some of the deaths observed above LE may be 'excess' deaths that would not otherwise have occurred at that age. The overall effect on YLL, and the cost per life year, will depend on the number of deaths above and below LE that are excess. Therefore, estimates of YLL are required which take account of the 'counterfactual' deaths that would have occurred even if the population in the PBC was not at risk through membership of the ICD codes that make it up, but faced the same mortality risks as the general population, accounting for the age and gender distribution of the PBC population.

[38] This is the life expectancy that reflects the age distribution of the general population, i.e., the average of the sum of the life expectancies conditional on age, over the current age distribution. It will always be higher than life expectancy at birth.

Ideally, with reliable information about the size of the population at risk in each PBC and its age and gender distribution it would be possible to estimate the number of deaths that would be expected to occur had this population not been at risk, based on mortality data for the general population. The difference between deaths observed across all ages and the deaths expected to have occurred in this matched 'normal' population would provide the number of 'excess' deaths by age and gender.[39] The YLL associated with each of these excess deaths is the life expectancy conditional on gender and on surviving to the age at which the excess death occurred. The total YLL for the at risk population is simply the sum of these YLLs over all excess deaths, which could occur at any age. This YLL is equivalent to the area between the survival curve for the population at risk in a PBC and the counterfactual survival curve for the same population but not at risk from membership of the PBC. The difficultly is that routinely available data do not provide any information about the size of the population at risk or its age and gender distribution. All that is routinely available are observed deaths (by age and gender). Therefore, it is not possible to directly estimate excess deaths or compare survival curves.

Even if the size of the at risk population is unknown we can still use information that might be available about its age and gender distribution (or make reasonable assumptions) to estimate a matched 'normal' LE using life tables for the general population - such a LE summarises the area under the counterfactual survival curve. Unfortunately, it is not possible to also calculate the LE for the population at risk in the PBC (or represent the survival curve) without information about the size of the at risk population - if it was possible the difference between these life expectancies would approximate the YLL per patient at risk in a PBC.

Fortunately, we can still recover a consistent estimate of YLL using observed deaths and a LE that represents the normal LE of a matched population that is not at risk. This requires all observed deaths - both those that occur below and those that occur above this LE to be taken into account. Those deaths occurring below LE generate YLL - compared to the average of a matched population not at risk. However, we must also account for those deaths that occur at ages above LE. These deaths generate life years 'gained' (YLG) compared to the average of a matched population not at risk. Therefore, the appropriate estimate is a net YLL (i.e., YLL – YLG). In effect, by subtracting YLG from YLL we take account of the fact that not all deaths below LE are excess deaths but some deaths above LE are (see Appendix C for more formal explanation of the equivalence of these ways of calculating YLL).[40]

Using the life expectancy of the general population

Routinely available data provides the age and gender of observed deaths but no information about the age and gender distribution of the at risk population itself. Using observed age and gender at death as an indication of the distribution of the at risk population will significantly overestimate the LE of a normal matched population insofar as a disease may be chronic (not all PBC mortality occurs on entry into the at risk population), and that PBC related mortality risk may increase with age (see Table C14 Appendix C).[41]

---

[39] These 'counterfactual' deaths will occur in the other PBCs insofar as all deaths are recorded in an ICD codes. Therefore, we take account of the unavoidable fact that everyone must die of something at some time. For example, even if all observed cancer mortality was avoidable and could in principle be eliminated with sufficient expenditure, lives would not be 'saved' but deaths delayed and reallocated to other causes. Note that the outcome elasticities are based on PBC mortality that is sensitive to changes in expenditure (i.e., is avoidable) at the margin so no assumptions about how much of the PBC mortality is avoidable is required.

[40] Simply taking the difference between a fixed LE and the age at death of deaths that occur below LE and ignoring those death that occur above LE, would only provide the correct figure if it is reasonable to assume that no deaths would have otherwise occurred prior to LE (so all 'normal' deaths must occur at LE) and that there are no deaths (survivors) beyond LE in the at risk population, i.e. all deaths below LE are excess deaths and there are no excess deaths above LE.

[41] If risk increases over the disease duration more deaths would be observed in groups that have been prevalent for some time (i.e., are older) than those that are incident. Also if PBC related mortality is higher for older age groups they will be overrepresented in observed deaths compared to a matched normal population. For both reasons LE, YLL and cost per life year would be overestimated using age at death as a proxy for the age distribution of the at risk population.

In the absence of additional external information the net YLL could be based on the life expectancy of the general population, reflecting its current age and gender distribution. These are reported in Table 4.3 and illustrate the impact of accounting for counterfactual deaths in the way described above. The YLL reported in column 5 of Table 4.3 are calculated the same way and are the same as the figures previously reported (column 5 of Table 4.2). That is, they do not account for deaths that would have otherwise occurred below LE or the very many deaths that occur above LE. With the exception of PBC18&19 many death occur above the LE of the general population (see column 4 in Table 4.3) in all PBCs. As a consequence there are LYG associated with all other PBCs (see column 6) so the net YLL in column 7 are lower than YLL based on the same life expectancy. Therefore, failure to account for counterfactual deaths would lead to an overestimate of the YLL associated with a PBC and the effects of expenditure on YLL. Consequently the cost per life year threshold would be underestimated (see Table 4.6).

Table 4.3: Net YLL using life expectancy of the general population

PBC		LE of Males [1]	LE of Females [2]	Average 2006-2008 Deaths <LE [3]	Deaths >LE [4]	YLL [5]	YLG [6]	Net YLL [7]
1	Infectious diseases	80.7	84.4	3,710	3,248	62,052	18,796	43,256
2	Cancer	80.7	84.4	95,213	35,597	1,345,038	175,350	1,169,689
4	Endocrine	80.7	84.4	4,000	2,764	65,016	15,864	49,152
7	Neurological	80.7	84.4	8,975	6,378	145,529	34,621	110,908
10	Circulatory	80.7	84.4	82,099	77,752	916,192	444,694	471,498
11	Respiratory	80.7	84.4	30,500	34,945	310,334	215,829	94,505
13	Gastro-intestinal	80.7	84.4	15,827	8,320	273,308	45,295	228,012
17	Genito-urinary	80.7	84.4	4,198	6,427	39,099	40,530	-1,431
18+19	Maternity & neonates	80.7	84.4	226	0	17,167	0	17,167

However, these figures are only correct insofar as the distribution of age and gender in each PBC is similar to the general population. For example, if the at risk population tends to be younger the correct LE for the PBC will be lower and the net YLL will also tend to be lower. Similarly if the at risk population tends to be older than the general population the correct LE will be higher and net YLL will also tend to be higher.[42] This explains the apparent net gain in YLL (negative net YLL) for PBC17 (Genito-urinary) where most deaths occur at ages greater than the LE of the general population so that LYG exceeds YLL. As we are able to show later (see Table 4.4) this is because the age distribution in this PBC tends to be older than the general population, i.e., the LE for a matched normal population should be higher with fewer deaths above and more below this LE.

Using additional information about age and gender distribution

It is evident that estimates of YLL require some account to be taken of counterfactual deaths. In the absence of routinely available information this requires examination of alternative sources of information which might provide a basis for more credible assumptions about the age and gender distribution of the PBC population than either, the distribution of observed deaths or the general population.[43] The WHO Global Burden of Disease (GBD) study, updated in 2008 using 2004 data (see Addendum 1 in Appendix C for more details)[44] provides a range of summary health indicators for the UK, which are, in part, based

[42] A higher (lower) LE will mean that there are more (less) deaths below LE, each generating more (fewer) YLL and fewer (more) deaths above LE each generating fewer (more) LYG.

[43] Although this research was not funded to purchase access to GPRD data we were able to examine a sample of it which comprised of 22,313,086 rows/patient–ICD10 events (3 digit) representing 4,229,910 patients with data on new diagnosis of diseases observed between 1 Jan 2006 and 24 June 2011 (see Addendum 1 in Appendix C). Although GPRD data could, in principle, provide this type of information the difficulties of reliability, face validity and interpretation of the sample data in the form available to us meant that it was not directly useful. We discuss the potential value of other sources of information, including GPRD in Chapter 5.

[44] We are aware that the 2000-2002 WHO GBD study and the update which was published in 2008 using 2004 data has itself recently been updated. However, the report and tools where not publically available at the time this research was conducted. We discuss the potential of future sources of information in Chapter 5.

on estimates of the incidence of sequelae associated with different types of disease by age and gender[45]. Therefore, the type of information used by WHO in the GBD Study to generate summary estimates for the UK can also be used to improve the assumptions required about the age and gender distribution of the PBC populations. Importantly, at this stage, we do not need to rely on estimates of the absolute size of the at risk population, but only the relative 'share' by age and gender.

GBD classifies diseases by U-codes, which are groups of three digit ICD-10 codes (see Addendum 1 in Appendix C for details of how U-codes map to ICD-10 codes).[46] Since we know which ICD codes contribute to each PBC we can map information from U-codes to PBCs via the ICD codes that contribute to each. The resulting average age and life expectancy for each PBC is reported in columns 3 and 4 of Table 4.4 using the information available from GBD in combination with life tables for the general population.

Table 4.4: Average age and life expectancy for PBCs based on GBD

PBC		Sex	Average age of general population [1]	LE of general population [2]	Average age in PBC (GBD) [3]	LE of at risk population (GBD) [4]
1	Infectious diseases	m	38.5	80.7	28.6	79.6
		f	40.8	84.4	30.2	83.6
2	Cancer	m	38.5	80.7	61.3	83.0
		f	40.8	84.4	52.3	84.7
4	Endocrine	m	38.5	80.7	44.2	81.0
		f	40.8	84.4	50.8	84.7
7	Neurological	m	38.5	80.7	24.8	79.6
		f	40.8	84.4	23.5	83.3
10	Circulatory	m	38.5	80.7	55.4	83.0
		f	40.8	84.4	57.9	86.5
11	Respiratory	m	38.5	80.7	32.1	80.3
		f	40.8	84.4	33.7	84.0
13	Gastro-intestinal	m	38.5	80.7	35.8	80.6
		f	40.8	84.4	41.9	84.5
17	Genito-urinary	m	38.5	80.7	63.2	83.5
		f	40.8	84.4	47.3	85.6
18+19	Maternity & neonates	m	38.5	80.7	3.0	78.7
		f	40.8	84.4	24.1	83.1

These summary estimates suggest that some of the PBC populations may be on average be older than the general population (e.g., Cancer, Circulatory and Genito-urinary PBCs) or younger (e.g., Maternity & neonates, Infectious diseases and Neurological). However, when trying to interpret these summaries it should be noted that the average age reported in Table 4.4 is the average over the ages at which sequelae occur within the ICDs contributing to the PBC. Therefore, a similar average age can reflect very different age distributions. Some reflect a markedly bimodal distribution, e.g., Respiratory, where there is high incidence at very young and older ages, or very different age distributions across the type of diseases that contribute to the PBC. For example PBC7 (Neurological) includes dementia which accounts for the vast majority of the PBC population older than 70. However, a greater proportion of the population is in

[45] WHO, through the National Burden of Disease toolkit reports UK specific information about the incidence and duration of sequelae associated with different types of disease by age and gender. Since it is possible that a patient may experience more than one of the types of sequelae reported in GBD we use the gender and age distribution of the sequelae with the highest prevalence (evaluated as incidence x duration) to evaluate the age and gender distribution within each disease, i.e., the minimum estimate of prevalence consistent with these figures (see Section C2.1.3 and Addendum 1 in Appendix C).

[46] Throughout the analysis in Chapter 4 mortality, life years and QALY were not assigned to procedural ICD codes (Section C2.1.3 Appendix C) as these are likely to be evident in other ICD codes related to the procedure. This means that no health effects are associated with PBC 22 Social Care (which only includes procedural ICD codes), although changes in expenditure on PBC 22 are included. This is likely to overestimate the threshold because any health effects associated with PBC 22 will not be reflected in the estimated outcome elasticities of other PBCs unless the effects happen to be correlated with changes in expenditure in those other PBCs.

much younger age groups with other conditions, especially migraine (see Addendum 1 Appendix C). When interpreting these summary estimates it should also be noted that the reported life expectancies are not the life expectancies at the average ages reported in column 3, but the average over the life expectancies for each age group within the contributing ICDs weighted by the age distribution of sequelae from GBD U-codes.

The implications for net YLL of using these PBC specific estimates of 'normal' life expectancy are reported in Table 4.5. As expected, the net YLL for those PBC with a LE greater than the general population are higher than those reported in column 5 in Table 4.3 (e.g., PBC10 Circulatory and PBC17 Genito-urinary, which now has positive net YLL). Similarly those PBCs with a LE less than the general population have lower net YLL than reported in column 5 in Table 4.3 (e.g., PBC1 Infectious diseases and PBC18 & 19 Maternity & neonates, where the effect of a lower LE is more modest as there are no deaths above either of the estimates of LE).

Table 4.5: Net YLL using life expectancy for each PBC

PBC		Average 2006-2008						
		LE of Males [1]	LE of Females [2]	Deaths <LE [3]	Deaths >LE [4]	YLL [5]	YLG [6]	Net YLL [7]
1	Infectious diseases	79.6	83.6	3,498	3,460	58,686	21,724	36,962
2	Cancer	83.0	84.7	101,203	29,607	1,473,733	126,549	1,347,184
4	Endocrine	81.0	84.7	4,068	2,696	66,283	15,058	51,225
7	Neurological	79.6	83.3	8,370	6,983	135,686	41,770	93,917
10	Circulatory	83.0	86.5	96,694	63,157	1,102,020	278,251	823,768
11	Respiratory	80.3	84.0	29,549	35,897	298,343	230,313	68,030
13	Gastro-intestinal	80.6	84.5	15,824	8,323	273,117	45,414	227,703
17	Genito-urinary	83.5	85.6	4,969	5,655	47,229	29,101	18,127
18+19	Maternity & neonates	78.7	83.1	226	0	16,801	0	16,801

The impact on the cost per life year threshold of the issues discussed in Sections 4.1, 4.2 and 4.3 are summarised in Table 4.6 (see Table C16 in Appendix C for detailed breakdown of changes in spend and YLLs across PBCs).

Table 4.6: Summary of cost per life year threshold

	Using cut-off in estimating YLL (ONS)		Using net YLL estimates	
	cut-off of 75 [1]	cut-off of LE of the GP [2]	Using LE of the GP [3]	Using LE of the PBC population (GBD) [4]
big 4 PBCs	£10,398	£5,487	£10,421	£8,080
11 PBCs (with mortality)	£20,031	£10,660	£19,928	£15,628
All 23 PBCs (zero health effects for remaining 12 PBCs)	£73,697	£39,218	£73,317	£57,497
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£22,639	£12,048	£22,523	£17,663

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

Using ONS data to calculate YLL in the same way as the published NHS IC figures, but overcoming some of the issues associated with the reporting of mortality at PCT level and the coverage of published estimates of YLL (see Section 4.2.1), generates similar estimates of a cost per life year threshold (see column 1 Table 4.6) to those reported in Chapter 3. Calculating YLL in the same way, but based on the life expectancy of the general population significantly overestimates YLL for the reasons set out in Section 4.2.2 so underestimates the cost per life year threshold (see column 2). Taking account of counterfactual deaths by calculating net YLL based on the life expectancy of the general population (see column 3) provides similar estimates to those reported in Chapter 3. Assuming that PBC populations have the same age and gender distribution as the general population when the, albeit limited, information that is available suggests otherwise, seems inappropriate. Therefore, our preferred central estimate of the cost per life year threshold is reported in column 4. These are lower than those based on the general population, reflecting the impact on net YLL of evidence that the population at risk in some key PBCs

(especially PBCs 2 and 10) tend to be older than the general population. In Section 4.2.5 we consider extreme upper and lower bounds that might be placed on this central estimate.

4.2.4 Inferring excess deaths

We have been able to establish a measure of net YLL, which takes account of deaths that would have occurred anyway below a normal LE for the PBC population (i.e., not all deaths observed in a PBC are excess) and that some deaths observed above this LE would not otherwise have occurred at that age (i.e., some of these deaths are excess). As explained in Section 4.2.3, net YLL calculated in this way is equivalent to first establishing the number of excess deaths at each age, then calculating YLL for each excess death (based on the LE conditional on the age at which each excess death occurred) and then summing these YLL across all excess deaths (i.e., across all ages). In other words, the estimates of net YLL imply a number of excess deaths required to generate them in each PBC. Therefore, it is possible to solve for the total number of excess deaths based on the net YLL and the average YLL per observed death.[47] The net YLL divided by the average YLL per death provides the number of excess deaths required, which on average will generate the estimated net YLL.[48]

The implied excess deaths associated with net YLL based on the LE of the PBCs (see column 7 Table 4.5) are reported in Table 4.7. With the exception of PBC18&19, excess deaths are some proportion of total observed deaths in each PBC. The proportion of excess deaths differs by PBC reflecting the distribution of deaths relative to the LE of the PBC.[49] For example, in those PBCs where a large proportion of deaths occur below LE (see column 3 and 4) excess deaths tend to be greater proportion of total deaths (e.g., PBC2, 13 and 10). Where most deaths occur above LE excess deaths as a proportion of total deaths tend to be lower (e.g., PBC1, 11 and 17).

Table 4.7: Excess deaths implied by net YLL

PBC		Net YLL [1]	YLL per observed death [2]	Excess deaths [3]	Total deaths [4]	% excess deaths [5]
1	Infectious diseases	36,962	13.4	2,797	6 958	40%
2	Cancer	1,347,184	14.1	95,715	130 810	73%
4	Endocrine	51,225	13.7	3,769	6 764	56%
7	Neurological	93,917	13.7	6,909	15 353	45%
10	Circulatory	823,768	10.5	79,218	159 851	50%
11	Respiratory	68,030	9.2	7,386	65 445	11%
13	Gastro-intestinal	227,703	15.2	15,199	24 147	63%
17	Genito-urinary	18,127	8.3	2,172	10 625	20%
18+19	Maternity & neonates	16,801	73.9	226	226	100%

Estimates of net YLL and changes in life years due to expenditure (see Table 4.5 and 4.6) have already accounted for the fact that not all deaths are excess and don't generate YLL. Nevertheless, solving for the number of implied excess deaths associated with these net YLL estimates allows a comparison of the cost per excess and observed PBC death avoided and an examination of the interpretation that can be placed of the life years expected to be gained from an excess or observed death averted. Since only deaths observed in the PBC can used to estimate the effects of expenditure (excess deaths are not directly

[47] The average of the sum of the YLLs for every observed death where the YLL for each observed death is the difference between age at death and LE conditional on age of death.

[48] In the absence of information about the age distribution of excess death this assumes that the average YLL associated with observed and excess deaths are similar. Insofar as excess deaths are thought likely to generate more YLL than observed deaths the number of excess deaths will tend to be overestimated. This would tend to underestimate the cost per excess death averted. However, the cost per life year estimates remain unchanged and do not require such an assumption.

[49] The impact of the age distribution of deaths and the age distribution of the at risk population (summarised as LE) on the calculation of excess deaths is not always obvious as both will affect the numerator (net YLL) as well the denominator (average YLL per death) in this calculation.

observed since they rely on an unobserved counterfactual population and would occur outside the PBC), the outcome elasticities can be interpreted as the proportionate change in observed PBC mortality due to a proportionate change in PBC expenditure. Equally, however, they can also be interpreted as the proportionate effect on excess death due to a proportionate change in expenditure so can be applied to either total observed or total excess deaths.[50]

The cost per excess death and the cost per PBC death averted are reported in Table 4.8 (see Table C19 in Appendix C for a detailed breakdown of changes in spend and excess or PBC deaths across PBCs). The cost per PBC death averted is, of course; significantly lower than the cost per excess death as excess deaths are only a proportion of total deaths (see Table 4.7). Also the cost per PBC death averted are substantially lower than those reported in Chapter 3 (see, Tables B8.22 and B8.23 in Appendix B), since these estimates do not restrict the effects of expenditure to PBC deaths under 75.[51] The cost per PBC or excess death averted (or life saved) should not be over interpreted because they are of little direct policy interest since lives are never saved (death is only delayed) and the significance of a death averted depends critically on how long it is averted for (the life years gained – see Table 4.6) and the quality of life in which additional years are lived (see Section 4.3).

Table 4.8: Summary of the cost per death averted threshold

	Cost per excess death averted, £ [1]	Cost per PBC death averted, £ [2]
big 4 PBCs	£91,129	£32,864
11 PBCs (with mortality)	£177,692	£64,774
All 23 PBCs (zero health effects for remaining 12 PBCs)	£653,748	£238,310
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£200,829	£73,208

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

However, establishing the number of excess and PBC deaths averted which are associated with net YLL is useful because it enables an assessment of the number of life years gained associated with each death averted. On average across all 11 PBCs each excess death averted is associated with 11.4 life years gained. These are reported for each PBC in Table C21 in Appendix C and range from 74.3 years per excess death for PBC 18 & 19 Maternity & neonates to 8.3 for PBC17 Genito-urinary. However, clinicians or the evaluative literature cannot distinguish whether an observed death is excess or not. What can be observed is whether groups of similar patients with and without access to a treatment survive and for how long. Therefore, it is the life years associated with each observed death that provides a context that can be interpreted based on experience and evidence of how effective those interventions that could be

[50] Observed PBC mortality that is sensitive to changes in expenditure can be regarded as 'avoidable' and it is only this mortality that contributes to the estimates of outcome elasticities (not all observed mortality is necessarily avoidable and sensitive to expenditure - such mortality will not contribute to the estimates). Not all observed mortality is excess when compared to the counterfactual population but this is unrelated to the question of how sensitive it is to expenditure, i.e., observed mortality will be just as sensitive to expenditure whether or not it is regarded as excess. Therefore, the estimated outcome elasticities can be applied to either observed PBC deaths or excess PBC deaths

[51] Recall from Chapter 3 and appendix B that the measure of mortality that is available at PCT level and used to estimate the outcome elasticities is restricted to deaths under 75, as are the published estimates of YLL associated with them (see Section 4.2.2). However, to restrict effects only to those under 75 would imply that there is no excess mortality above 75 or equivalently that there are no health effects of PBC expenditure above 75. Rather than assume no affects of NHS activity in older populations we apply the effects that can be observed to the whole PBC but account for deaths that would otherwise occurred in our estimate of net YLL in Section 4.2.3. In many respects whether or not PBC deaths at older ages are as sensitive to changes in expenditure is not critical since any observed deaths that might be averted at older ages are less likely to generate life years gained because they are more likely to have occurred anyway in that year (i.e., are excess so generate zero life years gained anyway). Therefore, they will have very limited impact on cost per life year or subsequently on cost per QALY estimates in Sections 4.3 and 4.4). For this, and the reasons given in the text, it is the cost per life year rather than cost per death averted, whether excess or observed, that is of primary interest.

invested or disinvested tend to be. The average life years expected to be gained associated with each observed PBC deaths averted takes account of that fact that some deaths that are avoided in the PBC are not delayed for very long but quickly occur[52] elsewhere and do not generate LY gained (i.e., they were not excess deaths). These are also reported for each PBC in Table C21 in Appendix C and range from 74.3 years per observed death for PBC 18 & 19 Maternity & neonates[53] to 1.0 for PBC11 Respiratory problems, i.e., the YLL per PBC death are much lower for those PBCs where a small proportion of observed deaths are excess. On average across all 11 PBCs each PBC death averted is associated with 4.1 life years gained.

4.2.5 Summary of cost per life year estimates

The sequence of analysis set out above has enabled an examination of the impact of the limitations associated with the incomplete reporting mortality data at PCT level and incomplete coverage of published YLL estimates. We have also been able to consider effects above 75 while taking account of that fact that many deaths would have occurred anyway, despite the limited information available about the population at risk within a PBC. The GBD Study does provide some information about the age and gender distribution of the population at risk in a PBC so offers some improvement over the other assumptions that would otherwise be required (i.e., that the distribution of age and gender is the same as the general population or follows the distribution of observed deaths). For this reason the cost per life year threshold in column 4 of Table 4.6 and repeated in lines 1 to 4 in Table 4.9 are regarded as the central or best estimates given the evidence available and the credibility of alternative assumption that could be made. As explained in Section 4.1, these are based on the conservative assumption that any health effects of changes in expenditure are restricted to one year, which, to some extent, may be offset by the more optimistic assumption any death averted returns the individual to the mortality risk face by the general population, matched for age and gender.

Table 4.9: Summary of the cost per life year threshold with upper and lower bounds

	Best estimate	
Effect of expenditure on mortality:	*1 year*	
YLL per PBC death averted:	*~ 4.1 YLL ***	
big 4 PBCs	£8,080	[1]
11 PBCs (with mortality)	£15,628	[2]
All 23 PBCs (zero health effects for remaining 12 PBCs)	£57,497	[3]
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£17,663	[4]
	Lower bound	
Effect of expenditure on mortality:	*Remainder of disease*	
YLL per PBC death averted:	*~ 4.1 YLL ***	
big 4 PBCs	£3,846	[5]
11 PBCs (with mortality)	£6,106	[6]
All 23 PBCs (zero health effects for remaining 12 PBCs)	£22,463	[7]
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£6,901	[8]
	Upper bound	
Effect of expenditure on mortality:	*1 year*	
YLL per PBC death averted:	*2 YLL*	
big 4 PBCs	£16,432	[9]
11 PBCs (with mortality)	£32,387	[10]
All 23 PBCs (zero health effects for remaining 12 PBCs)	£119,155	[11]
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£36,604	[12]

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal. ** see Tables C14, C15 and C18 in Appendix C

[52] What portion of observed deaths are regarded as excess depend on how time is discretised. The data available reports deaths in annual intervals so in this context 'quickly' means within one year. If deaths were reported in narrower time intervals then a greater proportion of observed deaths would be regarded as excess and in the limit with continuous time all observed deaths would be excess. Of course, the average YLL associated with them would be smaller and is approximated by the net YLLs reported in Table 4.5 per observed death (the effects of approximation is likely to be small but unavoidable as it is due to deaths being reported in annual intervals).

[53] This is the same as life years associated with excess deaths since all observed deaths in this PBC are excess.

It does not seem credible to imagine that NHS expenditure has no health effects in the 12 PBC which do not have sufficient mortality reported at PCT level to estimate outcome elasticities - what is implied by the estimate reported in line 3. Therefore, it is the estimates reported in lines 2 and 4 that are of policy interest. The estimate of £15,628 per life year (line 2) is restricted to the effects of changes in expenditure in the 11PBCs where outcome elasticities can be estimated. The threshold of £17,663 per life year uses the estimated health effects of expenditure in these PBC as a surrogate for health effects in the others, i.e., assuming that the effects that can be observed will be similar to those that cannot. However, no health effects are assigned to PBC23 (General Medical Services) on the basis that any health effects of this expenditure would be recorded in the other PBCs.[54]

The extreme upper and lower bounds for the cost per life year thresholds in Table 4.9 are based on making the necessary assumptions about duration of health effects and how long a death might be averted optimistic (providing the lower bound for the threshold) or conservative (an upper bound for the threshold). The lower bound (lines 5 to 8) is based on assuming that health effects are not restricted to one year but apply to the whole of the remaining disease duration of the population at risk in PBCs during the expenditure year.[55] Although this combines optimistic assumptions, it is possible, indeed likely, that at least some expenditure may have effects on the health outcomes of future patients that are not currently part of the population at risk in a PBC, e.g., investments or disinvestment in prevention will have an impact on populations that are incident to PBCs in the future. Such effects are not captured in any of the estimates presented in this chapter so all are conservative with respect to this type of health effect of changes in expenditure.

The upper bound (lines 9 to 12) is based on the combination of assuming that health effects are restricted to one year for the population currently at risk and that any death averted is only averted for the minimum duration consistent with the mortality data. The econometrics work used the average of 3 years of mortality (2006 to 2008), so the estimated outcome elasticities are based on differences in mortality that remain after averaging over three years. Therefore the estimated effects are based on differences in observed PBC deaths that must have been sustained, on average, for more than a minimum of 2 years.[56]

4.3 Adjusting life years for quality of life

The central or best estimates of the cost per life year threshold, which were presented in Table 4.9 (lines 2 and 4) take no account of the health related quality of life in which years of life, expected to be gained or lost through changes in expenditure, are likely to be lived. Even if attention is restricted to the direct health consequences of changes in mortality, estimates of the cost per life year will tend to overestimate the effects of changes in expenditure (underestimate the threshold) compared to a more complete

---

[54] It would be inappropriate to assign all the change in GMS expenditure to the estimate of cost per life year based only on the 11 PBCs with outcome elasticities because it would imply that GMS only contributes to these PBCs. Restricting attention to the 11 PBCs with outcome elasticities but allocating part of the change in GMS expenditure to them based on their proportional share of changes in overall expenditure would yield the same cost per life year as reported in line 4. It should be noted that including changes in GMS expenditure but not assigning health effects to this PBC is likely to overestimate the threshold because any health effects associated with GMS will not be reflected in the estimated outcome elasticities of other PBCs unless the effects happen to be correlated with changes in expenditure in those PBCs.

[55] Estimates of the duration of disease for each U-code are available from the GBD Study (see Table C22 and Addendum 1 in Appendix C). This information is also used in Sections 4.4.

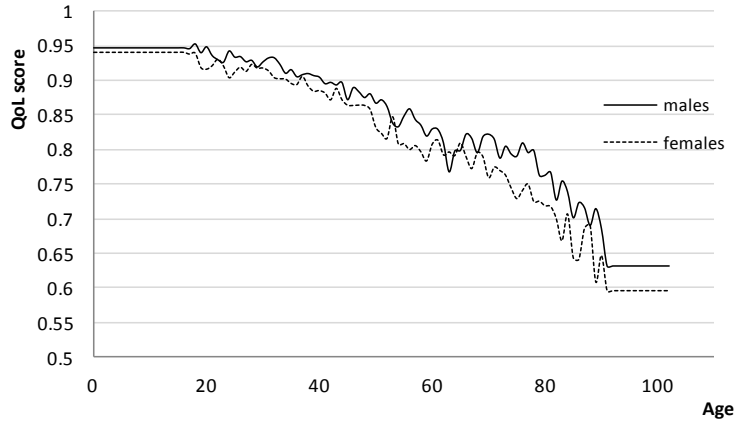
[56] Variation in mortality the first year of data will only contribute to these estimates if differences are sustained for a minimum of 3 years. Similarly variation in mortality in the second (third) year will only contribute if it is sustained for a minimum of 2 (1) years. If differences in mortality are similar each year (contribute equally to the estimates) then estimated effects must have been sustained on average for a minimum of 2 years. Since some of the variation in mortality in 1st year that is not sustained to the 3rd year will nevertheless be sustained for 1 or 2 years, 2 life years per death averted represents somewhat less than the minimum consistent with restricting life years gained to the observed mortality data. Of course, this is minimum difference in observed rather than unobserved counterfactual excess deaths. Nonetheless it can be interpreted as an upper bound given the data available and therefore the analysis that has been feasible.

measure of health that accounts for the quality in which of the years of life are expected to be lived. In this Section we examine the ways in which the life years reported in Section 4.2 can be adjusted for quality, taking account of information that is available about: i) how quality of life differs by age and gender (see Section 4.3.1), and ii) how the quality of life years associated with mortality changes might be effected by the types of diseases that make up each PBC (see Section 4.3.2). Throughout we continue to take account for counterfactual deaths in the way described in Section 4.2.3 by making the adjustment for quality to the life years associated with every observed death before calculating a quality adjusted net YLL. The implications for a cost per quality adjusted life year (QALY) threshold that only accounts for the health effects of mortality changes are presented in Section 4.3.3. In Section 4.4 we explore the ways in which the likely direct effects of expenditure on quality of life (other than through mortality) might also be taken into account.

4.3.1 Quality of life based on the general population

The most commonly used metric of health related quality of life in the UK is EQ5D,[94] which is specified in the NICE reference case for methods of technology appraisal.[1] This metric has 5 dimensions of quality each with three possible levels. Each of these 243 possible health states is valued relative to a score of one, which represents full or best imaginable health (the best score across all 5 dimensions), and a score of zero, which represents death, based on a representative sample of the UK population.[95] Therefore, insofar as the years of life expected gained or lost through changes in expenditure would be lived in this state of full health the cost per life year thresholds reported in Table 4.9 would also be the cost per QALY thresholds, albeit ones that only account for the health effects of mortality changes. However, unsurprisingly, there is good evidence that, on average, the general population is not in this state of full health. Therefore, the quality of life score associated with the health states experienced by the general population are less than 1, decline with age and differ by gender. These quality of life 'norms' for the general population by age and gender are illustrated in Figure 4.1 based on an analysis of data from the Health Survey for England (HSE).[57]

Figure 4.1: Quality of life for the general population by age and gender



These quality of life norms can be applied to the YLL associated with all observed deaths in each PBC, taking account of gender and age at death. The results are reported in column 4 to 6 of Table 4.10. There are two effects of adjusting life years for quality: i) since quality of life norms are always less than 1 the adjusted YLL and YLG are always lower than the unadjusted values in columns 1 and 2 (previously reported in Table 4.5); and ii) deaths above LE are necessarily at older ages with poorer quality of life norms than those below, so the difference between adjusted and unadjusted values is greater for YLG

[57] See Addendum 1 in Appendix C for a description on HSE data and section C2.2.1 of appendix C for the analysis of quality of life norms illustrated in Figure 4.1.

than YLL. The overall effect of quality adjustment on net YLL is the balance of these two effects. The overall effect of quality adjustment is to reduce the net YLL (compare Colum 6 and 3).[58]

Table 4.10: Net YLL adjust for the quality of life 'norms'

PBC		Unadjusted life years			Quality adjusted life years		
		YLL [1]	YLG [2]	Net YLL [3]	YLL [4]	YLG [5]	Net YLL [6]
1	Infectious diseases	58,686	21,724	36,962	47,481	14,618	32,864
2	Cancer	1,473,733	126,549	1,347,184	1,143,445	84,036	1,059,409
4	Endocrine	66,283	15,058	51,225	52,856	9,973	42,883
7	Neurological	135,686	41,770	93,917	109,349	28,262	81,087
10	Circulatory	1,102,020	278,251	823,768	848,046	183,330	664,717
11	Respiratory	298,343	230,313	68,030	231,578	154,743	76,835
13	Gastro-intestinal	273,117	45,414	227,703	216,256	30,277	185,979
17	Genito-urinary	47,229	29,101	18,127	35,929	18,947	16,982
18+19	Maternity & neonates	16,801	0	16,801	14,568	0	14,568

The quality adjusted net YLL figures in column 6 suggests that the health effects of mortality are lower than when relying only on unadjusted life years in Section 4.2. Therefore, the health effects of changes in expenditure on this more complete measure of health are lower. The implications of these adjustments on a cost per QALY threshold that only accounts for the direct health effects of mortality are reported in Table 4.11. As expected the cost per QALY threshold based on adjusting the life years gained or lost (column 2) is higher than a threshold based on unadjusted life years (column 1 and previously reported in Table 4.9).

Table 4.11: Summary of cost per QALY threshold based on population norms and mortality effects

	Cost per life year threshold [1]	Cost per QALY threshold Population norms [2]
big 4 PBCs	£8,080	£9,631
11 PBCs (with mortality)	£15,628	£18,622
All 23 PBCs*	£17,663	£21,047

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal except GMS.

4.3.2 Adjusting age related quality of life for disease decrements

Adjusting life years for age and gender related quality of life norms assumes that any life year gained through a change in expenditure would be lived in a similar quality of life to the general population. It is possible however, that patients benefiting from reduced mortality may, nevertheless, continue to be effected by the type of diseases that make up each PBC and experience the quality of life associated with the original disease.

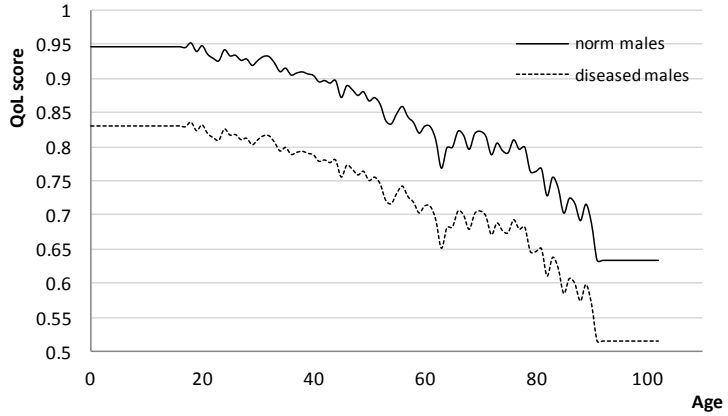
The Health Outcome Data Repository (HODaR)[96] provides over 30,000 observations of EQ-5D measures of quality of life by ICD code and the age and gender of the patients in the sample (see Addendum 1 Appendix C). Although this is a rich UK data set, there were a limited number of observations for some of the less common ICD codes. For this reason HODaR was supplemented with information from the Medical Expenditure Panel Survey (MEPS)[97] which also provides EQ-5D by ICD and reports the average age of respondents (see Addendum 1 Appendix C). These data provided a means of estimating the quality of life associated with each ICD code at the average age of respondents in the pooled sample.[59] The quality of life associated with each PBC can be expressed as an average of the

[58] The only exception is PBC11 (Respiratory) which has a large proportion of deaths occurring above the life expectancy of the PBC population (see Table 4.5).

[59] ICD estimates of the quality of life score and age were pooled across datasets by considering the number of patients from each dataset contributing to estimates, i.e. a weighted average.

quality of life associated with its component ICDs.[60] The quality of life effects of being in each PBC can then be expressed as a disease related decrement compared to the population norms at the same age (see Table C29 in Appendix C). This is illustrated for PBC1 (Infectious disease) in Figure 4.2, where the weighted average of quality of life scores across the component ICD codes was 0.667, at an average age average age of 54 for male respondents. Since the quality of life norms for males age 54 is 0.859 this suggests a decrement associated with membership of PBC1 of 0.192, which can then be applied to quality of life norms by age.[61]

Figure 4.2: Quality of life for males in PBC1 (Infectious disease) and the general population by age



Quality of life norms adjusted for disease related decrements can be applied to the YLL associated with observed deaths in each PBC, taking account of gender and age at death in the same way as Section 4.3.1.[62] The results are reported in column 4 to 6 of Table 4.12. The overall effect of quality adjustment that also applies a disease related decrement is to reduce the net YLL to a greater extent than adjustment with population norms alone (compare column 6 in Table 4.12 to column 6 in Table 4.10).

[60] The average quality of life scores across the ICDs which contribute to each PBC and the average age and gender of respondents were used to calculate a PBC disease related decrement based on quality of life norms from the general population. This 'PBC decrement' could then be applied to each observed death and the age at which each life year was gained or lost. In Section 4.4 information about the relative share of different types of disease (U-codes) within a PBC and the information about which ICDs are more likely to contribute to the effects of changes in PBC expenditure are explored.

[61] In principle it would be possible to estimate disease related disutility by age rather than assume a fixed decrement. HODaR does provide age for each reported quality of life score but MEPs only provides average age of respondents in published summaries. However, even with access to 'raw' scores and the age and gender of each, it is very unlikely that there would be sufficient data to estimate age related decrements in each of the component ICDs. It would, however, be possible to assume a proportionate rather than fixed decrement by age. Since the average age of respondents in the pooled HODaR and MEPs sample tends to be older than the age distribution of the PBC populations (see Table C29 and C13 in Appendix C) this would tend to increase the quality adjusted net YLL and reduce the cost per QALY threshold compared to the fixed decrement applied here.

[62] The quality of life score was applied to each observed death considering the age at which each life year was gained or lost (from ONS) the 'PBC decrements' from HODaR and MEPS.

Table 4.12: Net YLL adjusted for disease and age related quality of life

PBC		Unadjusted life years			Quality adjusted life years		
		YLL [1]	YLG [2]	Net YLL [3]	YLL [4]	YLG [5]	Net YLL [6]
1	Infectious diseases	58,686	21,724	36,962	37,055	10,793	26,262
2	Cancer	1,473,733	126,549	1,347,184	955,690	67,930	887,760
4	Endocrine	66,283	15,058	51,225	43,394	7,844	35,550
7	Neurological	135,686	41,770	93,917	68,893	15,842	53,050
10	Circulatory	1,102,020	278,251	823,768	656,145	135,241	520,905
11	Respiratory	298,343	230,313	68,030	169,269	106,505	62,764
13	Gastro-intestinal	273,117	45,414	227,703	163,593	21,677	141,916
17	Genito-urinary	47,229	29,101	18,127	29,749	15,152	14,598
18+19	Maternity & neonates	16,801	0	16,801	13,662	0	13,662

It should be noted that combining quality of life adjustments for both population norms and disease related decrements assumes that any life years gained due to a reduction in mortality will be lived in the diseased state until life expectancy, i.e., that all diseases are not just chronic but disease duration is lifelong. Inevitably this assumption means that the health effects of changes in mortality will be reduced. Consequently the cost per QALY threshold reported in Table 4.13 (column 2) will be higher than adjusting life years gained for population norms in Table 4.11.

Table 4.13: Summary of cost per QALY threshold based on disease related decrements

	Cost per life year threshold [1]	Cost per QALY gained Disease related decrements [2]
big 4 PBCs	£8,080	£12,109
11 PBCs (with mortality)	£15,628	£23,395
All 23 PBCs*	£17,663	£26,441

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal except GMS.

4.3.3 Summary of the cost per QALY threshold based only on mortality effects

The analysis to this point is summarised in Table 4.14. The three estimates of a cost per QALY threshold are based on assuming that each life year gained is either: lived in full health (see column 1, equal to the cost per life year estimates in Table 4.9), lived in a quality of life that reflects age and gender norms of the general population (column 2); or lived in a quality of life that reflects the original disease state (column 3).

Assuming that life years gained are lived in full health is not credible and should be regarded as an underestimate of the threshold, given what is known about quality of life norms for the general population (see Figure 4.1). Equally, assuming that all life years gained are lived in the quality of life of the original disease state does not seem credible either and is likely to overestimate the threshold since it assumes that all disease is not only chronic but lifelong and all life years would be lived in the diseased state until death.[63] Although adjusting life years gained for the quality of life of the general population taking account of age and gender (in column 2) is likely to underestimate a cost per QALY threshold based only on mortality effects, it probably represents the 'best' of the three alternative estimates available at this stage of the analysis (see Section 4.4.2 for how analysis based on measures of QALY burden allows this assumption to be relaxed).[64] The lower and upper bounds are based on combining optimistic and

[63] The information that is available about disease duration suggests that many types of disease that comprise the PBCs are not chronic and certainty not lifelong (see Table C22 in Appendix C). In Section 4.4 we take account of quality of life experienced while alive in the diseased state.

[64] In section 4.4.2 measures of QALY burden are used as the basis of estimating the health effects of changes in expenditure. This analysis applies the estimated proportionate effect of changes in expenditure on life year burden of disease to measures of the total QALY burden. This is equivalent to assigning a proportional adjustment to the quality of life with disease to life years gained.

pessimistic assumptions about the duration of health effects and how long a death might be averted as described in Section 4.2.5.

Table 4.14: Summary of QALY threshold estimates based only on mortality effects

	[1] *(QoL score =1)*	[2] *(QoL norm)*	[3] *(QoL diseased)*	
	Best estimate			
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per death averted:	*~4.1YLL ***	*~4.1YLL ***	*~4.1YLL ***	
QALYs per death averted	*~4.1QALYs*	*~3.5QALYs*	*~2.8QALYs*	
big 4 PBCs	£8,080	£9,631	£12,109	[1]
11 PBCs (with mortality)	£15,628	£18,622	£23,395	[2]
All 23 PBCs*	£17,663	£21,047	£26,441	[3]
	Lower bound			
Effect of expenditure on mortality:	*Remainder of disease*	*Remainder of disease*	*Remainder of disease*	
YLL per PBC death averted:	*~4.1YLL ***	*~4.1YLL ***	*~4.1YLL ***	
QALYs per death averted	*~4.1QALYs*	*~3.5QALYs*	*~2.8QALYs*	
big 4 PBCs	£3,846	£4,252	£5,319	[4]
11 PBCs (with mortality)	£6,106	£6,852	£8,568	[5]
All 23 PBCs*	£6,901	£7,744	£9,683	[6]
	Upper bound			
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per PBC death averted:	*2 YLL*	*2 YLL*	*2 YLL*	
QALYs per death averted	*2QALYs*	*~1.9QALYs*	*~1.5QALYs*	
big 4 PBCs	£16,432	£17,456	£21,747	[7]
11 PBCs (with mortality)	£32,387	£34,492	£42,967	[8]
All 23 PBCs*	£36,604	£38,983	£48,561	[9]

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal except GMS.
** see Tables C14, C15 and C18 in Appendix C

However, it should be noted that these cost per QALY thresholds only account for the direct health effects of changes in mortality due to changes in expenditure. Insofar as much, or at least some, of NHS activity and expenditure is intended to improve quality of life, not just mortality, then these estimates will underestimate total health effects and overestimate a cost per QALY threshold based on a more complete measure of possible health effects. In Section 4.4 we explore the ways in which the likely effects of expenditure on quality of life (other than through mortality) might also be taken into account.

4.4 Including quality of life effects during disease

The cost per QALY thresholds presented in Section 4.3 only account for the health (QALY) effects of changes in mortality due to changes in expenditure. It does not seem credible to suppose that all NHS activity and expenditure only influences mortality with no effect on the quality of life while alive and experiencing a disease. Insofar as changes in NHS expenditure will also affect quality of life as well as mortality then total health effects will be underestimated and the thresholds presented in Table 4.14 will overestimate the cost per QALY threshold. In this section we explore ways to also take account of those effects on health not directly associated with mortality and life year affects (i.e., the 'pure' quality of life effects) to estimate an overall cost per QALY threshold.

The routine reporting of quality of life outcomes are increasingly available at PCT level (see Addendum 1 in Appendix C for a description of these data). In principle, the variation in such measures of outcome across PCTs could be used to estimate outcome elasticities for quality of life rather than mortality effects using similar econometric methods to those described in Chapter 3 (see Section B8.8 in Appendix B for the results of an exploratory econometric analysis of these data). However, the currently limited coverage of routine reporting of these outcomes means that it is not feasible to estimate quality of life effects across all the PBCs using these data. In Section 5.8 we discuss how these data might be used to improve estimates of the threshold as the coverage and routine reporting of quality of life outcomes improves and how the analysis presented in Section 5.3 might help prioritise reporting in particular areas (i.e., those PBCs and ICD codes that have the greatest influence on estimates of the threshold).

Here we explore how estimates of effects of expenditure that can be observed (i.e., on mortality) can be used to infer the likely effects on what cannot be directly observed (quality of life), rather than making extreme assumptions that are not credible (e.g., assuming that changes in expenditure will have no effects on quality of life outcomes).

In Section 4.4.1 we use three alternative estimates of the ratio of QALYs to life years lost due to different types of disease as a means of inferring the change in QALYs that is likely to be associated with the estimated change in YLL, i.e., applying the total QALYs lost associated with each YLL with disease. This is consistent with regarding the estimates of the mortality and life year effects as a surrogate for a more complete measure of the health effects of a change in expenditure.

However, these ratios of QALYs lost to life years lost due to disease in those PBC where outcome elasticities could not be estimated cannot inform estimates of the threshold (there are no estimated life year effects with which to apply the ratios). Nonetheless, the sources of information on which ratios are based also provides much of the information required to calculate the QALY burden of disease in these areas, which can be used to inform estimates of the threshold. Therefore, in section 4.4.2 we use estimates of the QALY burden of disease, infer a proportionate effect on burden from the estimated effects on life years, and then apply this proportionate effect to the measures of QALY burden for all the other PBCs. In this way we can use all the information available about the mortality and quality of life effects of the different types of disease that make up each PBC, including those where mortality based outcome elasticities are not available.

4.4.1 Using ratios of QALYs to YLL

The ratio of the total QALYs to years of life lost (YLL) due to a disease indicates the number of QALYs associated with each YLL. Therefore, any change in YLL is likely to generate a number of QALYs indicated by the ratio - if it is reasonable to interpret the estimated effects on mortality and life years as a surrogate for a more complete measure of total health effects. For example, a disease with a ratio greater than 1 suggests that each YLL across the at risk population is associated with more than one QALY, i.e., where there are significant quality of life effects while experiencing the disease.[65] Therefore, a change in expenditure that leads to 1 life year gained in this type of disease maybe expected to generate more than one QALY and a greater QALY effect than the same life year effects in a disease where this ratio is less than 1, i.e., where most of the effect of disease is on mortality rather than quality of life. Therefore, information which allows these ratios to be estimated for the diseases that make up each PBC provides a means of accounting for the likely effect on quality of life other than through effects on mortality.

To understand the differences between the three ratios presented below it is useful to regard the total QALY lost to YLL ratio (R) for a particular disease as the sum of two ratios: i) the QALYs lost due to premature death to YLL ratio (R_{death})[66] and ii) the QALYs lost during disease (while alive) to YLL ratio (R_{alive}) (see Section C2.3.1 in Appendix C for more detailed explanation).

DALY to YLL ratios

The WHO GBD study provides UK specific estimates of the years of life lived with disability and the years of life lost due to different types of disease (classified by U-codes that can be mapped to ICD-10, see Section 4.2 and Addendum 1 in Appendix C). GBD uses Disability Adjusted Life Years (DALYs) as a measure of the burden of disease. This DALY measure has two components: i) the years of life lived with disability (YLD), which incorporates weights (between zero and one) to reflect the scale of disability experienced each year and the number of years lived with disability over the durations of disease; and ii)

[65] Insofar as YLL would not have been lived in full health (see Section 4.3), the quality of life effects during disease must offset the less than full quality of life of the YLL to generate a ratio greater than one. Therefore, ratios less than one are possible even when disease has measurable quality of life effects for those experiencing it.

[66] The analysis in Section 4.3 already implies an R_{death} ratio at PBC level – see the following main text.

the years of life lost (YLL). The total DALY associated with a disease is simply YLL+YLD. Therefore, the DALY to YLL ratio is (YLL+YLD)/YLL or equivalently YLL/YLL + YLD/YLL. Since the first term (YLL/YLL = R_{death}) must equal one and the second (R_{alive} = YLD/YLL) must be ≥ 0, a ratio based on DALYs must necessarily be bounded by below by one. This is illustrated in Table 4.15a for four different types of diseases (classified by U-codes) which reflect diseases where mortality is the major component (e.g., U016) and where the impact of disease on the quality of life while alive is the major component (e.g., U141).

Table 4.15a: Examples of DALY to YLL ratios

Ucode	DALY ratios	(R_{death} + R_{alive})
U037 (Other infectious diseases)	1.23	(1+0.23)
U016 (Tetanus)	1.00	(1+0)*
U061 (Mouth and oropharynx cancers)	1.05	(1+0.05)
U141 (Spina bifida)	2.34	(1+1.34)**

* Given the short disease duration, it is only mortality effects that contribute to the ratio
* Quality of life effects during disease contribute significantly to estimates of the ratio

Adjusting DALYs for quality of life norms

The use of DALY ratios bounded below by one essentially assumes that YLL would have otherwise been lived in a state of full health. As was discussed in section 4.3.1 this is not credible given information available about the quality of life in the general population (see Figure 4.1). It would lead to over estimating the QALYs associated with mortality and life year effects and underestimating the cost per QALY threshold. Therefore, it is important to adjust these DALY ratios for the quality of life norms by age and gender in the same way as described in Section 4.3.1. The effect of this adjustment [67] is illustrated in Table 4.15b. Now those types of disease where mortality rather than quality of life with the disease is the major component can have ratios less than one. Indeed the first term of these ratios (R_{death}) is consistent with, and is implied by, the analysis in Section 4.3.1 where the ratio of quality adjusted net YLLs to unadjusted net YLLs represents this ratio on average for each PBC.

Table 4.15b: Examples of modified DALY to YLL ratios

Ucode	Modified DALY ratios	(R_{death} + R_{alive})
U037 (Other infectious diseases)	1.01	(0.78+0.23)
U016 (Tetanus)	0.78	(0.78+0
U061 (Mouth and oropharynx cancers)	0.83	(0.78+0.05)
U141 (Spina bifida)	2.18	(0.85+1.34)

Using quality of life estimates (based on HODAR and MEPS)

The disability weights used in the DALY measure are not based on the same description of health states as the EQ5D measure, nor are the weights based on a representative sample of the UK population responding to choice based elicitation questions. EQ5D based quality of life decrements (adjustments to age related quality of life norms) associated with different types of disease can be estimated from HODaR and MEPS data (previously described in Section 4.3.2).[68] These disease related quality of life decrements be can be calculated for each U-code (based on the contributing ICD codes) so can be used to replace the DALY disability weights in R_{alive} reported in Tables 4.14a and 4.14b.[69] This final adjustment is illustrated

[67] Reflecting the quality of life norms for the general population in Figure 4.1 and the distribution of ages and gender within each U-code (see Addendum 1 in Appendix C).
[68] Since quality of life effects of different disease states are expressed as age related decrements (see Figure 4.2) we do not require the HODaR and MEPS samples to necessarily be representative of the age distribution of the population at risk in the groups of ICD codes that make up each U-code.
[69] The average quality of life scores across the ICDs which contribute to each U-code (see Addendum 1 for how ICD codes map to U-codes) and the average age and gender of respondents from HODaR and MEPS were used to calculate a disease decrement for each U-code, based on quality of life norms from the general population. These U-code disease decrements can then be applied to the age and gender distribution of each U-code, based on

in Table 4.1.4c and turns, what were originally, DALY ratios into EQ5D QALY ratios.[70] For these reasons we regard the QALY to YLL ratios rather than DALY or modified DALY ratios as the preferred basis of estimating a cost per QALY threshold that provides a more complete picture of the likely health effects of changes in expenditure.

Table 4.15c: Examples of QALY to YLL ratios (HODaR and MEPS)

Ucode	QALY ratios (HoDAR and MEPs)	(R_{death} + R_{alive})
U037 (Other infectious diseases)	1.37	(0.78+0.60)
U016 (Tetanus)	0.78	(0.78+0
U061 (Mouth and oropharynx cancers)	0.80	(0.78+0.02)
U141 (Spina bifida)	1.88	(0.85+1.03)

Allocating effects at PBC level to ICD codes

Table 4.15 illustrate how QALY ratios can be calculated for and differ by U-code.[71] Unsurprisingly, these ratios differ across the type of diseases that make up each PBC (see Table C45 in Appendix C). When using this information to estimate a cost per QALY threshold the mortality and life year effects observed at PBC level must be allocated in some way to the component ICD codes before ratios are applied to LY effects and the resulting QALY effects are summed across all the contributing ICD codes.[72] For this reason it is important to consider how other information might inform the different ways in which the effects observed at PBC level might be generated by the distribution of impacts at ICD level, i.e., where investment or disinvestment is likely to occur within the PBC and therefore which ICDs are likely to contribute most to overall health effects.

An important and complementary element to the econometric analysis of routinely reported information at PBC level was to investigate whether other information, commonly available at a local level within the NHS, might provide a useful indication of where, within a PBC, investment or disinvestment is more likely across the NHS. The details of this investigation are reported in Addendum 2 in Appendix C. The review of local data sources suggested that there is very little routinely collected data on investment and disinvestment by local NHS organisations beyond the high-level aggregate data on spending by PB which are used in the econometric analysis. Although more disaggregated data on spending decisions about specific services relevant to particular ICD codes could in principle be acquired through additional

---

information from GBD about the prevalence and age distribution of each - using information about the incidence of sequelae associated with them (as described in Section 4.2.3) and information about the durations of disease (see Table C.22 Appendix C).

[70] For example, the evidence about quality of life from HODaR and MEPS suggests that the impact of U037 on quality of life is greater than indicated by DALY disability weights. The quality of life effects of U141, although still very significant, are lower than indicated by DALY disability weights.

[71] Information about the size and age and gender distribution is only available at U-code level. Therefore U-code ratios are applied to all the ICD codes that contribute to a particular U-code. Note that, unlike ICD codes, U-codes do not map directly to PBCs so some ICDs in different PBCs may belong to the same U-code and therefore have the same U-code ratio. Some ICDs are not included in the U-code classification of disease. Most of these are procedural codes where we do not assign life year and QALY effects anyway (any health effects would be evident in other ICD codes), so it was not necessary to impute ratios for them (84 out of 1562). Of the others most were associated with PBC16 with a zero outcome elasticity so did not require imputation either (186 out of 1562). Imputation based on the median ratio across the ICDs within the PBC was required for the remaining (482 out of 1562). Eighty eight of these cannot be mapped into U-codes. The remaining 394 were associated with U-codes where the ratio was undefined because the denominator (YLL) was zero. In both these cases, values were imputed based on the median ratio across the ICDs within the PBC. Since the distribution of ratios within a PBC tend to be highly positively skewed, imputation based on the median is likely to be conservative with respect to health effects and especially in the latter case where mortality effects appear to be a much less important aspect of the disease.

[72] It is important to note that it would be inappropriate to calculate an average of the ratios within a PBC and then apply this 'average ratio' to life year effects at PBC level, rather than calculate QALY effects at ICD level by applying the relevant ratio. The results, however, can be presented as an implied PBC ratio (i.e., a ratio of averages), see Table C.43 in Appendix C.

primary research (surveys or Freedom of Information requests) this would be costly and with a risk that information acquired in this way may not be complete, consistent or representative.

In the absence of useful information at a local level it is possible to assume that a change in PBC expenditure will be allocated equally (on a per patient basis) across the component ICD codes, i.e., any investment or disinvestment is equally likely across the population at risk within the PBC. Hospital Episode Statistics (HES) (see Addendum 1 in Appendix C) provides information about the costs associated with each ICD by PCT so it is possible to establish which ICDs contribute most to the variability in HES costs within a PBC across PCTs. Those that contribute most to this variance might be expected to be more likely to have been subject to differential investment or disinvestment across PCTs.[73]

There are differences in relative weight assigned to ICD based on the size of the population or its contribution to variance in HES costs. If investment or disinvestment within a PBC tends to focus on ICD codes representing areas of marginal value the health effects of a change in PBC expenditure maybe overestimated and a cost per QALY threshold underestimated when allocating effects equally across the population at risk within each PBC. However, weighting ICDs based on HES data is likely to favour those ICDs which represent more severe disease requiring more hospital care. This may over represent ICDs with lower QALY to YLL ratios if mortality effects tend to be a major component of these types of disease and maybe conservative with respect to the health effects of changes in expenditure.

The implications for a cost per QALY threshold that uses the estimated mortality and life year effects as a surrogate for a more compete measure of the likely heath effects (i.e., that includes quality of life as well as quality adjusted life year effects) is summarised in Table 4.16. These results use the contribution to variance in HES costs to 'weight' the different ICD codes within a PBC (allocate the life year effects), before applying the QALY ratios associated with each ICD (see Table C41 in Appendix C).

Table 4.16: Summary of the QALY threshold using QALY to YLL ratios

	DALY ratios [1]	Modified DALY ratios [2]	QALY ratios, (HODaR and MEPS) [3]
big 4 PBCs	£5,402	£6,419	£5,990
11 PBCs (with mortality)	£9,958	£11,718	£10,297
All 23 PBCs	£11,254	£13,244	£11,638*

* Preferred analysis

The QALY to YLL ratio implied by this analysis for all 11 PBC with outcome elasticities is 1.52, which suggests that every (unadjusted) life year is associated with 1.52 QALYs on average across these PBCs. However, this implied QALY ratio differs across these PBCs, ranging from 0.79 in PBC2 to 15.05 in PBC18+19 (see Table C43 in Appendix C). Since all the analysis in this Section seeks to use the estimated mortality and life year effects as a surrogate for a more complete measure of likely health effects, it is the cost per QALY threshold for all 23 PBCs that is most relevant. As expected this threshold (£11,638), is lower than a cost per QALY threshold based only the quality adjusted life year effects (£21,047 in Table 4.14 that assumes no effects of NHS expenditure on quality of life itself). This difference gives some indication of the relative importance of QALY effects due to avoidance of premature death and the QALY effects of avoiding disability during disease.

Table 4.17 reports how the estimated QALY effects for each PBC can be decomposed into that part associated with quality adjusted life year effects and that part associated with 'pure' quality of life effects. These results appear credible for the first 11PBCs, where those for which mortality is the major concern

[73] Unfortunately total PBC costs are not available at ICD level across PCTs so could not be used for this purpose. Costs from HES data are only a component of total PBC costs (41% of total PBC costs for the 11 PBCs where mortality effect can be estimated) and contribute less to the variability in PBC costs across PCTs (HES contribute only 23% of the variability for the 11 PBCs where mortality effect can be estimated)

have a much greater share of total QALY effects associated with avoidance of premature death (e.g., PBC2 and PBC10) compared to those where quality of life is the major concern (e.g., PBC 7).[74]

Table 4.17: Decomposing estimated QALY effects by PBC

PBC		QALY change (total)	QALY change (death)	% QALY gained due to avoidance of premature death	% QALY gained due to avoidance of disability while alive
2	Cancer	1,699	1,641	97%	3%
10	Circulatory	6,713	4,856	72%	28%
11	Respiratory	3,215	923	29%	71%
13	Gastro-intestinal	3,605	1,193	33%	67%
1	Infectious diseases	27	11	40%	60%
4	Endocrine	2,036	323	16%	84%
7	Neurological	342	52	15%	85%
17	Genito-urinary	12	6	52%	48%
16	Trauma & injuries*	0	0	NA	NA
18+19	Maternity & neonates*	273	15	6%	94%
3	Disorders of Blood	1,087	547	50%	50%
5	Mental Health	19,828	9,979	50%	50%
6	Learning Disability	2,990	1,505	50%	50%
8	Problems of Vision	2,348	1,181	50%	50%
9	Problems of Hearing	621	313	50%	50%
12	Dental problems	2,282	1,148	50%	50%
14	Skin	1,021	514	50%	50%
15	Musculo skeletal	1,469	739	50%	50%
20	Poisoning and AE	426	215	50%	50%
21	Healthy Individuals	1,781	896	50%	50%
22	Social Care Needs	6,566	3,304	50%	50%
23	Other	0	0	NA	NA

The ratios of QALYs to YLL due to disease in those PBC where outcome elasticities could not be estimated cannot be used to inform estimates of the threshold because there are no estimated life year effects with which to apply the ratios. Therefore, as in previous sections, the estimated effect of expenditure on health for the 11 PBCs with outcome elasticities is applied to the estimated changes in PBC expenditure for the other 12 PBCs (excluding GMS for the reasons given in Section 4.2), i.e., assuming that the health effects that can be observed of a change in expenditure will be similar to those that cannot. However, the use of QALY ratios also implies that the share of total health effects between quality adjusted life year effects and that part associated with 'pure' quality of life effects are also similar to those PBC with estimated outcome elasticities. Summing the different types of health effects across these 11PBCs suggests that 50% is due to avoidance of premature death and 50% due to avoidance of disability. This is clearly not credible when applied to the other PBCs, e.g., mental health, vision and hearing are likely have a much greater share of total health effects associated with quality of life effects and very little associated with premature mortality.

The problem is that using QALY to YLL ratios means that much of the information that is available about the other 12 PBCs cannot be used to inform the estimates of the cost per QALY threshold. Fortunately, the sources of information on which ratios are based also provide much of the information required to calculate the QALY burden of disease in these areas. Section 4.4.2 explores how measures of burden can be used to estimate a cost per QALY threshold that captures the likely effects of a change in expenditure on all aspects of health while using all the information that is available about all the PBCs.

[74] It should be noted that the implied QALY ratio of 1.52 for the 11 PBC with outcome elasticities is a ratio of QALYs to unadjusted YLL. The proportion of total QALY effects due to premature deaths for the same PBCs (50% in Table 4.17) also implies a ratio - equal to two. However, this is a ratio of total QALY effects to quality adjusted YLL. The difference between these two ratios is the denominator, i.e., quality adjusted YLL are lower than unadjusted YLL.

4.4.2 Using estimates of the QALY burden of disease

In this Section we use estimates of the QALY burden of disease, infer a proportionate effect on burden from the estimated effects on life years, and then apply this proportionate effect to the measures of QALY burden for all PBCs. In this way we can use all the information available about the mortality and quality of life effects of the different types of disease that make up each PBC, particularly those where mortality based outcome elasticities are not available. Applying a proportionate effect to measures of QALY burden of disease is equivalent to assuming that any effects on life years are lived at quality of life that reflects a proportionate improvement to the quality of life with disease[75] It also allows quality of life effects of changes in expenditure to be included; also based on proportionate improvement in the quality of life with disease.

The total QALY burden of disease for the population with disease in a particular year includes: i) the quality adjusted years of life lost due to all the disease related mortality that could occur in this population over their remaining duration of disease and ii) the reduction in quality of life while alive also for their remaining disease duration. However, applying the estimated proportionate effects on mortality and life years to such a measure of total burden would provide an estimate of the effects of a change in expenditure, not just in one year, but in all the remaining years of disease for the population at risk in that year. Recall from Section 4.2 that we have adopted the conservative assumption that changes in expenditure will only have health effects in one year for the population with disease in that year. Therefore, it is not a measure of total burden that is required, but a measure of the QALY burden of disease during one year for the population with disease (prevalent and incident) in that year. The estimated outcome elasticities can then be appropriately applied to this measure of burden.[76]

The information from GBD used to derive QALY ratios in Section 4.4.1 includes information about the YLL and duration of disease for those incident to a U-code, i.e., the measure of QALY burden from the information included in the ratios is a measure of the total burden of the disease but only for the population that is incident (rather total population with disease) in one year. Assuming that incidence is stable over the disease duration this is also equivalent to the QALY burden of disease during one year for the population with disease (i.e., those that are incident and prevalent) in that year.[77]

However, in moving from ratios to absolute measures of burden it becomes more important to examine and then adjust for any inconsistency between information about YLL and size of the incident population from GBD (which is available by U-codes and can be mapped to ICDs), and the information about net YLL and observed deaths for each PBC based on ONS data as described in Section 4.2.3 (see Table C44 in Appendix C).[78]

---

[75] In Section 4.3 each life year gained could be assumed to be lived in full health, lived in a quality of life that reflects age and gender norms of the general population or lived in a quality of life that reflects the original disease state. Applying an estimated proportionate effect on the life year burden of disease to measures of QALY burden of disease implies a proportionate improvement in the quality of life with disease applied to any life year effects. Therefore, basing estimates on measures of QALY burden provides are more conservative estimate of the QALY effects of changes in mortality than the best estimate reported in Section 4.3, which was based on quality of life norms.

[76] Of course it would be possible to solve for a lower outcome elasticity that could be applied to total burden which would return the required estimate of total QALY effects restricted to one year - see Section 2.1 in Appendix C

[77] So long as estimates of the quality of life decrement of disease from HODaR and MEPS are representative of average effects across those earlier (incident) and later (prevalent) in their disease duration an assumption of constant quality of life decrement with respect to disease duration is not required.

[78] There are a number of reasons for potential inconsistencies: i) GBD is based on earlier years of mortality data; ii) the imprecision of mapping from U-codes to PBC via ICD codes; and iii) the YLL reported in GBD are calculated in the same way as published NHS IC estimates (see Section 4.2.2 and 4.2.3) and will tend to overestimate the net YLL (see Table 52 in Appendix C). The YLL by U-code, reported in GBD, that are mapped to ICDs are adjusted by these proportionate differences to ensure that the YLLs associated with all contributing ICD codes are consistent with (do not over estimate) the net YLL for the PBC as a whole. However, due to the earlier years of data and imprecision in mapping from U-codes to ICDs there might also be some inconsistency in estimates of the total

The implications for the cost per QALY threshold of using information about the QALY burden of disease for all PBCs rather than QALY ratios for those where an outcome elasticity can be estimated are reported in Table 4.18. The QALY effects of a change in PBC expenditure are a weighted average of the QALY effects within each of the ICDs that contribute to the PBC. The figures reported in column 2 are based on weighing the effects at ICD level by the proportion of the total PBC population within each contributing ICD code, rather than the contribution to variance in HES costs.[79]

Table 4.18: Summary of the cost per QALY threshold

	Cost per QALY gained*	
	QALY ratios, (HoDAR and MEPs) [1]	QALY Burden (HoDAR and MEPs) [2]
big 4 PBCs	£5,990	£3,036
11 PBCs (with mortality)	£10,297	£5,128
All 23 PBCs	£11,638	£15,701*

* Preferred analysis

The cost per QALY threshold for the 11PBCs with outcome elasticities is lower using a measure of QALY burden (£5,128) rather than the QALY ratios (£10,297) described in Section 4.2.1. This is in part because GBD calculates YLL in the same way as published NHS IC figures so will tend to overestimate a net YLL which accounts for counter factual deaths (see Section 4.2.3). This will make little difference to the first term in the QALY ratio (R_{death}) used in Section 4.2.1 since an overestimate of YLL affects both denominator and numerator of the ratio. However, the second term (R_{alive}) is likely to be underestimated. Therefore the ratios used in section 4.4.1 will tend to underestimate the QALY effects of expenditure and overestimate the cost per QALY threshold (see Table 4.18). We are able to adjust the GBD based measure of QALY burden for this overestimation of net YLL in calculating the QALY threshold reported in column 2).[80]

Since the purpose of this Section is to use the estimated mortality and life year effects as a surrogate for a more complete measure of likely health effects, it is the cost per QALY threshold for all 23 PBCs that is of most relevance. The cost per QALY threshold for all 23 PBCs is based on applying the proportionate effects on the QALY burden of disease, based on the observed effects of changes in expenditure on mortality in the 11 PBC with outcome elasticities,[81] to the QALY burden of disease in the other PBCs. This generates a much higher cost per QALY threshold (£15,701) than one based on applying the estimated QALY effects of changes in expenditure, using QALY ratios for the 11 PBC with outcome elasticities, to changes in expenditure in the others (£11,638). The reason is that the QALY burden of disease in the other PBC is, in general, lower than the QALY burden of disease across those PBCs where outcome elasticities can be estimated (see Table C45 in Appendix C). Therefore, applying the same proportionate effects to a lower QALY burden generates a smaller health effect of a change in

---

incidence of disease for a PBC. Insofar as disease related mortality risk is stable, the same number of deaths should be observed in GBD and ONS data for the same at risk population. The PBC deaths recorded in GBD and those observed in ONS data (see Table 52 in Appendix C) are similar but nonetheless the proportionate difference is used to adjust the scale of quality of life burden while alive based on GBD information (equivalent to adjusting estimates of incidence). Notable exceptions are PBC1 and PBC18+19 where the discrepancies are due to imperfect mapping from U-code to PBC via ICD codes.

[79] HES costs are a much smaller proportion of total PBC expenditure for the 11 PBCs where a mortality effects could not be estimated (HES costs account for less than 15% of total PBC expenditure) and account for very little of a the variability in PBC costs across PCTs (the contribution that variance in HES costs makes to variance in PBC expenditure in this group of PBCs is less than 8%). Therefore, allocating PBC level effects to ICDs based on contribution to variance in HES costs is less appropriate when information about QALY burden in this groups of PBCs is used to inform the estimate of the overall threshold.

[80] See previous footnote and Table 52 in Appendix C.

[81] Note that this is the ratio of total change in health to total change in expenditure across these PBC (rather than an average ratio) and the contribution that each of these PBCs make to these total effects on health and expenditure depends on the estimated expenditure as well as outcome elasticities.

expenditure.[82] In essence the difference between these estimates is that in column 1 the absolute effect on health associated with an absolute change in expenditure is extrapolated to the other PBCs, where as in column 2 it is the relative effect on health of an absolute change in expenditure that is extrapolated. Since we know that QALY burden differs between (and within) PBCs and especially between the groups of PBCs with and without estimated outcome elasticities (see Table C45 in Appendix C), [83] it is the values based on QALY burden in column 2 that are regarded as most credible and represent our central or best estimate.

A detailed breakdown of changes in expenditure and changes in QALYs across all PBCs is provided in Table C48 in Appendix C when the analysis is based on QALY ratios and when based on QALY burden of disease. A comparison of these values confirms that QALY effects for the other PBC are lower and therefore the cost per QALY for each of these PBCs are in general much higher when based on a proportionate effect on QALY burden. Of course, we have not directly observed quality of life effects in these PBC but inferred them from the proportionate effects that we can observe. Insofar as investment and disinvestment opportunities in these PBCs might have been more valuable (offered greater improvement in quality of life)[84] than suggested by the implied PBC thresholds, then overall QALY effects will tend to be underestimated and the cost per QALY threshold overestimated. For the reasons discussed in previous sections, we regard all the cost per QALY threshold reported in column 2 of Table 4.18 as on balance conservative with respect to overall health effects of a change in expenditure. However, the estimate of £15,701 maybe especially conservative with respect to health effects (i.e., overestimated) based, as it is on an extrapolation of the proportionate effects to measures of burden on these PBC, rather than observations of the direct impact of changes in expenditure on quality of life in these types of disease. This is especially so in PBC 5 Mental Health Disorders, which accounts for a large proportion of the change in overall expenditure (30%) and where a review of the evidence suggests that the investment and disinvestment opportunities in this PBC are likely to have been more valuable than the implied PBC cost per QALY of £60,111 (see Addendum 3 Appendix C)[85]. The lower cost per QALY threshold for the 11PBCs with outcome elasticities (£5,128) might be regarded as more secure in this respect but they only account for a proportion (27%) of any change in overall expenditure (see Table C53 in Appendix C).[86]

---

[82] Applying the absolute health effect of expenditure from the 11 PBCs with outcome elasticities implies different (higher) proportionate effects in the other PBCs

[83] The QALY burdens per incident patient are reported in this Table for each PBC, including the median and range across the contributing ICD codes. However, these values should not be over interpreted as the 'average' QALY burden for the PBC depends on how PBC effects are allocated to ICDs and the 'average' burden for groups of PBCs depends on how a change in overall expenditure is shared between them, i.e., the expenditure elasticities estimated for each PBC in Chapter 3 and Appendix B.

[84] See Addendum 3 in Appendix C for an examination of the value of investment and disinvestments that may have been available in PBC5 (Mental Health Disorders), which accounts for much of the change in overall expenditure. This qualitative analysis suggests that these may well be more valuable than the implied PBC cost per QALY of £60,111 reported in Table C56 in Appendix C.

[85] See footnote above.

[86] It is not possible to estimate expenditure equations for all 23 PBCs simultaneously (see Section 5.8), so the 23 independently estimated expenditure elasticities may not necessarily account for all of a change in overall spend, i.e., the sum of changes in PBC expenditure based on a 1% change in total spend and the estimated PBC expenditure elasticities is less than a 1% change in total spend. Previously in Chapter 3 and Section 4.2, 4.3 and 4.4.1 any remaining change in total spend was assigned to the other 11 PBCs where outcome elasticities could not be estimated (in these Sections expenditure elasticities for these PBCs were not estimated because the same health effect of expenditure was assumed for these PBCs so it did not matter how spend was allocated between them). However, in this section it does matter how the remaining change in expenditure is allocated between the other 11 PBCs as they have different QALY burdens so different implied health effects of expenditure. Therefore, the remaining change in total spend is allocated between these 11 PBCs reflecting the relative share of changes in expenditure based on their estimated expenditure elasticities. This does mean that a greater proportion of a change in overall expenditure tends to be allocated to this group of PBCs. Since these PBCs tend to have lower QALY burden and a higher implied PBC cost per QALY this will tend to overestimate the overall cost per QALY threshold.

Table 4.19 reports how the estimated QALY effects based on measures of QALY burden for each PBC can be decomposed into that part associated with life year effects adjusted for quality and that part associated with 'pure' quality of life effects. These results are very similar to those reported in Table 4.17 which were based on QALY ratios for the 11 PBCs with an estimated outcome elasticity. Those PBCs for which mortality is the major concern have a much greater share of total QALY effects associated with avoidance of premature death (e.g., PBC2 and PBC10) compared to those where quality of life is the major concern (e.g., PBC 7). The differences tend to favour QALYs gained though avoidance of disability, which reflects the underestimation of the effects on 'pure' quality of life when using QALY ratios based on estimates of YLL from GBD (see the discussion above).[87] The QALY to YLL ratios that are implied by this analysis are reported in Table C50 Appendix C. As expected the implied QALY ratio across all 11PBCs with outcome elasticities is higher (3.05[88]) then reported in Section 4.4.1 because the previous bias against quality of life effects by using QALY ratios based on unadjusted GBD information has been removed.

Table 4.19: Decomposing estimated QALY effects by PBC

PBC		QALY change (total) [1]	QALY change (death) [2]	% QALY gained for premature death [3]	% QALY gained for disability while alive [4]
2	Cancer	1,501	1,393	93%	7%
10	Circulatory	5,908	4,054	69%	31%
11	Respiratory	19,869	758	4%	96%
13	Gastro-intestinal	2,776	1,024	37%	63%
1	Infectious diseases	53	9	18%	82%
4	Endocrine	4,887	269	5%	95%
7	Neurological	963	43	4%	96%
17	Genito-urinary	24	5	22%	78%
16	Trauma & injuries*	0	0	NA	NA
18+19	Maternity & neonates*	10	7	69%	31%
3	Disorders of Blood	689	35	5%	95%
5	Mental Health	3,397	296	9%	91%
6	Learning Disability	125	25	20%	80%
8	Problems of Vision	240	9	4%	96%
9	Problems of Hearing	434	3	1%	99%
12	Dental problems	489	0	0%	100%
14	Skin	107	39	37%	63%
15	Musculo skeletal	1,697	84	5%	95%
20	Poisoning and AE	54	9	16%	84%
21	Healthy Individuals	23	4	16%	84%
22	Social Care Needs	0	0	NA	NA
23	Other	0	0	NA	NA

In Section 4.4.1 the ratios of QALYs to YLL due to disease in those PBC where outcome elasticities could not be estimated could not be used to inform estimates of the threshold or indicate how any total health effects in these other PBCs are likely to be 'shared' between life year effects adjusted for quality and that part associated with 'pure' quality of life effects (see Table 4.17). By applying the observed proportionate effects of changes in expenditure to measures of QALY burden of disease in these other PBCs the likely share of any effects on QALYs between avoidance of premature mortality and avoidance of disability more closely reflect the nature of these types of diseases (see Table 4.19). As expected, a much greater proportion of QALY effects are associated with quality of life during the disease compared to the 11PBCs where mortality based outcome elasticities could be estimated. The share of effects in particular PBCs are also much more credible. For example, in PBC5 Mental Health Disorders the

[87] The exception is PBC 18 &19. The reason is that there are significant adjustments made based on differences in observed and recorded mortality (to adjust for differences in classification when mapping from U codes to PBCs via ICDs) as well as differences in YLL due to the GBD method of calculation (see Table 52 in Appendix C).

[88] The implied QALY ratios across these 11 PBCs range from 0.70 in PBC2 Cancer to 14.86 in PBC7 Neurological.

overwhelming share of QALY effects are associated with quality of life itself and for others, such as PBC12 Dental problems, PBC9 Problems of Hearing and PBC8 Problems of Vision; almost all effects are associated with quality of life rather than mortality and life years. For this, and the other reasons discussed above, the analysis based on measures of QALY burden are regarded as the best estimate of a cost per QALY ratio that reflects a more complete picture of the likely health effects of changes in overall expenditure.

4.4.3 Summary of the cost per QALY threshold

The results of the three sequential steps of analysis described in this Chapter are summarised in Table 4.20. In Section 4.2 we explored ways in which the estimated effects on mortality from the econometrics work in Chapter 3 might be better translated in to life year effects by overcoming some of the limitations of mortality data available at PCT level and taking account of counterfactual deaths. The results of this analysis were reported in Table 4.9 and are repeated in column 1 of Table 4.20.[89] In Section 4.3 we considered how the estimated life year effects might be adjusted for the quality of life in which they are likely to be lived, taking account of the gender and the age at which life years are gained or lost (see Table 4.14). The results of this analysis are repeated in column 2 below. Finally in Section 4.4 we explored ways to also take account of the likely effects of changes in expenditure on quality of life during disease as well as the effects associated with mortality and life years (see column 3). These estimates provide our central estimate of a cost per QALY threshold, because they make best use of available information while the assumptions required, which on balance are likely conservative with respect to health effects, appear more reasonable than the other alternatives available.[90]

Table 4:20: Summary of cost per QALY threshold estimates

	[1]	[2]	[3]	
QoL associated with life extension:	*1*	*Norm*		
QoL during disease:	*0*	*0*	*Based on burden*	
			Best estimate	
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per death averted:	*~ 4.1 YLL*	*~ 4.1 YLL*	*~ 4.1 YLL*	
QALYs per death averted:	*~ 4.1 QALY*	*~ 3.5 QALY[1]*	*~ 12.6 QALY*	
big 4 PBC's	£8,080	£9,631	£3,036	[1]
11 PBCs (with mortality)	£15,628	£18,622	£5,128	[2]
All 23 PBCs	£17,663	£21,047	**£15,701**	[3]
			Lower bound	
Effect of expenditure on mortality:	*Remainder of disease duration*	*Remainder of disease duration*	*Remainder of disease duration*	
YLL per death averted:	*~ 4.1 YLL*	*~ 4.1 YLL*	*~ 4.1 YLL*	
QALYs per death averted:	*~ 4.1 QALY*	*~ 3.5 QALY*	*~ 12.6 QALY*	
big 4 PBC's	£3,846	£4,252	£674	[4]
11 PBCs (with mortality)	£6,106	£6,852	£860	[5]
All 23 PBCs	£6,901	£7,744	**£2,785**	[6]
			Upper bound	
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per death averted:	*2 YLL*	*2 YLL*	*2 YLL*	
QALYs per death averted:	*~ 2 QALY*	*~ 1.9 QALY*	*~ 6.1 QALY*	
big 4 PBC's	£16,432	£17,456	£6,292	[7]
11 PBCs (with mortality)	£32,387	£34,492	£10,626	[8]
All 23 PBCs	£36,604	£38,983	**£32,537**	[9]

[89] The cost per life year threshold in Table 4.9 can be interpreted as cost per QALY thresholds conditional on the assumption that all life years are lived in full health and the quality of life with disease is zero (equivalent to death).
[90] Note that the proportionate difference between the estimates in column 3 and columns 1 and 2 are greater in lines 1 and 2, reflecting the additional health effects from considering the likely impact of changes in expenditure on quality of life during disease. These differences are less marked in line 3 because the effects in those PBCs where an outcome elasticity can be estimated are extrapolated to the other PBCs using proportionate effect on QALY burden and measures of QALY burden in these other PBCs (see the discussion in Section 4.4.2 for a more details).

The estimate of £5,128 per QALY (line 2) is restricted to the effects of changes in expenditure in the 11PBCs where outcome elasticities can be estimated. Although this might be regarded as more secure these PBCs only account for a proportion of a change in overall expenditure (approximately 28%, see Table 61 in Appendix C). The threshold of £15,701 uses the estimated proportionate effects of expenditure on the QALY burden of disease in these PBC as a surrogate for proportionate effects in the others, i.e., assuming that the effects that can be observed will be similar to those that cannot. As discussed in Section 4.4.2 there are reasons to suspect that this may underestimate health effects in these PBCs which have most influence on the overall threshold. As in previous sections, no health effects are assigned to PBC23 (General Medical Services) on the basis that any health effects of this expenditure would be recorded in the other PBCs.[91] Therefore, the best or central estimate of cost per QALY threshold is £15,701 (column 3, line 3). However, this estimate reflects changes in undiscounted QALYs associated with changes in expenditure. Although all the health effects of a change in expenditure are restricted to one year (so no discounting is necessary) some of the quality adjusted life year effects of a change in mortality in that year will occur in future years, so in principle should be discounted. However, discounting these life year effects, even at the higher rate of 3.5% recommended by NICE, only increases the cost per QALY threshold to £15,940 (see Table C52 in Appendix C for discounted values).

As in previous Sections of this Chapter, the upper and lower bounds for the cost per QALY thresholds in column 3 are based on making the necessary assumptions about duration of health effects and how long a death might be averted optimistic (providing the lower bound for the threshold) or conservative (an upper bound for the threshold). The lower bound (lines 4 to 6) is based on assuming that health effects are not restricted to one year but apply to the whole of the remaining disease duration of the population at risk in PBCs during one year. Although this combines optimistic assumptions, it is possible that at least some part of a change in expenditure may prevent disease so will have an impact on populations that are incident to PBCs in the future. Such effects are not captured in any of the estimates presented in this Chapter so all are conservative with respect to this type of health effects of expenditure. The upper bound (lines 7 to 9) is based on the combination of assuming that health effects are restricted to one year for the population currently at risk and that any death averted is only averted for 2 years (see Section 4.2.5).

[91] It would be inappropriate to assign all the change in GMS expenditure to the estimate of cost per QALY based only on the 11 PBCs with outcome elasticities because it would imply that GMS only contributes to these PBCs. Restricting attention to the 11 PBCs with outcome elasticities but allocating part of the change in GMS expenditure to them based on their proportional share of changes in overall expenditure would yield a slightly higher cost per QALY than reported in line 2. It should be noted that including changes in GMS expenditure but not assigning health effects to this PBC is likely to overestimate the threshold because any health effects associated with GMS (or PBC 22 see Footnote 48 and 56) will not be reflected in the estimated outcome elasticities of other PBCs unless the effects happen to be correlated with changes in expenditure in those PBCs.

Chapter 5: Implications for a policy threshold

5.1 Introduction

The three sequential steps of analysis, which provide a cost per life year threshold (see Section 4.2 of Chapter 4) through a cost per life year adjusted for quality (see Section 4.3) to a cost per QALY threshold (see Section 4.4), have been explained in this Chapter using the analysis of 2006 expenditure and mortality data from 2006 to 2008 (see Section 3.5.1 in Chapter 3 and Section B8.5 in Appendix B) to illustrate the implications for the threshold estimates. At each step we explored the different ways that routinely available data could be used and how additional information could improve our estimates. In doing so we identified a preferred analysis at each stage based on which made the best use of available information, whether the necessary assumptions appeared more reasonable than the alternatives available, and which provided a more complete picture of the likely health effects of a change in expenditure. Although other assumptions and judgments are possible that retain some level of plausibility, they do not necessarily favour a higher threshold. Indeed, when considered together, they suggest that on balance the central or best estimate presented in Chapter 4 and in Table 5.1 below is, if anything, likely to be an overestimate (see Section 5.4 for a more detailed discussion and summary). In Section 5.8 we discuss how some of these remaining uncertainties might be resolved through access to additional and better data and the type of analysis that would then be possible.

5. 2 Re-estimating the cost per QALY threshold using more recent data

The same methods of analysis can be applied to the econometric analysis of the 2008 expenditure and 2008 to 2010 mortality data (see Section 3.5.3 in Chapter 3 and Section B11 in Appendix B). The differences between the 2006 analysis reported in Chapter 4 and the analysis of expenditure in 2008 reported below are the: i) total PBC expenditure; ii) estimated expenditure elasticities; iii) estimated outcome elasticities; iv) observed PBC deaths by age and gender; and v) life expectancy by age and gender. The other information about quality of life norms (see Section 4.3.1), disease related decrements in quality of life (see Section 4.3.2) and the information from GBD about incidence and duration of disease remain unchanged between 2006 and 2008 (we discuss how these estimates might be improved through access to more recent and better data in Section 5.8).

It should be noted that important improvements were made to the classification and collection of PBC expenditure data that took place after the 2006 data were collected. Therefore, the differences in threshold estimates between 2006 and 2008 partly reflect this (see Section 3.5.4 and B11.4 in Appendix B) so should not be over interpreted. The results of the analysis of 2007 and 2008 expenditure are comparable in this respect, providing insights into how the threshold might change over time and with changes in the overall budget. The implications of this analysis on the need for periodic reassessment are discussed in Section 5.6. For the purposes of this methodological research the 2008 expenditure and 2008 to 2010 mortality data were the latest to be analysed. Since it is the analysis of the most recent data that is of most policy relevance, our discussion throughout this Section is based on analysis of 2008 expenditure, although the same sensitivity analysis (see Section 5.3) and analysis of uncertainty (see Section 5.4) is available for 2006 and 2007 expenditure (see Section C.2.5 in Appendix C).

It is unnecessary to repeat all the analysis presented in Sections 4.2 to 4.4 (the details of each stage of the analysis of 2008 data can be found in Appendix C). Instead the results of the three sequential steps of analysis are summarised in Table 5.1. They include: i) the cost per life year (column 1)[92] based on the methods of analysis outlined in Section 4.2; ii) the cost per life year adjusted for quality of life (column

[92] The cost per life year threshold in column 1 can be interpreted as cost per QALY thresholds conditional on the assumption that all life years gained or lost are lived in full health but the quality of life with disease is zero (equivalent to death).

2)[93] based on the methods of analysis outlined in Section 4.3; and iii) the cost per QALY (column 3) based on the methods of analysis outlined in Section 4.4.2. These estimates, in column 3, take account of the likely effects of changes in expenditure on quality of life during disease as well as the effects associated with mortality and life years; making best use of available information, while the assumptions required appear more reasonable than the other alternatives available. For this reason these estimates remain our central or best estimates for all the waves of expenditure and mortality data.

Table 5.1: Summary of cost per QALY threshold estimates (expenditure in 2008)

	[1]	[2]	[3]	
QoL associated with life extension:	*1*	*Norm*		
QoL during disease:	*0*	*0*	*Based on burden*	
			Best estimate	
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per death averted:	*~ 4.5 YLL*	*~ 4.5 YLL*	*~ 4.6 YLL*	
QALYs per death averted:	*~ 4.5 QALY*	*~ 3.8 QALY*	*~ 12.7 QALY*	
big 4 PBC's	£10,220	£12,338	£4,872	[1]
11 PBCs (with mortality)	£23,360	£28,045	£8,308	[2]
All 23 PBCs	£25,214	£30,270	**£18,317**	[3]
			Lower bound	
Effect of expenditure on mortality:	*Remainder of disease duration*	*Remainder of disease duration*	*Remainder of disease duration*	
YLL per death averted:	*~ 4.5 YLL*	*~ 4.5 YLL*	*~ 4.6 YLL*	
QALYs per death averted:	*~ 4.5 QALY*	*~ 3.8 QALY*	*~ 12.7 QALY*	
big 4 PBC's	£5,083	£5,811	£1,194	[4]
11 PBCs (with mortality)	£8,579	£9,861	£1,175	[5]
All 23 PBCs	£9,260	£10,644	**£2,832**	[6]
			Upper bound	
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per death averted:	*2 YLL*	*2 YLL*	*2 YLL*	
QALYs per death averted:	*~ 2 QALY*	*~ 1.4 QALY*	*~ 5.6 QALY*	
big 4 PBC's	£23,346	£26,138	£11,040	[7]
11 PBCs (with mortality)	£52,936	£59,151	£18,827	[8]
All 23 PBCs	£57,136	£63,844	**£41,507**	[9]

The estimate of £8,308 per QALY (column 3, line 2) is restricted to the effects of changes in expenditure in the 11 PBCs where outcome elasticities can be estimated. However, these PBCs only account for a proportion of a change in overall expenditure (approximately 35%, see Table 5.2 below). As was explained in Section 4.4.2 and 4.4.3; the QALY threshold of £18,317 (column 3, line 3) uses the estimated proportionate effects of expenditure on the QALY burden of disease in the 11 PBCs as a surrogate for proportionate effects in the others, (i.e., assuming that the effects that can be observed will be similar to those that cannot) and represents our central or best estimate. As in previous sections, no health effects are assigned to PBC23 or 22 (General Medical Services and Social Care) on the basis that any health effects of this expenditure would be recorded in the other PBCs.[94] Although this estimate of £18,317 reflects changes in undiscounted QALYs associated with changes in expenditure, discounting the quality adjusted life year effects only increases the cost per QALY threshold to £18,613.[95]

[93] The cost per life year adjusted for quality of life in column 2 can be interpreted as cost per QALY threshold conditional on the assumption that the quality of life with disease is zero (equivalent to death); effectively ignoring any effects on those who survive with disease.

[94] It would be inappropriate to assign all the change in GMS expenditure to the estimate of cost per QALY based only on the 11 PBCs with outcome elasticities because it would imply that GMS only contributes to these PBCs. Restricting attention to the 11 PBCs with outcome elasticities but allocating part of the change in GMS expenditure to them based on their proportional share of changes in overall expenditure would yield a slightly higher cost per QALY than reported in line 2. It should be noted that including changes in GMS expenditure but not assigning health effects to this PBC is likely to overestimate the threshold because any health effects associated with GMS or PBC 22, Social Care (see Footnote 48), will not be reflected in the estimated outcome elasticities of other PBCs unless the effects happen to be correlated with changes in expenditure in those PBCs.

[95] The effects of discounting are modest because: i) the health effects of a change in expenditure are restricted to one year (where no discounting is necessary); ii) most of the total QALY effect occurs in that year; iii) it is only some of the life year effects (adjusted for quality) of a change in mortality in that year that occur in future years that need

The upper and lower bounds for the cost per QALY thresholds in column 3 in Table 5.1 are based on making the necessary assumptions about duration of health effects of expenditure and how long a death might be averted optimistic (providing the lower bound for the threshold) or conservative (an upper bound for the threshold). The lower bound (lines 4 to 6) is based on assuming that the health effects of expenditure are not restricted to one year but apply to the whole of the remaining disease duration of the population at risk in PBCs during one year. Although this combines optimistic assumptions, it is possible that at least some part of a change in expenditure may prevent disease so will have an impact on populations that are incident to PBCs in the future. Such effects are not captured in any of the estimates presented in this report so all estimates are conservative in this respect (the possibility of a longer and more complex lag structure for the effects of expenditure are discussed in Section 5.8). The upper bound (lines 7 to 9) is based on the combination of assuming that health effects are restricted to one year for the population currently at risk and that any death averted is only averted for 2 years (see Section 4.2.5).

The estimated QALY effects associated with each PBC can be decomposed into that part due to life year effects adjusted for quality and that part associated with effects on quality of life during disease. The proportionate share of these different aspects of the total health effect are the same as reported in Table 4.19; where those PBCs for which mortality is the major concern have a much greater share of total QALY effects associated with avoidance of premature death (e.g., PBC2 and PBC10) compared to those where quality of life is the major concern (e.g., PBC 7).

5.3 Which PBCs matter most?

Which PBCs have the greatest influence on the overall threshold depends, to a large extent, on how a change in overall expenditure is allocated to the different PBCs (see column 1 in Table 5.2),[96] i.e., those that account for a greater share of the change in expenditure will tend to have the greater influence. However, it also depends on the proportionate effect of a change in PBC expenditure on the QALY burden associated with the PBC[97] and the scale of the QALY burden (for the population at risk) associated with the type of diseases that make up each PBC[98]. These determine the cost per QALY associated with each PBC (see column 4 below and Table C80 in Appendix C). The share, attributable to each PBC, of the total health effects of a change in overall expenditure (see column 2 of Table 5.2) is the combined effect of all of these. The proportionate impact on the overall cost per QALY threshold of a 10% change in PBC health effects in column 3 gives an indication of how sensitive the overall threshold is to the estimate of health effects associated with each PBC. It starts to suggest where further efforts to improve estimates of the overall threshold might be most usefully directed.

---

to be discounted; and iv) these need to be discounted only over 4.6 years on average (see Tables C89 and C90 in Appendix C for discounted values).

[96] Which is determined by the estimated expenditure elasticities (the proportionate change in PBC expenditure due to a change in overall expenditure) and total PBC expenditure (see Chapter 3 and section B11 in Appendix B)

[97] Which are determined by the outcome elasticities (the proportionate effects on mortality and YLL of a proportionate change in PBC expenditure (see Section 4.4.2 for details of how these estimates can be applied to measures of QALY burden in all PBCs).

[98] See Section 4.4 for how PBC level effects can be allocated to the contributing ICD codes and how measures of QALY burden for each ICD code can be established

Table 5.2: Impact of each PBC on the overall cost per QALY threshold (2008)

PBC		% Share of change in overall expenditure [1]	% Share of total health effects (QALY) [2]	Elasticity of the threshold* [3]	PBC cost per QALY [4]
2	Cancer	3.24	3.50	0.35	£16,997
10	Circulatory	5.50	14.32	1.43	£7,038
11	Respiratory	3.32	30.45	3.05	£1,998
13	Gastro-intestinal	2.32	5.83	0.58	£7,293
1	Infectious diseases	2.37	2.08	0.21	£20,829
4	Endocrine	1.37	8.04	0.80	£3,124
7	Neurological	4.33	14.48	1.45	£5,480
17	Genito-urinary	3.36	1.40	0.14	£43,813
16	Trauma & injuries*	5.58	0	0	NA
18+19	Maternity & neonates*	4.95	0.03	0.00	£2,969,208
3	Disorders of Blood	2.92	1.89	0.19	£28,305
5	Mental Health	25.32	9.31	0.93	£49,835
6	Learning Disability	1.47	0.34	0.03	£78,854
8	Problems of Vision	2.75	0.66	0.07	£76,850
9	Problems of Hearing	1.24	1.19	0.12	£19,070
12	Dental problems	4.09	1.34	0.13	£55,916
14	Skin	2.79	0.29	0.03	£174,775
15	Musculo skeletal	5.14	4.65	0.47	£20,254
20	Poisoning and AE	1.32	0.15	0.01	£163,766
21	Healthy Individuals	5.01	0.06	0.01	£1,483,012
22	Social Care Needs	4.26	0	0	NA
23	Other	7.35	0	0	NA

* The proportionate change in the overall cost per QALY threshold due to a 10% increase or decrease in the health effects associated with the PBC. These elasticities are correct up to a 50% change in health effects.

Although the 11 PBCs where outcome elasticities could be estimated only account for 36% of the change in overall expenditure they account for 80% of the overall health effects. Within this group some PBCs contribute more than others. For example, PBC11 (Respiratory) accounts for a greater share of total health effects and has a higher elasticity (3.05%) than PBC10 (Circulatory) even though the latter accounts for a greater part of a change in overall expenditure. The reason is that the cost per QALY associated with changes in expenditure in PBC11 is lower than PBC10 and much lower than the overall threshold (so generates more health effects for the same, or even smaller, change in expenditure).[99] The elasticities in column 3 are instructive, e.g., the elasticity for PBC11 suggests that even if the health effects of a change in expenditure in this PBC were over estimated by 30% the overall threshold would increase by 9.15% to £19,993. All other PBCs have much less influence in this respect. Nonetheless PBC10 is important compared to others as it does contribute a large share of total health effects and has one of the highest elasticities (1.43%).[100] Also PBC7 (Neurological), although accounting for a smaller share of a change in overall expenditure, does contribute a large share of total health effects with an elasticity of 1.45% and a relatively low cost per QALY associated with changes in PBC expenditure.[101]

[99] Within PBC11: Chronic lower respiratory diseases (J40-J47) accounts for 85% of the QALY effects of a change in PBC expenditure; Lung diseases due to external agents (J60-J70), 4%; Other diseases of upper respiratory tract (J30-J39), 4%; Other respiratory diseases principally affecting the interstitium (J80-J84), 1%; and Other diseases of pleura (J90-J94), 1%. The other ICD codes each contribute less but together account for 4% of the health effects of a change in PBC11 expenditure.

[100] Within PBC10: Ischemic heart diseases (I20-I25) accounts for 55% of the QALY effects of a change in PBC expenditure; Cerebrovascular diseases (I60-I69), 21%; Other forms of heart disease (I30-I52), 7%; Congenital malformations and deformations circulatory system (Q20-Q28), 3%; and Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified (I80-I89), 3%. The other ICD codes each contribute less but together account for 8% of the health effects of a change in PBC10 expenditure.

[101] Within PBC7 Episodic and paroxysmal disorders (G40-G47) accounts for 73% of the QALY effects of a change in PBC expenditure; Extrapyramidal and movement disorders (G20-G26), 8%; Other degenerative diseases of the nervous system (G30-G32), 5%; Other disorders of the nervous system (G90-G99), 3%; and Nerve, nerve root and plexus disorders (G50-G59), 2%. The other ICD codes each contribute less but together account for 9% of the health effects of a change in PBC7 expenditure.

The other 12 PBCs, where outcome elasticities could not be estimated account for the greater part of a change in overall expenditure (64%) but only 20% of the overall health effects, i.e., the cost per QALYs associated with a change in expenditure in these PBCs is, in general, much higher. Of course, we have not directly observed quality of life effects in these PBCs but inferred them from the proportionate effects that we can observe. Insofar as investment and disinvestment opportunities in these PBCs might have been more valuable (offered greater improvement in quality of life) than suggested by the implied PBC thresholds in column 4, the overall QALY effects will tend to be underestimated and the overall cost per QALY threshold will be overestimated.

The overall threshold of £18,317 maybe especially conservative (i.e., likely to be overestimated) with respect to health effects in PBC5 (Mental Health Disorders), which accounts for a large proportion of the change in overall expenditure (25%) and contributes most to the overall health effects (9%) compared to these other PBCs. The cost per QALY associated with this PBC (£49,835) is based on an extrapolation of estimated proportionate effects to a population based measures of QALY burden in this PBC, rather than observations of the direct impact of changes in expenditure on quality of life in the types of diseases that make up the PBC. Evidence that is available suggests that the investment and disinvestment opportunities in this PBC are likely to have been much more valuable than this implied cost per QALY. A review of the evidence of the cost effectiveness of the investment and disinvestment opportunities that have been available in mental health during this period is reported in Addendum 3 Appendix C. A search for evidence about interventions in those ICD codes that contribute most to the PBC (based on prevalence or the contribution to the variance in PBC costs), suggests that pharmacological, psychological and social interventions for depression are all more cost effective (in general much less than £10,000 per QALY) than the overall threshold and significantly more valuable than the implied QALY threshold for this PBC. Based on the contribution that each ICD makes to variance in PBC costs across PCTs, it is schizophrenia that contributes most. Although interventions that may have been invested or disinvested in schizophrenia are, in general, less cost effective (in general less than £24,000 per QALY) than those available for depression, they are still much more valuable than the implied cost per QALY of this PBC in Table 5.2.[102]

It is very important not to misinterpret the cost per QALY associated with each PBC in column 4 of Table 5.2. These are not cost effectiveness thresholds. That is, they do not represent the QALYs likely to be forgone due to costs imposed (e.g., by the approval of a new and more costly technology by NICE) in a particular PBC because NHS expenditure is not devolved and constrained to PBC specific budgets. Rather the overall expenditure is constrained through government decisions about public expenditure, but within the NHS resources (at the margin at least) can be reallocated in anything other than the very short run across different activities and disease areas. For example, the additional net NHS costs of approving a new but more costly technology in PBC10 (Circulatory) will not be restricted to the circulatory PBC (5.5% will, see column 1 in Table 5.2) but are likely to be reallocated in the same way as an equivalent reduction in overall expenditure (i.e., the shares of a change in overall expenditure in column 1).[103] Therefore, the relevant cost per QALY threshold for a technology in the Circulatory PBC is not £7,038 but the overall threshold of £18,317.

---

[102] Although the published evidence suggests that investment and disinvestment opportunities in this PBC tend to be much more valuable than the implied cost per QALY, we have little information on the particular investments and disinvestments that were actually made by PCTs. The review of local data sources (see Addendum 2, Appendix C) revealed very little routinely collected information about specific investments and disinvestments beyond more aggregate measures of spending. In common with other PBCs, there will inevitably be inefficient, ineffective or even iatrogenic practice (e.g., due to poor diagnosis and inappropriate prescribing). Insofar as these types of activities are sensitive to changes in PBC expenditure this will tend to increase the cost per QALY associated with changes in expenditure in this PBC. Whether both the extent of these inefficiencies and their sensitivity to changes in expenditure are sufficient to increase the cost per QALY above £49,835 is unclear, although it seems unlikely. Note that the effects of the scale and sensitivity to expenditure of inefficient or even harmful practice in the other PBCs where outcome equations could be specified are already captured in the estimated outcome elasticities.

[103] In principle at least, with sufficient panel data which would allow a more complex lag structure and simultaneous estimation of expenditure and outcome elasticities across all PBCs; it might be possible to isolate the short run effects of a change in expenditure in one PBC across all the others. In the absence of such data and so long as

The primary purpose of Table 5.2 is to identify which PBCs have greatest influence on the estimate of the overall threshold and examine whether the implied values for the other PBCs is likely to lead to under or over estimation. There are differences in the implied cost per QALY ratio between PBCs, including some with very high implied cost per QALY (e.g., PBCs 18, 19, 20 and 21 reflecting small estimated health effects in the denominator), although they have limited influence on the overall estimate of the threshold. These differences in the implied cost per QALY across PBCs should not be over interpreted. For example, these differences could be interpreted as evidence of a misallocation of resources (e.g., reallocating expenditure from PBCs with higher to lower cost per QALY would improve health) if the purpose of the NHS and PCTs is to maximise unweighted QALYs. However, rather than a misallocation these differences (between the first 11 PBCs) might indicate that the actual quality of life effects of expenditure are proportionally greater (lower) than mortality effects in those with higher (lower) cost per QALY, or that the health effects in these PBCs are more socially valuable with a greater implicit weight attached to QALYs gained or lost in these areas (e.g., maternity and neonates). The higher cost per QALY for the remaining PBCs may reflect that the actual quality of life effects of changes in expenditure maybe more than proportional to QALY burden (e.g., evidence from mental health PBC suggests that investment and disinvestment opportunities may have been more valuable that the implied PBC cost per QALY of £49,835). Also it was not possible to estimate the health effects of changes in PBC expenditure simultaneously across PBCs. Consequently the effects of changes in expenditure in one PBC may be recorded in ICD codes relevant to other PBCs, so it is possible that PBCs with higher implied cost per QALY may be contributing health effects to other (recipient) PBCs.[104]

Whether these differences are regarded as evidence of a misallocation or not, however, is unimportant for an estimate of a cost per QALY threshold that reflects the health effects of how changes in overall expenditure are currently expected to be allocated. Whether nor not PCTs do or should maximise QALYs has no influence on the current estimate of the threshold, given that NICE currently uses an unweighted QALY threshold.[105] Also, insofar as local objectives do change or national policy does reallocate expenditure, the impact of these and other changes that will take place over time will be reflected in estimates of the threshold in subsequent periods once these changes have taken place (see Section 5.6).

5.4 How uncertain are the estimates and what are the implications?

There are a number of sources of uncertainty which contribute to an assessment of how uncertain a central or best estimate of the cost per QALY threshold might be. There are three reasons why uncertainty in the estimate of the threshold might be of policy interest: i) the uncertainty in the parameters that determine the threshold might influence the mean or expected value of the threshold if they have a non linear relationship to the threshold or when they have a multi linear relationship but are correlated with each other; ii) the consequences of over or underestimating the threshold differ so the uncertainty may have an influence on the extent to which a policy threshold (a single value that can be

---

adjustments are expected take place quickly relative to the time horizon of the effects of the new technology on NHS cost and outcomes (i.e., marginal NHS resource can be reallocated in the medium term) using the overall cost per QALY threshold for technologies relevant to any PBC is reasonable and more so than other alternative assumptions that might be made.

[104] The health effects of a change in expenditure in a 'contributory' PBC will not be reflected in the estimated health effects of change in expenditure in the 'recipient' PBCs unless they happen to be correlated with changes in expenditure in the 'recipient' PBCs, i.e., all changes in expenditure are assigned to PBCs but all the health effects may not be. This suggests that the health effects are likely to be underestimated and the overall threshold underestimated.

[105] The quite general theoretical framework in Chapter 3 assumes that PCTs maximise some unspecified welfare function where health (not necessarily QALYs) is one of its arguments (see Section 3.3). The type of econometric analysis conducted would remain the same irrespective of the measure of health or weights that might be place on different type of health gained or lost. The assumption required is that mortality is related to the 'health' argument however 'health' might be specified. We make no comment on whether QALY maximisation ought to be the objective of PCTs, nor is that required to estimate a threshold for NICE which is currently based on cost per (unweighted) QALY.

compared to the incremental cost effectiveness ratio of a new technology) should differ from the mean or expected value of the central or best estimate; and iii) in conjunction with other methods of analysis[106][98] it can indicate the potential value of gathering more information to improve these estimates in the future. Of course, hypothesis testing and the traditional rules of inference associated with it, such as statistical significance, p-values and confidence intervals, have no relevance when making unavoidable decisions about policy relevant quantities based on information currently available and the best use thereof.[99]

An assessment of parameter uncertainty

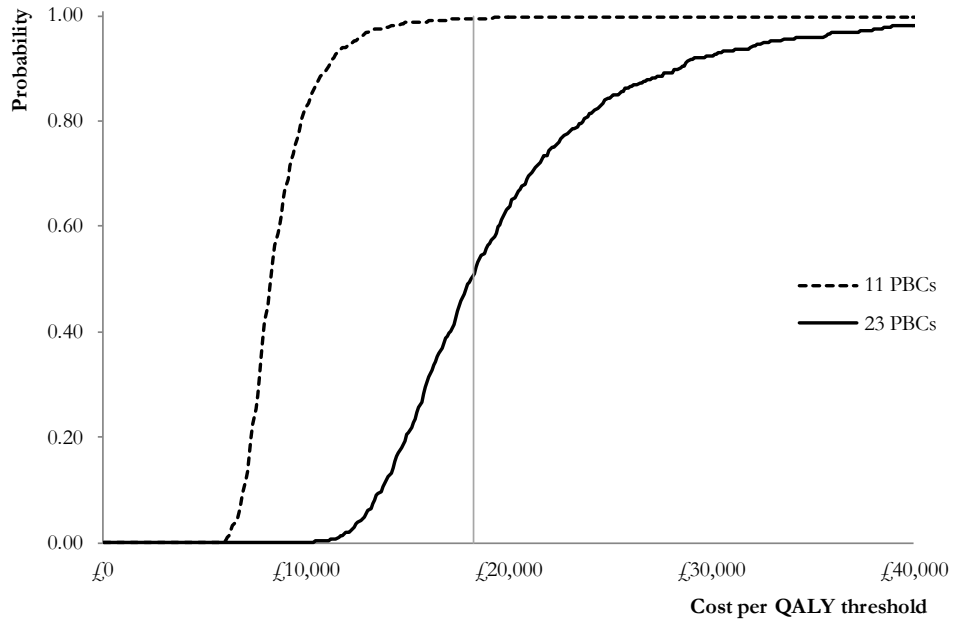
Two sets of parameters are critical to the threshold, the expenditure elasticities estimated for each of the 23 PBCs, and the outcome elasticities estimated for 11 of these. These parameters are estimated with uncertainty, indicated by the standard errors on the relevant coefficients in the econometric analysis outlined in Chapter 3 and detailed in Appendix B. Since these statistical models estimate coefficients using normality on the relevant scale, normal distributions can be assigned to each of these estimated coefficients, each with a mean and standard deviation based on the results of the econometric analysis.[107] These distributions, represent the uncertainty in the mean estimate of each of the parameters and can be propagated through the various calculations required to estimate and overall cost per QALY threshold (i.e., through the sequence of analysis detailed in Section 4.2 to 4.4) using Monte Carlo simulation which randomly samples from the assigned distributions. The results of each random sample represent one possible realisation of the overall threshold, given the uncertainty in estimates of the mean parameter values that determine it. By repeatedly sampling, a distribution of potential values that the overall threshold might take can be revealed. The results of this simulation are illustrated in Figure 4.3 which illustrates the cumulative probability density function for a cost per QALY threshold based only on the 11 PBCs with estimated outcome elasticities and for all 23 PBCs. It represents the probability (on the y-axis) that the threshold lies below a particular value.

It has already been noted that restricting attention only to changes in expenditure in those 11 PBCs where an outcome elasticity can be estimated results in a much lower estimate of the threshold than considering all changes in expenditure across all PBCs This lower estimate of £8,308 per QALY is much less uncertain but these PBCs only account for 36% of a change in overall expenditure, so it is the higher estimate, for all 23PBCs, that is of most relevance for policy (see Sections 4.4.3 and 5.2). The fact that this estimate is more uncertain simply reflects the quality and quantity of data currently available. Since useful analysis should endeavour to faithfully characterise uncertainty in policy relevant quantities, rather than select those quantities or questions for which precise estimates are possible, it is the more uncertain estimate for all 23 PBCs that should be of primarily interest. The values that are used to generate Figure 5.1 are available in Table C81 in Appendix C. They indicate that the probability that the overall threshold is less than £20,000 per QALY is 0.64 and the probability that is less than £30,000 is 0.92.

[106] A form of value of information analysis could be applied to these estimates in subsequent research, ideally capturing some of the other sources of uncertainty. Such analysis has firm foundations in statistical decision theory and has been applied to health care decisions. More recently it has been applied to the decisions faced by NICE when considering whether there is sufficient evidence to support the approval of a new technology

[107] The Monte Carlo simulation is in essence Bayesian, where the standard errors from the frequentist econometric analysis are used to assign normal prior distributions with means equal to the point estimates and a standard deviation equal to the estimated standard errors. This is equivalent to a fully Bayesian analysis with initially uninformative priors which are updated through the analysis of expenditure and mortality data.

Figure 5.1: Cumulative probability density function for the cost per QALY threshold



The implications of uncertainty

Integrating this parameter uncertainty into the estimates of the overall threshold does not change the mean or expected value of the cost per QALY threshold.[108] This is to be expected as the expenditure and outcome elasticities have a multi linear relationship to the overall threshold and the analysis sampled independently from the distributions assigned to estimated coefficient. We did investigate the potential correlation between the expenditure and outcome elasticities by repeatedly re-estimating both based on randomly sampling with replacement the 152 PCTs - creating bootstrapped data sets where the original PCTs could appear more than once or not at all in the re-sampled data. This analysis indicated a small positive correlation between outcome and expenditure elasticities in 4 PBCs using 2006 expenditure data (see Section B.10 in Appendix B). Such levels of correlation will tend to have a modest but positive influence on the mean value of the cost per QALY threshold.[109]

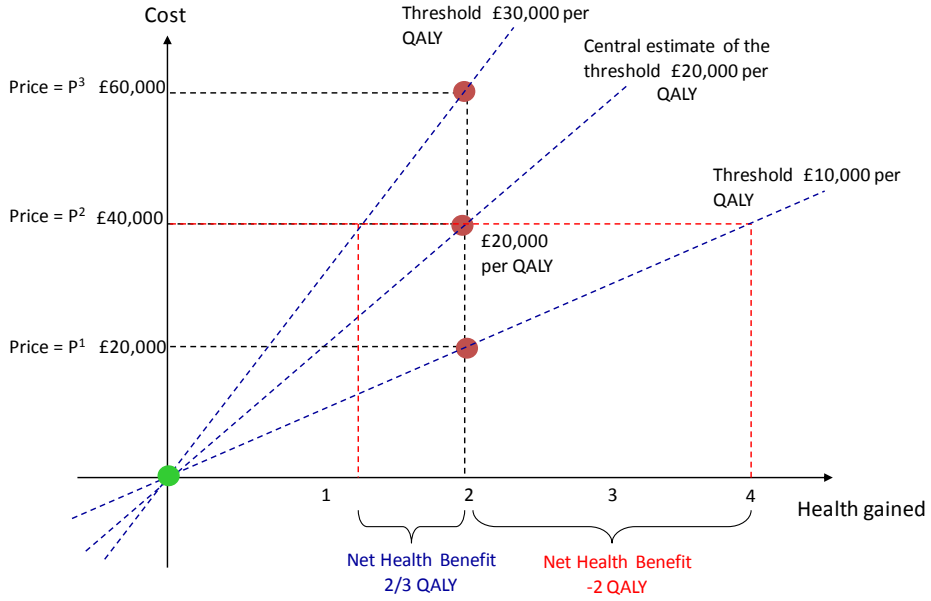
Uncertainty in the estimate of the overall threshold means that a policy threshold set at its mean or expected value may be inappropriate. Insofar as the consequences (to the NHS) of under or over estimation are symmetrical, then the expected or mean value would be the appropriate policy threshold irrespective of the scale of uncertainty. However, the consequences of overestimating the threshold are more serious than underestimating it. This is illustrated in Figure 5.2 which is similar to Figure 2.1 presented in Chapter 2.

---

[108] Note that the mean of the simulated values is not the mean of the sampled ratios but the ratio of the mean sampled values for the numerator and denominator. Deterministic and simulated values are the same for 2006, 2007 and 2008 expenditure data (other than negligible Monte Carlo error from 1000 samples). Also note that in constructing the cumulative probability density function in Figure 5.1 and the histograms of values in Appendix C it is important to identify whether sampled negative values favour a low value for the threshold or and unbounded one (there were no negative values sampled in the simulation of values for all 23 PBCs).

[109] Positive correlation suggests that a high spend elasticity will be associated with a high outcome elasticity (i.e., less negative, implying a smaller heath effect of a change in expenditure) resulting in a higher estimate of the threshold. It also suggests that when spend elasticity is low outcome elasticity will also tend to be lower (i.e., more negative, implying a larger health effect of a change in expenditure) resulting in a lower estimate of the threshold. Although realisations of spend elasticities higher and lower than the mean estimate are equally likely, higher spend elasticities provide a greater 'weight' associated with higher estimates of the threshold (where outcome elasticity is also high) when calculating the mean threshold. For these reasons a positive correlation will tend to increase the mean estimate of the threshold.

Figure 5.2: Consequences of over and underestimating the overall threshold



It shows the impact on net health benefit if the central estimate of £20,000 is in fact an overestimate and the threshold should be £10,000 per QALY. In these circumstances the technology should not have been approved at price P^2. This overestimation leads to a loss of net health benefit of 2 QALYs as a consequence. Alternatively, the central estimate of £20,000 may be an underestimate and the threshold should be £30,000 per QALY. In these circumstances the technology could just as easily have been rejected or approved based on the central estimate and price P^2. However, if the threshold is £30,000 per QALY rather than £20,000 it should be approved. If it was rejected this underestimation leads to a loss of net health benefit of 2/3 of a QALY as a consequence, i.e., less than the loss associated with the same scale of overestimation. If the scale of under or overestimation of the central estimate is equally likely (the distribution of possible values of the threshold is symmetrical) then using the mean or expected value as a policy threshold (one that can be compared to the incremental cost effectiveness ratio of a new technology) will lead to a loss of net benefit.[110] A policy threshold that represents the maximum the NHS can afford to pay for QALY gains offered by a technology will be lower than the mean of the cost per QALY threshold (i.e., lower than £18,317) to compensate for the more serious consequences of overestimating the 'true' value.[111] Importantly this remains the case even if effects are expressed in terms of their equivalent consumption value (net money benefit based on 'willingness to pay') rather than a measure of net health benefit.[112]

---

[110] Only a negative skew in the distribution of the threshold would tend to offset the implications of the non linear relationship between net health benefit and the value of the threshold. However, in this case the mean estimate is very similar but slightly greater than median values (see Section C.2.3.1 in Appendix C) indicating a small positive skew, which reinforces the implication that the policy threshold should be below the expected or mean value.

[111] Rather than solve for this type of 'certainty equivalent' a probabilistic analysis of the cost-effectiveness of a technology which integrated the uncertainty associated with the cost per QALY threshold as well, would take account of these issues, i.e., the technology would be cost-effective if it offered the highest expected net benefit when averaged over all Monte Carlo simulations, including sampling from the distribution of the cost per QALY threshold.

[112] Although health benefits can be expressed in terms of consumption (in money) using some consumption value of the health effects (willingness to pay), NHS costs must be first converted into health forgone, using an uncertain estimate of the threshold, before these are also expressed in consumption (money terms) using the same consumption value of health, i.e., the non linear effect of the threshold remains unavoidable. Failure to account for the threshold and the implications of its uncertainty would only be reasonable in a heath care system where expenditure was not constrained and/or all costs fell on private consumption.

How much lower a policy threshold should be set below the mean or expected value depends on three considerations: i) the scale of uncertainty in the estimate of the threshold (greater the uncertainty implies a lower policy threshold); ii) the scale of the incremental costs relative to incremental health benefits offered by the technology (policy threshold should only be equal to mean estimate if there are no additional NHS costs associated with the technology); and iii) the skewness of the distribution of cost per QALY threshold (a positive skew tends to offset these effects - see Figure C8 in Appendix C). The overall scale of the impact on a policy threshold will be specific to the additional NHS costs associated with a technology as well as the other sources of uncertainty discussed below and possible correlations between expenditure and outcome elasticities discussed above. We have not quantitatively integrated all these considerations in to an analysis of an appropriate policy threshold, although this maybe possible in future research.

Structural uncertainty

The uncertainty associated with the parameters estimated in the econometric models is only one, and not necessarily the most important, source of uncertainty associated with the cost per QALY threshold. The parameter uncertainty presented above is conditional on the econometric model being 'correct'. In particular, that the instruments used to identify the causal effect on health of changes in expenditure are valid. Although all the models passed the relevant tests of validity, there remains some uncertainty about the validity of the instruments used, i.e., there remains structural or model uncertainty (see Chapter 3 for an overview).[100] For this reason we undertook an analysis of how sensitive estimates of outcome elasticities might be to instrumental validity (see Section B9.4 in Appendix B). We were also able to specify a distribution for the measure of instrumental validity used in this sensitivity analysis, i.e., how 'likely' each value might be (see Section B9.5 in Appendix B). Therefore, there are two 'levels' of uncertainty: i) the parameter uncertainty (uncertainty in estimated coefficients given a particular 'level' of instrumental validity) and the structural uncertainty in the level of instrumental validity. Both sources of uncertainty were integrated by randomly sampling the distribution of measures of instrumental validity and then, conditional on this sampled value, re-estimating outcome equations and sampling the estimated coefficients. This analysis in Section B9.5 of Appendix B shows that model or structural uncertainty constitutes a greater part of the overall uncertainty associated with the outcome elasticities, so fully integrating this source of uncertainty is likely to have a significant impact on the extent to which a policy threshold should be lower than the mean or expected value of the cost per QALY threshold. Importantly, this additional structural uncertainty has little effect on the point estimates of the outcome elasticities, i.e., the central estimate of the cost per QALY threshold is robust to uncertainty in instrumental validity in the econometric models.

Other sources of uncertainty

Of course the parameter and structural uncertainty associated with the econometrics work outlined in Chapter 3 is itself only one source of uncertainty associated with the estimated cost per QALY threshold. Each of the steps of analysis in Section 4.2 to 4.4 explored the different ways routinely available data could be used and how additional information could improve the estimates. We identified a preferred analysis (or scenario) at each stage based on which made the best use of available information, whether the assumptions required appeared more reasonable than the other alternatives available, and which provided a more complete picture of the likely health effects of a change in expenditure. Insofar as the preferred analysis is the only plausible scenario, there would be no other sources of uncertainty. However, other assumptions and judgments are possible, which although they may be judged less credible might nonetheless have some probability of being the most credible (given evidence currently available). Therefore, there will be uncertainty between these alternative 'scenarios' as well as within each (the parameter and model uncertainty described above).[98] Although in principle this can be integrated into the analysis even in the absence of data to test alternative views[101]– we do not do so here since assigning probabilities to alternative scenarios would be somewhat speculative and inevitably disputed. Instead we offer a summary of the qualitative considerations. Of course any increase in the uncertainty associated with the central estimate of the cost per QALY will impact on the extent to which a policy threshold should be lower than the mean. However, a critical issue is whether consideration of other 'scenarios' might change this central estimate, e.g., if scenarios that lead to a lower estimate are judged

more credible than those that lead to higher ones. In other words the question is whether on balance the central or best estimate of £18,317 in Table 5.1 is likely to be an under or overestimate of the cost per QALY threshold.

Most of the considerations have been discussed in detail throughout Chapter 4 so are only briefly summarised here. The key assumptions made in Chapter 4 that underpin the central estimate of the cost per QALY threshold reported in Table 5.1 are briefly summarised in Box 1, including brief indication why such an assumption was required, the likely qualitative effect that each is likely to have on estimates for the health effects of changes in expenditure and where these are introduced in Chapter 4.

Box 5.1: Summary of assumptions and their likely impact on the central estimate of £18,317 per QALY

Assumption	Justification	Likely impact on estimates	Reference
1. Deaths averted by a change in expenditure returns an individual to the mortality risk of the general population (matched for age and gender).	No data to directly estimate the effects on survival. Appears more credible than restricting life year effects of changes in mortality to the period of observed variation in mortality outcomes.	Overestimate health effects	Sections 4.1 and 4.2.5 Footnotes 35, 58 and 127
2. Expenditure and outcome elasticities are uncorrelated.	Expenditure and outcome equations where estimated separately.	If the small but positive correlation between outcome and expenditure elasticises found in 4 PBCs was applied to all PBCs it is likely to have a modest but positive impact on the expected value of the threshold.	Section 5.6 Footnote 111
3. Mortality effects of changes in expenditure (reported at PCT level) can be applied to all mortality recorded in a PBC.	Assuming no health effects of expenditure in areas of disease where mortality is not recorded as PCT level or in over 75 age groups appears arbitrary and less plausible than basing estimates of effects that cannot be observed on what can.	Although including over 75 mortality may overestimate the effects on observed PBC deaths (if mortality in older ages groups is less sensitive to changes in expenditure) it has a much more limited impact on life year effects (i.e., including mortality above life expectancy reduces Net YLL)	Sections 4.2.1, 4.2.2, 4.2.3 and 4.2.4 Footnotes 36, 53, 54
4. The PBC QALY effects are a weighted average of effects within each of the ICDs that contribute to the PBC based on the proportion of the total PBC population within each contributing ICD codes.	PBC costs are not available at ICD level across PCTs. Although costs from HES data are available at ICD level they are only a small component of total PBC costs and contribute very little to the variability in PBC costs across PCTs especially when considering PBCs where mortality effects could not be estimated.	There is no information about how changes in PBC expenditure are allocated to particular ICD codes so the effect is unclear. However, it may overestimate health effects if investment within a PBC is focused on ICD codes where expenditure has greater health effects and disinvestment focuses on ICD codes with less health effects.	Sections 4.4.1 and 4.4.2 Footnotes 75, 81 and 125
5. Health effects of changes in expenditure are restricted to the population at risk during one year.	It was not possible to estimate a longer and more complex lag structure. Assuming that estimated health effects could be applied to the whole remaining duration of disease for the population at risk appears less plausible.	Underestimate health effects	Sections 4.1, 4.2.5, 4.4.3 and 5.8 Footnotes 35, 126 and 131
6. Health effects restricted to the PBC in which expenditure changes. No health effects associated with changes in GMS expenditure (or PBC22, Social Care).	It was not possible to estimate outcome equations for PBCs simultaneously so estimated outcome elasticities do not account for health effects due to changes in expenditure in other PBCs.	Likely to underestimate health effects because effects of changes in expenditure in 'contributory' PBCs will not be reflected in estimates of health effects in other (recipient) PBCs unless they happen to be correlated with changes in expenditure in these PBCs.	Sections 4.2.3, 4.2.5, 5.3 and 5.8 Footnotes 37, 48, 56, 93, 96, 106 and 134
7. Remaining change in total spend is assigned to the group PBCs where mortality effects could not be estimated	It was not possible to estimate expenditure equations for all 23 PBCs simultaneously so expenditure elasticities for all 23 PBC do not account for all of a change in overall spend.	May overestimate the threshold because a greater proportion of a change in expenditure tends to be allocated to PBCs with lower QALY burden and a higher implied PBC cost per QALY.	Sections 4.4.2 and 5.3 Footnotes 88 and 122

8. Same proportional effect on QALY burden of disease as the estimated proportional effect on the life year burden of disease.	Estimates of effects on mortality and life years are used as a surrogate for effects on quality of life. Appears more plausible than assuming no effects of NHS expenditure on quality of life outcomes	May under estimate the quality of life effects of changes in expenditure in these PBCs if effects are more than proportional to mortality and life year effects or over estimate them if they are less then proportional.	Sections 4.3.3 and 4.4.2 Footnote 66
9. Life year effects are lived at a quality of life that reflects a proportionate improvement to the quality of life with disease.	Consistent with using estimated mortality and life year effects as a surrogate for more complete measure of health outcome (QALYs). Appears more plausible than assigning quality of life norms or a quality of life with disease to life year effects.	This assumption is more conservative than assigning quality of life norms to life years (assuming that all disease is acute) but less conservative than assigning quality of life with disease (assuming that all life years would be lived in the diseased state until death)	Sections 4.3.3 and 4.4.2 Footnotes 66 and 77
10. Proportional effect on QALY burden of disease in PBCs where mortality effects could not be estimated is assumed to be the same as the overall proportional effect on the life year burden of disease across those PBCs where mortality effects could be estimated.	Consistent with using estimated mortality and life year effects as a surrogate for health effects (QALYs) where mortality effects cannot be directly estimated. Appears more plausible than assuming no health effects of NHS expenditure in these PBCs.	May under estimate the QALY effects of changes in expenditure in these PBCs if effects are more than proportional to QALY burden of disease. Other evidence suggests that the effect of this assumption maybe to underestimate health effects in key PBCs (Mental Health).	Sections 4.4.2, 4.4.3 and 5.3 Footnotes 84, 86, 92 and 123

On the one hand, there are some reasons why the health effects might be overestimated and the central estimate of the QALY threshold underestimated (e.g., see items 1 to 4 in Box 5.1). Calculating the life years lost that account for deaths that would have otherwise occurred as described in Section 4.2.3 and 4.2.4 is equivalent to assuming that those deaths averted by a change in expenditure returns the individuals to the mortality risk of the general population (matched for age and gender) Although this appears more credible than the alternative assumptions that could be made (e.g., restricting life year effects of changes in mortality to the period of observed variation in mortality outcomes), it is likely to be optimistic with respect to the life year effects of a changes in mortality, tending to underestimate the cost per QALY threshold.

On the other hand there are a number of reasons why the central estimate might be overestimated (e.g., see assumptions 5 to 7 in Box 5.1). The health effects of a change in expenditure are restricted to the population at risk during one year. This is undoubtedly pessimistic in three respects: i) it means that effect on quality of life during disease only occur for one year (the effect of investment that might have long term effects on quality of life, e.g., hip replacement are excluded); ii) mortality effects are also restricted to one year, so the full effect investments that reduce mortality for patients throughout their disease duration, not just in the first year, will not be captured; and iii) changes in expenditure that reduce incidence into the at risk population in the future (i.e., prevention of disease) will not be captured either. A more formal and longer lag structure in the estimation of outcome elasticities would be likely to capture more health effects of a change in expenditure.

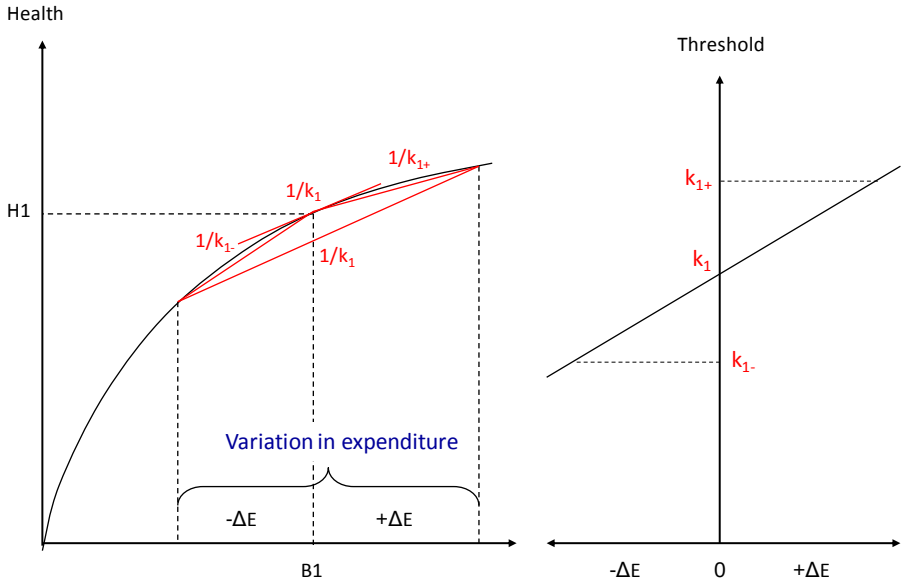
The effect of other assumptions that have been necessary are more ambiguous although some evidence suggests their net effect maybe conservative with respect to health effects of changes in expenditure (e.g., assumptions 8 to 10 in Box 5.1). The observed effects of a change in expenditure on mortality and life years in the 11 PBCs where outcome elasticities could be estimated was used as a surrogate for health effects in the other 12 PBCs (excluding GMS), i.e., the estimated effects of a change in expenditure that could be observed were used to inform those effects that currently, at least, cannot. This approach is not necessarily optimistic with respect to overall health effects. In fact there are good reasons to believe it may underestimate them (overestimate the threshold). As discussed previously in Sections 4.4.3 and 5.2; if this means of extrapolating from observed to unobserved effects is rejected then threshold estimate could be based only on the health effects of changes in expenditure in those PBCs where outcome elasticities can be estimated. This generates a much lower cost per QALY threshold (£8,308) even if that portion of GMS expenditure was allocated to these 11 PBCs (see Section 4.2.5). Alternatively, taking account of the greater proportion of the change in expenditure allocated to the other 12 PBCs but assuming that there are no health effects of expenditure in all these other PBCs is not plausible. In fact the evidence that is available about the value of investment and disinvestment opportunities in the most important of these other PBCs (PBC 7 Mental Health Disorders), suggests that the health effects of changes in expenditure in this PBC may be underestimated and the central estimate of the threshold overestimated (see Section 5.3 and Addendum 3 in Appendix C).

In addition, we have also shown that the uncertainty associated with our central estimate (from all sources) means that an appropriate policy threshold is likely to be below its mean or expected value. Finally, in Section 5.5 we explore how the threshold is likely to differ when considering opportunities to make investments (i.e., an increase in overall expenditure, or cost saving accruing to the NHS) and when disinvestment is required (a reduction in overall expenditure or costs imposed on the NHS). This analysis shows that a cost per QALY threshold relevant to technologies which impose costs on the NHS is likely to be less than our central estimate of £18,317. Therefore, although other assumptions and judgments are possible that retain some level of plausibility, they do not all favour a higher threshold. Indeed, when considered together, they suggest that on balance the central or best estimate of £18,317 presented in Table 5.1 is, if anything, likely to be an overestimate. In Section 5.8 we discuss how some of these remaining uncertainties might be resolved through access to additional and better data and the type of analysis that would then be possible.

5.5 Impact of investment, disinvestment and non marginal effects

The central estimate of the cost per QALY threshold in Table 5.1 is based estimates of the health effects of changes in expenditure across all 152 PCTs, some of which will be making investments (where expenditure is increasing) and others making disinvestments (where expenditure is reduced or growing more slowly). The cost per QALY threshold, however, is likely to differ across these different types of PCTs. This is illustrated in Figure 5.3 where the total observed variation in expenditure includes the impact of disinvestment (-ΔE), e.g., where costs are imposed on the NHS by the approval of a more costly technology; and investment (ΔE), e.g., where cost savings are accruing to the NHS. The central estimate of the cost per QALY threshold is the health effect of a change in expenditure across this variation in expenditure (k_1)[113]. One would expect that, other things equal, more expenditure (expanding the budget from B1) would increase health but at a diminishing rate. Therefore, the amount of health displaced by disinvestment, or a reduction in expenditure, would be expected to be greater, i.e., the threshold associated with -ΔE (k_1-) will be lower than the central estimate, k_1. Equally, the health gained from investments, or an increase in expenditure, would be expected to be lower, i.e., the threshold associated with ΔE (k_1+) will be higher than k_1.

Figure 5.3: Investment, disinvestment and budget impact



We have been able to examine this by re-estimating the outcome and expenditure elasticities separately for those PCTs where their actual budget is under the target allocation from the Department of Health resource allocation formula (i.e., those under greater financial pressure and more likely to be disinvesting than investing), and those that are over target (under less financial pressure and more likely to be investing than disinvesting). The detail of this analysis (based on 2006 expenditure and restricted to the 'big 4' PBCs) are reported in Section B8.9 in Appendix B. The results confirm what would be expected given Figure 5.3 and the discussion above - the outcome elasticities are smaller (in absolute terms) for all 4 PBCs in the group of PCTs above their target allocation and larger for all 4 PBCs in those below. Therefore, the health effects of changes in expenditure are greater in all these PBCs when PCTs are under more financial pressure and are more likely to be disinvesting then investing. The cost per life year estimates for these PBCs are reported in Appendix B: £10,604 for all PCTs combined (k_1); £8,441 for those PCTs under their target allocation (i.e., k_1- associated with –ΔE); and £14,083 for PCTs over their

[113] What can be estimated is the health effect over the observed variation in expenditure. This will also be the 'true' marginal effect (tangency at a budget of B1) if health returns to expenditure diminish at a constant rate (the second derivative is constant) as illustrated in Figure 5.3. Since nothing is 'truly' marginal the important question is how the threshold changes with the sign and scale of the non marginal budget impact associated with approval of a new technology.

target allocation (i.e., k_1+ associated with $+\Delta E$). Although these cost per life year estimates are not based on the same calculations as Section 4.2, they do start to indicate the scale of the effect on a threshold that is most relevant for new technologies that impose net costs on the NHS.

Expenditure elasticities for these PBCs also differ between these groups of PCTs - they are higher for those under their target allocation. These PBCs together consistently offer the greatest value in terms of cost per death averted, life year or QALY (see Table 5.1 and 5.2). This suggests that budget impact not only displaces more valuable activities within each PBC (outcome elasticities are larger) but that overall expenditure tends to be reallocated to more valuable PBCs. The effect of this reallocation on the overall threshold is not captured in the cost per life year estimate reported above, which are restricted to these 4 PBCs. Therefore, extending this type of analysis to all PBCs in future research is likely to show that the effect on the cost per QALY threshold of both the sign and scale of changes in overall expenditure will be greater. Subsequent work might enable a quantitative assessment of how the relevant threshold should be adjusted for the scale of the budget impact of technologies appraised by NICE.

Although further work is needed to fully specify the quantitative effect of the scale of non marginal impact of new technologies on an appropriate threshold, the qualitative impact seems clear. Firstly, the central estimate of the threshold is likely to be an overestimate for all technologies which impose net costs on the NHS (almost all technologies appraised by NICE have positive incremental NHS costs and all effective technologies that will be subject to value based pricing will impose net costs on the NHS).[4, 5, 16] Secondly the appropriate threshold to apply should be lower for technologies which have a greater impact on NHS costs.

5.6 How does the threshold change with overall expenditure?

The same methods of analysis can be applied to the econometric analysis of the 2007 expenditure and 2007 to 2009 mortality data (see Section B10 in Appendix B). This provides an opportunity to consider how the cost per QALY threshold is likely to have changed from 2007 to 2008 as overall expenditure has increased. This can provide some insights into how the threshold might be expected to change over time, as, for example, overall expenditure changes and productivity in the NHS might be expected to rise with innovation in health technologies, clinical practice and service delivery. This has implications for a judgement about the appropriate the frequency of periodic reassessment of the cost per QALY threshold.

It is not necessary the case that the threshold will rise with overall expenditure or even with NHS prices. This is illustrated in Figure 5.4 where the threshold at budget B1 is represented by k_1. If overall expenditure increases to B2 then, over things equal, the threshold would also be expected to increase (i.e., k_1 now overestimates the health effects of a change in expenditure at B2).[114] Increasing overall expenditure from B1 to B2 is equivalent to eliminating the same amount of waste in Figure 5.4, i.e., by re–allocating resources devoted to activities unproductive of health. Again, other things equal, the threshold would be expected to increase (k_1 now overestimates the health effects of a change in expenditure at B1) once the waste has been eliminated. However, insofar as the productivity of those activities that are valuable to the NHS also improve through innovation in health technologies, clinical practice and service delivery, the threshold will tend to fall. Figure 5.4 illustrates a situation where the effects of eliminating waste (NHS stopping doing things it should not be doing) and, at the same time, improving productivity (NHS getting better at doing things it should do) means that the overall threshold is unchanged.

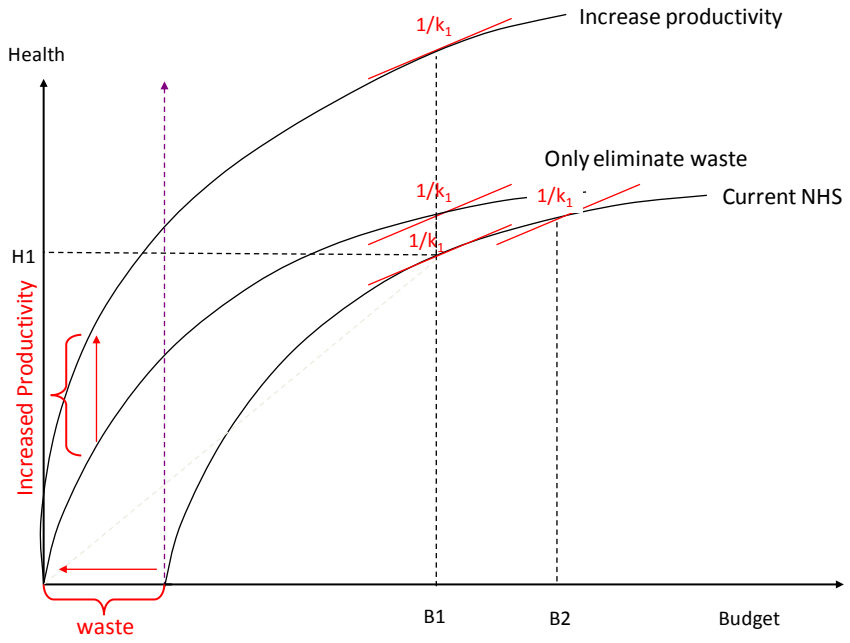
In making an assessment of whether the threshold is likely to increase with the NHS budget it is also necessary to consider whether there is discretion over how additional resources can be spent. For example, if any growth in the overall budget is spent on national initiatives or other activities that cannot or cannot easily be disinvested, then the additional costs of technologies approved by NICE must be accommodated by displacing other activities elsewhere. Therefore, it is growth in expenditure on more

[114] Due to the diminishing marginal returns illustrated in Figure 5.4 (see section 5.5 for further explanation).

'discretionary' parts of NHS expenditure and changes in the productivity and input prices of those health care activities which more likely to be displaced which are most relevant.

Over recent years much of the real budget growth in the UK NHS has been devoted to national initiatives that are not easily displaced, e.g. new contracts for General Practitioners and consultants, national waiting time targets, information technology initiatives, etc.[102] It also includes Technology Appraisal guidance issued by NICE itself, which has a funding mandate. Therefore, any real growth in what remains may have been more modest, so it is more likely to have been offset by any growth in the productivity of displaceable activities, e.g. drugs, devices, procedures and other services. Similarly, although there has been a general rise in input prices for the UK NHS, much of this inflation has been driven by staff as well as capital and overhead costs, some of which cannot be easily displaced. What are more relevant are the prices of inputs which could be displaced, an important element of which is drug prices. Although branded drug prices have tended to rise, at the same time there has been generic entry on patent expiry with dramatic reductions in prices for important classes of drugs.[103] Therefore, it is not self evident that the threshold has grown over recent years, despite real increases in the NHS budget.

Figure 5.4: Impact of changes in budget and productivity



The central estimates of the cost per QALY threshold for 2007 and 2008 expenditure years are reported in Table 5.3. In comparing these estimates of the QALY threshold it should be noted that important improvements were made to the classification and collection of PBC expenditure data that took place after the 2006 data were collected. Therefore, the differences in threshold estimates for 2006 and 2007 partly reflect this (see 3.5.4 in Chapter 3 and B11.4 in Appendix B) so should not be over interpreted. The results of the analysis of 2007 and 2008 expenditure are comparable in this respect.

Although overall expenditure increased by 6% between 2007 and 2008 which represented real growth of 2% in 2007 prices,[115] the overall threshold for all 23 PBCs fell by 2% in nominal terms and by 5% in real terms.

[115]2008 expenditure expressed in 2007 NHS prices based on 3.9% NHS inflation from the HCHS index – see Section B11.5 in Appendix B.

Table 5.3: Growth in the cost per QALY threshold (2007 to 2008)

	Cost per QALY threshold (2007) [1]	Cost per QALY threshold (2008) [2]	Nominal growth (%) [3]	Cost per QALY threshold (2008) 2007 NHS prices [4]	Real growth (%) [5]
big 4 PBC's	£4,549	£4,872	7%	£4,689	3%
11 PBCs (with mortality)	£8,513	£8,308	-2%	£7,996	-6%
All 23 PBCs	£18,624	£18,317	-2%	£17,629	-5%

The reasons are complex but reflect changes in productivity, which differs across PBCs (changes in outcome elasticities), but also a general reallocation of a change in overall expenditure (changes in expenditure elasticities) towards those PBCs that appear more valuable in 2008.[116] Given the sources of uncertainty described above, subtle differences between 2007 and 2008 should not be over interpreted. However, this analysis does suggest that the overall threshold will not necessary increase with growth in the real or even nominal NHS budget. In conjunction with the results of the analysis described in Section 5.4 it does suggest that the threshold is more likely to fall at a time when real budget growth is flat or falling and PCTs find themselves under increasing financial pressure.

Within the NICE Technology Appraisal process, the future incremental costs of a technology are expressed in real terms (at current prices) prior to discounting. Therefore, the estimates that are relevant to NICE decisions are: i) the nominal threshold in the current year[117] and ii) some assessment of the real growth in the threshold over the time horizon where incremental NHS costs are incurred. If there is an expectation of real growth (or fall) in the threshold over time then one way to incorporate this is through a higher (lower) discount rate applied to future cost.[104] Indeed, an expectation of changes in the real threshold over time also suggests something about the social rate of time preference heath revealed by budget allocations decisions.[105] However, incorporating an expected growth or decline in the threshold over time by adjusting discount rates is likely to be problematic once it is recognised that the expected incremental costs imposed by a technology are rarely uniform over time.

This discussion and the results reported in Table 5.3 suggest that there is little empirical support for an assumption that there will have been growth in the nominal threshold between 2008 and 2012.[118] Growth in the nominal or real threshold seems much less likely in the future with the prospect of reduced budget growth, increased pressures to improve productivity and downward pressure on input prices. Since how the nominal or real threshold is likely to change over time cannot be assumed to follow prices or overall expenditure nor empirical estimates or theoretical predictions of a growth in the private consumption value of health (willingness to pay), it becomes especially important to be able to regularly update estimates of the cost per QALY threshold based on routinely available data (see section 5.8).

5.7 What type of health forgone by approval of a new technology?

The methods of analysis described in Chapters 3 and 4 and discussed in this chapter can identify, not only how many QALYs are likely to be forgone across the NHS as a consequence of approving a technology which imposes incremental costs on the NHS, it can also indicate where those QALYs are likely to be forgone and how they are made up, i.e., the additional deaths, life years lost (unadjusted and adjusted for quality of life) and the quality of life impacts on those with disease.

For example, in 2011, NICE considered whether ranibizumab for the treatment of diabetic macular oedema should be approved for widespread use in the NHS (TA237).[106] Initially this technology was rejected by NICE on the grounds that, at its current price, it would be unlikely to be cost effective. In

---

[116] See Table C55 and C82 in Appendix C for a summary of outcome and expenditure elasticities and total expenditure by PBC in 2007 and 2008. Also compare Table C80 in Appendix C to Table 5.2 above for an indication of these net effects on the share of health effects and changes in expenditure.

[117] If the growth rate in the nominal threshold between 2007 and 2008 was applied the current 2012 threshold would be expected to be £16,895

[118] See above

2012, however, a rapid review of TA237 approved ranibizumab if use was restricted to the most cost effective sub group (those with central retinal thickness ≥400 micrometres) and after a Patient Access Scheme (PAS) for this subgroup of patients was offered (details of the PAS which provides a discounts to the NHS is commercial in confidence).[107]

The appraisal and guidance documents[106-108][119] provide the information required to estimate the additional NHS costs of treating this sub group of patients each year (see Addendum 4 to Appendix C for details of this example). Up to 44,000 NHS patients would be eligible for treatment with ranibizumab each year based on its licensed indication.[108] However, the subgroup of patients where ranibizumab was ultimately approved is likely to be 23,000 each year. This suggests that the approval of ranibizumab in this subgroup at the original appraisal price set in 2011 (i.e., without a PAS) would impose just over £80m of additional NHS costs for treating the eligible population each year.

Based on the 2008 central estimate of the cost per QALY threshold (£18,317 in Table 5.1) the approval of ranibizumab without a PAS would have been likely to displace 4,367 QALYs elsewhere in the NHS. However, the analysis which underpins the threshold estimate can also be used to identify where the additional NHS cost of £80m are likely to impact and where and what type of health effects are likely to be forgone. This is illustrated in Table 5.4. For example, the estimated expenditure elasticities and total PBC expenditure indicates how these costs will tend to effect spending in each of the 23 PBCs (see column 1).[120] The estimated outcome elasticities allow this change in spending in each PBC to be translated into a change in deaths and life year effects for the 11PBCs where mortality effects could be estimated (see columns 2 and 3). Applying the estimated proportional effect on the mortality burden of disease to measures of QALY (including the other PBCs) provides an estimate of the total QALY effect of the change in spend in each PBC (see Column 4).[121] The comparison of life year and total QALY effects allows the distinction to be made between QALY effects due the life year effects of additional deaths and QALY effects due only to quality of life (see column 5 and 6).

[119] All relevant documentation is available at http://guidance.nice.org.uk/TA237 and http://guidance.nice.org.uk/TA/Wave23/41

[120] The independently estimated expenditure elasticities for all 23 PBC do not account for all of a change in overall spend. The remaining change in total spend was assigned to the group PBCs where mortality effects could not be estimated. This will tend to overestimate the effect on spend in these PBCs and underestimate the effects on spending in the 11 PBC where mortality effects could be estimated (see Section 4.4.2 and Footnote 88).

[121] Although there was insufficient mortality available at PCT level to estimate outcome elasticities for the other PBCs, the measure of QALY burden in some of these PBCs does include some mortality (based on ONS data). Therefore, applying a proportionate effect to measures of QALY burden of will include some mortality and life year effects although they represent only a small proportion of the total QALY effects.

Table 5.4: Heath forgone across PBCs due to the approval of ranibizumab (£80m budget impact)

PBC	PBC description	change in spend (m) [1]	Additional Deaths [2]	Life years forgone [3]	QALYs forgone Total QALYs forgone [4]	QALYs forgone Due to premature death [5]	QALYs forgone Quality of life effects [6]
2	Cancer	£2.59	22	217	153	141	11
10	Circulatory problems	£4.40	132	672	625	427	198
11	Respiratory problems	£2.66	78	93	1,330	58	1,272
13	Gastro-intestinal	£1.86	15	143	255	94	161
	Big 4	£12	246	1,126	2,362	721	1,641
1	Infectious diseases	£1.89	4	31	91	21	70
4	Endocrine problems	£1.10	4	29	351	19	332
7	Neurological problems	£3.47	7	38	632	25	608
17	Genito-urinary problems	£2.69	13	19	61	12	49
16	Trauma & injuries	£4.46	0	0	0	0	0
18+19	Maternity & neonates	£3.96	0	2	1	1	0
	11 PBCs	£29	275	1,245	3,500	798	2,701
3	Disorders of Blood	£2.33	1	6	82	4	78
5	Mental Health Disorders	£20.25	12	55	406	35	371
6	Learning Disability	£1.18	1	4	15	3	12
8	Problems of Vision	£2.20	0	2	29	1	28
9	Problems of Hearing	£0.99	0	1	52	0	52
12	Dental problems	£3.27	0	0	59	0	59
14	Skin	£2.23	2	7	13	5	8
15	Musculo skeletal system	£4.11	3	15	203	10	193
20	Poisoning and AE	£1.05	0	2	6	1	5
21	Healthy Individuals	£4.01	0	1	3	0	2
22	Social Care Needs	£3.41	0	0	0	0	0
23	Other	£5.88	0	0	0	0	0
	All (23 PBCs)	**£80**	**295**	**1,337**	**4,367**	**859**	**3,509**

The results reported in Table 5.4 suggests that approval is likely to result in 295 additional deaths (most of which are likely to occur in Circulatory, Respiratory and Cancer PBCs – see column 2), and 1,337 life years forgone (most of which are likely to occur in Circulatory, Cancer and Gastro-intestinal PBCs – see column 3).[122] However, the impact of approval of this technology on QALYs forgone due to premature death (column 5) only accounts for a proportion of the total QALY effects (column 4). Most (3,509) are associated with quality of life forgone during disease (column 6). These quality of life impacts are most likely to occur in Respiratory, Neurological and Mental Health PBCs. The PBC level effects in Table 5.4 can also be examined at ICD level, whilst recognising the caveats discussed in Section 4.3 and 4.4.[123] For example within in the respiratory PBC it appears to be Chronic lower respiratory diseases (J40-J47) where most additional deaths, life years and quality of life are forgone. In the Mental Health PBC the additional deaths appear to be associated with disorders due to psychoactive substance use (F10-F19) and Mood (affective) disorders (F30-F39) (see Addendum 4 in Appendix C). However, it should be recognised that these effects which are based on the central estimate in Table 5.1 are likely to underestimate the health forgone given the discussion in Section 5.4 and especially in 5.5.

[122] The differences in contribution to deaths compared to life years reflects the distribution of age at death and the age and gender distribution of the population at risk in the ICD codes that contribute to each PBC (see Section 4.2 and addendum 1 in Appendix C).

[123] Information about the age, gender and the incidence of sequelae associated with different diseases within a PBC are only available for u-codes which can be mapped to groups of three digit ICD codes. Also allocating PBC level effects to ICD codes was based on the proportion of the total PBC population within each contributing ICD codes because PBC costs are not available at ICD level across PCTs. Although costs from HES data are available at ICD level they are only a small component of total PBC costs and contribute very little to the variability in PBC costs across PCTs especially in those PBCs where mortality effects could not be estimated (also see Footnote 75 and 81 and Addendum 1 in Appendix A).

The impact of a reduction in the price of this technology, either through value based pricing or the PAS that was offered during the rapid review,[107] can also be examined in the same way. The PAS was commercial in confidence but we will consider a scenario where a 30% reduction in NHS costs was applied for this subgroup of patients. Such a discount would be expected to save 1,310 QALYs including 89 deaths averted, 401 life years (258 when adjusted for quality) and quality of life effects during disease equivalent to 1,053 QALYs, when compared to approval of the technology at the original price (see Addendum 4 Appendix C for more details on this scenario analysis).

In many respects this starts to make 'real' the previously abstract notion that additional NHS costs are the health and opportunities of other unknown NHS patients. The methods of analysis presented in this report go some way to proving a empirically based and explicit quantification of the scale of opportunity costs the NHS faces when considering whether the health benefits associated with new technologies are expected to offset the health that is likely to be forgone elsewhere in the NHS. It also starts to make the other NHS patients, who ultimately bear the opportunity costs of such decisions, less abstract and more 'known' in social decisions. Since who happens to be known or unknown is only a matter of perspective, time and ignorance,[109] ethical and coherent social decisions require that both should be treated in the same way. The methods of analysis discussed in this chapter have contributed to removing some of the 'ignorance' and making the unknown more real.

5.8 Future research and improving estimates of the threshold

There are a number of ways in which this research could be usefully extended based on existing data and the information currently available, most of which have been discussed in previous sections of this chapter. Here we consider the scale of the evaluation problem in this context, examining what, in principle, would be required to resolve some of the key uncertainties discussed in Section 5.4, before a more detailed examination on of how additional routine data, greater access to existing data or data that are likely to become available might improve estimates of the cost per QALY threshold in the future. Two important questions remain when attempting to translate the estimated proportionate effects on mortality due to a change in expenditure into a more complete measure of the health effects (see Section 4.1 and 5.4). These are: i) whether the health effects of a change in expenditure in one year should be restricted to one year or extend over a longer period; and ii) the extent to which any death averted by expenditure in one year returns an individual to the mortality risk of the general population matched for age and gender. The central or 'best' estimates presented in Chapter 4 and 5 are based on combining the conservative assumption that the health effects of changes in one year of expenditure are restricted to one year[124] with the more optimistic assumption that any death averted by expenditure in one year returns the individual to the mortality risk of the general population.[125] The combination of assumptions that underpin the central estimates appear to be on balance conservative (see Box 5.1 and discussion in Section 5.4) and are certainty more credible than the implausibly pessimistic or optimistic assumptions that underpin the upper and lower bounds for the threshold that are also reported in Chapter 4 and 5. Key questions remain, however: why cannot routine data resolve some of these remaining uncertainties and what would be required to found a central estimate of the cost per life year or cost per QALY

[124] This is implicit in the estimates of outcome elasticities presented in Chapter 3. Although 3 years of mortality data are used in the analysis of each year of expenditure, these are averaged to an annual value prior to estimating outcome elasticities, so the estimated outcome elasticities represent the proportionate effect on mortality in one year due to a proportionate change in expenditure. This is likely to underestimate effects on mortality since expenditure that reduces mortality risk (or reduces the QALY burden of disease) for an individual in one year may well also reduce their risk (reduce QALY burden) over subsequent years; possibly over the whole of their remaining disease duration. Expenditure may also prevent disease in future patient populations. Therefore, total health effects will be underestimated and the cost per life year or QALY threshold will be overestimated.

[125] The years of life gained associated with each death averted are based on what would have been their life expectancy taking account of their age and gender (using life tables for the general population).

threshold only on econometric estimates rather than, in part at least, resting on judgments about the credibility of these alternative assumptions?[126]

A brief reiteration of the scale of this evaluation problem and the approaches to estimation that might be taken illustrates the quite profound difficulties and, therefore, the unavoidable need for explicit and accountable judgement and assumptions.[127]

A longer and more complex lag structure

Of course, a longer and more complex lag structure exploiting the PBC panel data set (i.e., both cross section and time series observations) that is becoming available over time, could in principle at least, identify the effect of a change in expenditure taking place in year *t* on health in years *t, t+1,.., t+n*. However, the health effects in subsequent years would need to be isolated from the effects of change in expenditure also occurring in subsequent years (which would also have both immediate and lagged health effects). The health effects of changes in expenditure in year *t* would need to be isolated from the lagged effects of changes in expenditure in previous years. Depending on the length of time series data available it may be possible to specify and estimate a richer empirical model to account for the lagged health effects of past expenditure and of lagged expenditure effects of past health outcomes.[128]

Although this is not a problem of principle it does pose difficulties as there are very real limits to the current time series because: i) there are a limited number of observations in the cross section (152 PCTs); ii) the definition (and boundaries) of PCTs has changed and has recently changed again with the formation of Clinical Commissioning Groups; iii) there are a limited number of years of observation in the time series (especially if lags are long); and iv) as noted in Chapter 3 and 5, the quality of PBC reporting has changed over time (recall that estimates from 2006 and 2007 PBC expenditure were not as comparable as 2007 and 2008). Nonetheless, as the panel data evolves over time there will be more opportunities to explore whether judgments about the duration of effects on mortality can be informed using the type of analysis presented in the report. Insofar as there are later lagged health effects this will tend to reduce the estimate of the cost per death averted and cost per life year and cost per QALY threshold.[129]

In many respects the problem of duration of mortality effects is a relatively straight-forward one compared to the second issue of how changes in mortality (whether immediate or lagged) translate into life years. In principle, estimating the effect of change in expenditure on life years is really estimating the effect of changes on expenditure on the survival curves of the population at risk through membership of ICD codes that contribute to each PBC. Even if the issue of lags is set aside, and attention is restricted to mortality effects in the expenditure year, translating these effects into life years would require observations on the entire survival curve of the at-risk population. This poses two profound difficulties i) we would need detailed information about the members of the at-risk population (patient identifiers) and ii) sufficient time to follow up the entire cohort from expenditure change to death (also accounting for

---

[126] It should be recognised that the purpose is to inform an assessment of the threshold for decisions that have not yet been made (i.e., prediction for decisions not yet made rather than a description of the past). Therefore, irrespective of the availability of evidence or the sophistication of analytic methods, the need for assumptions or scientific value judgements can never be avoided but only better informed.

[127] The nature of prediction to inform decisions and combined with the reality of a forever unobserved counterfactual makes judgement unavoidable - see footnote above.

[128] For example, a more structural approach of estimating an outcome equation jointly with an expenditure equation, both with appropriately specified lag structures and controlling for unobserved PCT effects might be possible although changes to PCT boundaries, recording of PBC data and the recent formation of Clinical Commissioning Groups makes the time series problematic.

[129] The health effects of previous changes in expenditure in *t-n* will not be reflected in estimates of the health effect of changes in expenditure in *t* unless they happen to be correlated with changes in expenditure in *t*. Therefore, excluding a longer lag structure for the health effects of changes in expenditure in *t-n,..t,..,t+n* is likely to underestimate the effects of changes in expenditure in *t*.

other changes that are likely to take place during that time). Even if these data were available and such heroic estimation was possible, any estimate would be so historic that it would be of limited policy relevance. This is not a problem unique to this research but remains a problem for all estimates of the years of life lost due to disease. It may be possible to use external, non-routine, historic data sources where patient identifiers are available to inform a judgement about whether changes in mortality in critical ICD codes (e.g., respiratory) tend to return patients to mortality risks similar to those of the general population or not. If historic evidence suggests that they remain at higher mortality risk it might indicate the likely scale of over-estimation if life year effects are based on the mortality risk of the general population. However, this would not be without major problems of distinguishing causality from selection effects.

The evolving panel data do have another advantage that could be exploited in the future. Currently it is only cross sectional variation (i.e., between PCTs) that contributes to the estimates of outcome elasticities. This means that changes in expenditure that all PCTs tend to make together, that might have very large health effects (they all tend to invest in obviously valuable activities at the same time) or limited health effects (they all disinvest in some activities that are not valuable at the same time) may not be fully reflected in the current estimates.[130] However, using variation in expenditure and outcome in both cross section and time series could more confidently pick up the full effects of simultaneous investment and disinvestment. The likely net effect on the overall threshold is unclear and will depend on whether PCTs tend to be more coordinated when investing in valuable activities (tending to reduce the threshold) or when disinvesting in ineffective ones (tending to increase the threshold).[131]

Simultaneous estimation across PBCs

Although expenditure equations are estimated for all 23 PBCs and outcome equations for the 11 PBCs where there are sufficient mortality data, these are estimated separately; each accounting for other PBC expenditure and other PBC need (see Section 3.3). The correlations between expenditure and outcome elasticities within each PBC were also estimated by repeatedly re-sampling the data set and re-estimating expenditure and outcome elasticities (see B8.11 in Appendix B). Although the estimate of the overall threshold accounts for changes in expenditure across all 23 PBCs with health effects estimated in 11 and inferred in the others, it is possible that changes in expenditure in one PBC may have health effects in others. Although total deaths across all 23 PBCs are accounted for, unless the possible 'external' mortality effects in other PBCs happen to be associated with variation in expenditure in those PBCs then these health effects will not be reflected in the estimated outcome elasticities. This seems likely to underestimate the total health effects of changes in expenditure unless positive health effects are thought to be offset by expenditure in one PBC damaging health outcome in others (e.g., adverse events associated with treatment or other iatrogenic effects).[132] To account properly for these possible effects would require estimating the interaction of changes in expenditure in each PBC on all the others while still accounting for possible endogeneity. Unfortunately, with only 152 observations in the cross section (PCTs), this type of simultaneous estimation is currently not feasible.

Throughout Chapters 3, 4 and 5 we have not imputed health effects for PBC 23 (General Medical Services) or procedural ICD codes on the grounds that the health effects of this type of expenditure will appear in ICD codes that contribute to other PBCs. However, the health effects of this type of expenditure (PBC23 and 22) will only be reflected in the estimated outcome elasticities insofar as the

---

[130] These effects will be picked up in the cross sectional variation, at least partially, so long as there is some variation in the health effects achieved and scale of simultaneous investment or disinvestment across PCTs.
[131] This would be particularly interesting when re-considering the subgroup analysis in Section 5.5 with panel data.
[132] We have taken account of competing risks or counterfactual deaths (which might appear in any of the PBCs in our calculation of net years of like lost - see Section 4.2.3). The health effects of a change in expenditure in a 'contributory' PBCs will not be reflected in the estimated health effects of change in expenditure in the 'recipient' PBCs unless they happen to be correlated with changes in expenditure in the 'recipient' PBCs, i.e., all changes in expenditure are assigned to PBCs but all the health effects may not be. This suggests that the health effects are likely to be underestimated and the overall threshold underestimated (see Section 5.3 and Footnote 106).

variation in outcomes reported in other PBCs, due to variation in GMS expenditure, happens to be associated with variation in expenditure in those other PBCs. Therefore our approach to GMS (and Social Care, PBC 22) expenditure is likely to be conservative with respect to overall health effects; tending to overestimate the cost per life year and cost per QALY threshold.

Exogenous shocks and quasi experiments

One response to these difficulties would be to look for exogenous budgetary shocks to the whole health care system and then estimate the health effects of the shock at a macro level. In principle this is very attractive since it would avoid all the difficulties of endogeneity and identifying valid instruments, exploring sensitivity and structural uncertainty. If a complete measure of health outcome were available at a health system level it would also avoid much of the complexity of working at a PBC and ultimately at ICD level.

Unfortunately there are a number of difficulties. Although the NHS budget is set each year through an essentially political process (so each year's change in budget might be regarded as an exogenous shock), insofar as public expenditure decisions are to some extent influenced by public sector performance then these apparent 'shocks' are endogenous in a very similar way to PCT expenditure decisions about particular PBCs but just at a higher level of aggregation. However, even if some arbitrary exogenous change to overall expenditure could be identified there are other serious difficulties. There is no comprehensive measure of outcome relevant to all NHS activities currently reported. This has two implications: i) the mortality data that are available is only relevant to approximately 36% of a change in overall expenditure (see Section 5.3); and ii) how mortality translates into life years and QALY depends critically on where those effects occur (the ICD codes that contribute to each PBC). In addition there are very good reasons why one would expect covariates (especially measures on need) and instruments to differ between different programmes of care. For all these reasons this research has focused on using routinely available data at its lowest level of aggregation.

By doing so we not only provide an estimates of a threshold based on a more complete measure of health effects, we are also able to indicate what type of health is effected and where they are most likely to occur. This provides a means to update estimates of the threshold should other aspects of social value be applied to measures of health or other aspects of social value be included in the future (e.g., consumption and other public expenditure effects). For example any 'weights' that might be assigned to different types of QALY gains or consumption and other public expenditure effects associated with health effects and the patient characteristics associated with ICD codes (e.g., QALY burden, years of life lost or other patient characteristics, such as age and gender) can be included in the current framework and a threshold re-estimated for 'weighted' QALYs or, give an estimate of the consumption value of a QALY, a threshold benefit cost ratio that includes consumption as well as health effects.

Evolving Programme Budget Category Data

Each year offers another wave of PBC expenditure data which means that a potentially useful panel data set is developing. This offers some useful opportunities that have been described above. However, with only 152 PCTs in the cross section, there is a limit to how much of the remaining uncertainty might be resolved. The utility of this evolving panel will also be limited by the formation of Clinical Commissioning Groups (CCGs) rather than PCTs as an important locus of expenditure decisions. Changes in PCT boundaries and the formation of CCGs will make the time series problematic unless CCGs can be mapped to previous PCT boundaries. However, updating expenditure and outcome elasticities based on variation in expenditure and outcomes across CCG would be possible (it would provide more observations in the cross section) so long as PBC expenditure and mortality outcomes are reported at CCG level.

Of course it would also be useful to be able to observe PBC expenditure at a lower level of aggregation (ideally at ICD code) as this would avoid the assumption necessary to allocate PBC level effects to ICD codes based on either estimates of the size of the at risk population or the crude (unadjusted for

covariates) contribution to variance in PBC expenditure. Since the only expenditure data that are available by ICD (and therefore PBC) for each PCT are HES based estimates of cost, the relevance of measures of contribution to variance in PBC expenditure depends on what proportion of PBC costs are accounted for by HES. However, HES cost are only a small component of total PBC expenditure and contribute very little to the variability in PBC expenditure across PCTs especially when considering PBCs where mortality effects could not be estimated (see Section 4.4 and Footnotes 75 and 81). Greater disaggregation within PBCs would be particularly useful as the examination of information routinely collected by PCTs was not particularly helpful in identifying what investment and disinvestments within a PBC explain the variation in PBC expenditure across PCTs (see Addendum C2 in Appendix C).

Extending measures of health outcome

Currently the only routinely collected health outcome data that can be matched to expenditure by PBC category at PCT level is mortality. For this reason outcome equations could only be estimated for 11 of the 23 PBCs. As discussed in Sections 4.3 and 4.4, this represents only one aspect of health outcome and is not particularly relevant to many disease categories and much of the care that the NHS offers when the primary purpose is to improve health experience and quality of life rather than to extend survival. Therefore, the estimated proportionate effects of expenditure on the QALY burden of disease in these 11 PBCs) were used as a surrogate for proportionate effects in the others, i.e., assuming that the proportionate effects that can be observed will be similar to those that cannot (see Section 5.2).

Of course, with access to a more complete measure of health outcome, which is routinely reported at PCT level and that can also be associated with PBC expenditure, it would be possible to use the same econometric methods to estimate the health effects of a change in expenditure across all PBCs, rather than imputing them in those PBCs where mortality is not the most relevant measure of health outcome.

The English NHS Patient Reported Outcomes (PROMs) programme was introduced in 2009 and routinely collects self-reported health status of patients receiving surgery for four elective procedures: knee and hip replacement, groin hernia repair, and varicose vein surgery. The data that are collected include both condition specific questions (the Oxford Hip Score, Oxford Knee Score and the Aberdeen Varicose Vein score; no condition specific instrument is available for hernia) as well as the generic instrument, the EQ-5D (both the EQ-5D profile, and the patient's global assessment of their health, the EQ-VAS. Patient-level data from the PROMs programme are freely available and can be linked to the HES database which provides a potential link to PBCs. Standardised reports on the PROMs data, including the average (case-mix adjusted) performance of providers, are regularly published by the NHS Information Centre, currently on a quarterly basis. Although currently offering very limited coverage for our purposes, there are plans to extend the PROMs programme in the future, with work underway or being planned around the potential use of PROMs in a wide range of long term conditions, primary care, in cancer survivorship, cardiovascular services, muscular skeletal, and cosmetic surgery.

In Appendix B, Section B8.8 we demonstrate how the econometric methods set out in Chapter 3 can be extended to these other non-mortality based outcome measures. EQ-5D utility scores (pre and post an operative procedure) from the PROMs programme are used to generate a non-mortality-based outcome measure, which we use to estimate our outcome model. Although the Department of Health does not report the number of patients undergoing an eligible operation by commissioner (PCT) it was possible to use the HES dataset to obtain this information. Routine reporting of procedure or intervention by commissioner in the PROMs data-set would seem a simple but important and valuable extension, especially as data are extended to primary care where HES cannot be used to substitute for this omission.

With data for both the average health gain per operation and the number of operations, we were able to estimate 'the health gain per head of population' for hip and knee replacements as defined above. This estimated outcome elasticity can then be used as an outcome measure for changes in expenditure in the

'problems of the musculoskeletal system' programme (i.e., PBC15)[133]. However, translating the short term impact of an intervention on quality of life, which can be estimated from PROMs data, into an estimate of the longer term effects on quality of life remains problematic.

Table B8.26 in Appendix B reports the estimated outcome equation for PBC15 (musculoskeletal system) using the PROMs based outcome measure. The result is intuitively plausible; an increase in expenditure improves health outcomes but, for a given spend, more need reduces the gain. The diagnostic statistics suggest that expenditure is endogenous and that the instruments are valid. They also suggest that the instruments are relevant and there is no evidence that the instruments are weak. Therefore, it is feasible to extend our modelling approach beyond those programmes with mortality outcomes should PROMs be extended more widely. Insofar as PROMs can contribute to a more secure estimate of the overall cost per QALY threshold in the future, the sensitivity analysis discussed in Section 5.3 starts to indicate where this type of information might be most useful.

Musculoskeletal is an important PBC, accounting for over 5% of a change on overall expenditure and almost 5% of the change in health outcomes. However, of those PBCs without mortality outcomes, it is PBC5 (Mental Health) that is most critical (see Table 5.2 in Section 5.3).

Measures of anxiety and depression are already routinely collected before, during and at the end of interventions as part of Improving Access to Psychological Therapies (IAPT), which is an NHS programme rolling out services across England offering interventions approved by NICE for the treatment of depression and anxiety disorders. By March 2011 IAPT services were offered in 142 of 151 PCTs. A requirement of the programme is to complete the Patient Health Questionnaire-9 (PHQ9, a measure of depression)[110] and Generalised Anxiety Disorder Assessment 7 (GAD7, a measure of anxiety)[111]. Both of these disease-specific measures can be linked to SF20 and further work could, in principle, link these scores to EQ5D. This is a rich, valuable and evolving data-set which potentially provides much of the information required to extend the econometric modelling to the mental health PBC. The experience with PROMs data suggests that this would be feasible, and the analysis in Section 5.3 indicates that this could make a significant contribution to strengthening the assessment of the overall threshold. It would also contribute to an assessment of the cost effectiveness of this programme both nationally and by PCT, which would be of value in its own right. Unfortunately, despite the collection of these data for every patient encounter for a number of years, unlike PROMs, these data have not yet been made publically available.[134] Of course, the services offered by the IAPT programme do not account for all the variation in expenditure in the mental health PBC. Nevertheless, access to data that have been and continue to be collected by practitioners and NHS patients, could provide estimates of changes in mental health outcomes due to changes in some types of mental health expenditure, which would be a significant advance.[135]

Incidence and duration of disease

Section 4.2 sets out the series of steps required to translate mortality effects into life years while taking account of competing risks or counterfactual deaths. This analysis used ONS data on deaths by age and

---

[133] The vast majority of hip and knee replacements are for osteoarthritis which is included in PBC15

[134] These data have only been collated centrally since April 2012 despite IAPT sites collecting these data at individual patient encounters for many years. In April 2012, the IAPT data standard was approved by the NHS Information Standards Board as a nationally mandated data standard. Data is now collected centrally on a monthly basis from over 200 service locations. The first report on the quality of IAPT data was published in November 2012 but the quarterly IAPT data reports, which were scheduled to be released at the same time, do not appear to have been made available. There is a commitment to make the dataset publicly available during 2012/13, although the timing and details of what will be available (summaries or patient level data and whether it will include the waves of data collected since 2006) and who might have access (commissioners, service providers or independent researchers) remains unclear (see www.iapt.nhs.uk and www.ic.nhs.uk/iapt).

[135] Similar difficulties will arise, however, when translating the observed impact of a therapy on quality of life, before and immediately after the intervention, into longer term effects.

gender in the ICD codes that contribute to each PBC, as well as life expectancies by age and gender for the general population. Some information was also required about the age and gender distribution of the population at risk in the ICD codes that contribute to each PBC (see Table 4.4 and 4.5). In Section 4.2.3 this was based on age and gender distribution of estimates of incidence from the WHO Global Burden and Disease (GBD) study. The same information was also used in Section 4.3.1 to adjust life years for the quality of life norms of the general population by age and gender. In Section 4.4.2 the measures of QALY burden of disease also used information about the duration as well as incidence of disease from the same GBD study. These estimates, published in 2008, were based on 2004 UK data and proved to be the best available source of this type of information given the resources available for this research. However, the GBD study has recently been updated with the findings first publically presented in December 2012.[112] The methodology of the new study as well as sources of information used have been much improved and any subsequent research on the threshold could integrate these new and improved estimates.

However, the GBD study is not the only potential source of information about estimates of incidence of disease by age and gender and disease duration across all the ICD codes that contribute to the 23 PBCs. For example, the General Practice Research Database (GPRD) (recently renamed the Clinical Practice Research Datalink (CPRD)) contains over 3 million active patient records drawn from approximately 400 primary care practices in the UK. GPRD is jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA).The database has clinical and prescription data and can provide information to support pharmaco-vigilance (indication, utilization, and risk/benefit profiles of drugs) and formal pharmaco-epidemiologic studies, including information on demographics, medical symptoms, therapy (medicines, vaccines, devices), and treatment outcomes.

Although this research was not funded to purchase access to GPRD data, we were able to examine a sample which comprised of 22,313,086 rows/patient–ICD10 events (3 digit) representing 4,229,910 patients with data on diagnosis of diseases observed between 1 Jan 2006 and 24 June 2011 (see Addendum C1 in Appendix C). Although GPRD data could, in principle, provide the type of information required the difficulties faced and the interpretation of the sample of data in the form available to us meant that it was not directly useful. The particular problems faced included: i) read rather than ICD codes reported in the data set, although mapping is and was possible; ii) being able to identify when an episode of disease ended; iii) estimating duration of disease from the sample of data when observations were censored by the limited years of data available to us; and iv) confidently identifying incident patients in diseases of longer duration despite two years of wash out prior to extracting the sample. GPRD is quite clearly a rich and valuable data set. However, our experience suggests that, to make best use of these data, specialist knowledge and experience of these data is really needed as well as access to a much larger sample than we were able to acquire with the limited resources available. Therefore, although GPRD could well help to improve estimates of incidence by age and gender and duration of disease, it would require additional well-resourced research including excess to specialist expertise and experience with this particular data set.

Recommendations for research

The priorities for further research that may be feasible based on data which are, or will become, available can be summarised as follows:

1. Any growth in the nominal or real threshold cannot be assumed (see Sections 5.5 and 5.6), so it will be important to update estimates of the threshold with more recent and future waves of expenditure and mortality data.
2. If other aspects of social value are applied to health benefits of a new technology they must also be attached to the type of health that is likely to be forgone due to additional NHS costs. For example, the value based pricing scheme due to be introduced by the Government in 2014 may include some additional weight for health benefits in diseases which impose a large health burden and/or where there are wider social benefits for patients, their carers and the wider economy. The methods developed in this research can be extended to allow the same weights to be also

attached to the type of health that is forgone and estimate the wider social benefits that are likely to be lost when the NHS must accommodate the additional costs of new drugs.

3. We have demonstrated that these methods of analysis can be applied to quality of life data collected as part of PROMs. This type of analysis could be applied to these data in key PBCs as PROMs is rolled out providing some evidence about the quality of life effects of changes in PBC expenditure.
4. A key PBC is Mental Health. Currently outcomes data that could be linked to measures of quality of life are routinely collected in primary care. In principle the same methods of analysis can be applied to these data once they are made available providing some evidence about the quality of life effects of changes in mental health expenditure.
5. Improved and more recent estimates of incidence (by age and gender) and duration of disease will soon be available from the recently published updated Global Burden of Disease study. These data could be used when the threshold is re-estimated for later waves of expenditure data. Alternatively, estimates could be based on CPRD data. However, our experience suggests that utilising CPRD data would need research that is well resourced with access to specialist expertise and experience with this particular data set.
6. Estimating a more complex lag structure based on the evolving panel data would provide valuable evidence about the duration of the health effects of changes in expenditure. The recent release of census data for 2011 may allow a panel model to be estimated allowing better control for unobserved heterogeneity across PCTs as well as exploiting variation in outcomes, expenditure and other covariates over time. There are, however, significant challenges including the formation of Clinical Commissioning Groups (CCG) in 2013, which will make the time series problematic for waves of expenditure and outcomes after 2012.
7. If PBC expenditure and outcome data are available at CCG level (as well as covariates and suitable instruments), it might become possible to estimate outcome and expenditure equations simultaneously across PBCs. This would enable more of the likely health effects of changes in expenditure to be reflected in the analysis.

5.9 Conclusions and implications for practice

The methods of analysis presented here go some way to providing an empirically-based and explicit quantification of the scale of opportunity costs the NHS faces when considering whether the health benefits associated with new technologies are expected to offset the health that is likely to be forgone elsewhere in the NHS. As such it provides a basis for determining the appropriate threshold for NICE decisions as well as those made centrally by the NHS and Department of Health more generally.

Since 2004 NICE has used a threshold range of £20,000 to £30,000 per QALY. It has been widely recognised for many years that this range is not based on evidence. The central estimate of the cost per QALY threshold (£18,317 per QALY based on 2008 expenditure) suggests that the upper bound to this range is almost certainty too high and the lower bound is also likely to be an overestimate (see Section 5.2). For example, the analysis of the uncertainty associated with the estimated expenditure and outcome elasticities indicates that the chance the threshold is less than £20,000 per QALY is 64 per cent and the chance that it is less than £30,000 is 92 per cent (see Section 5.4).

The central estimate is based on identifying a preferred analysis at each stage based on the analysis that made the best use of available information, whether the assumptions required appeared more reasonable than the other alternatives available, and which provided a more complete picture of the likely health effects of a change in expenditure. Although other assumptions and judgments are possible that retain some level of plausibility, they do not all favour a higher threshold. Indeed, when considered together, they suggest that on balance the central estimate of £18,317 is, if anything, likely to be an overestimate (see Section 5.4).

Although there is substantial uncertainty associated with the estimate of the overall threshold (including parameter, structural and other sources of uncertainty), a policy threshold set at its mean or expected value may be inappropriate because the consequences for the NHS of overestimating the threshold are more serious than underestimating it (see Section 5.4). In principle, a policy threshold (a single value that

can be compared to an ICER) should be set below its mean value to take account of the non linear relationship between the threshold and the additional net health benefit offered by a technology. The analysis of PCTs that are under more or less financial pressure (above or below their target resource allocation) starts to indicate the quantitative effect of the scale of the non marginal impact of new technologies on an appropriate threshold (see Section 5.5). It suggests that the central estimate of the threshold is likely to be an overestimate for all technologies which impose net costs on the NHS (almost all technologies appraised by NICE have positive incremental NHS costs and all effective technologies that will be subject to value based pricing will impose net costs on the NHS) and that the threshold might be lower for technologies which have a greater impact on NHS costs.

The research found no evidence that the threshold had increased with real growth in the NHS budget or with NHS prices (2007 to 2008) (see Section 5.6). There is little empirical support for an assumption that there will have been growth in the nominal threshold between 2008 and 2012. Since how the nominal or real threshold is likely to change over time cannot be assumed to follow prices or overall expenditure nor empirical estimates or theoretical predictions of a growth in the private consumption value of health (willingness to pay), it becomes especially important to be able to regularly update estimates of the cost per QALY threshold based on routinely available data (see section 5.8).

The methods of analysis can identify, not only how many QALYs are likely to be forgone across the NHS as a consequence of approving a technology which imposes incremental costs on the NHS, it can also indicate where those QALYs are likely to be forgone and how they are made up, i.e., the additional deaths, life years lost (unadjusted and adjusted for quality of life) and the quality of life impacts on those with disease (see Section 5.7). In doing so the study starts to make the other NHS patients, who ultimately bear the opportunity costs of such decisions, less abstract and more 'known' in social decisions. Since who happens to be known or unknown is only a matter of perspective, time and ignorance, ethical and coherent social decisions require that both should be treated in the same way. These methods contribute to removing some of the 'ignorance' and making the unknown more real.

These methods also allow other aspects of health outcome to be incorporated in the estimate of the threshold. This has implications for the Government's proposals to move to a system of value-based pricing for new prescription pharmaceuticals, which may include some additional weight for health benefits in diseases which impose a large health burden and/or where there are wider social benefits for patients, their carers and the wider economy. The methods developed in this research will allow the same weights to be also attached to the type of health that is lost and estimate the wider social benefits that are likely to be lost when the NHS must accommodate the additional costs of new drugs.

The methods of analysis can be used as a framework for further empirical work as additional and more appropriate data emerge in the NHS (see Section 5.8). They also offer a basis for threshold estimation in other health care systems which face constraints on the growth of health care expenditure and use cost effectiveness analysis to inform resource allocation decisions.

References

1. NICE, *Guide to the methods of technological appraisal.* ref: N1618, 2008.
2. Health, D.o., *A New Value-Based Approach to the Pricing of Branded Medicines - a Consultation.* . London Department of Health, 2010.
3. Health, D.o., *A New Value-Based Approach to the Pricing of Branded Medicines: Government Response to Consultation.* London: Department of Health, 2011.
4. Claxton, K., et al., *Value based pricing for NHS drugs: an opportunity not to be missed?* British Medical Journal, 2008. **336**(7638): p. 251-254.
5. Claxton, K., M.J. Sculpher, and S. Carroll, *Value-based pricing for pharmaceuticals: Its role, specification and prospects in a newly devolved NHS.* CHE Research Paper 60. http://www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP60.pdf. York: Centre for Health Economics, University of York, 2011.
6. Culyer, A., et al., *Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence.* J Health Serv Res Policy, 2007. **12**(1): p. 56-8.
7. McCabe, C., K. Claxton, and A.J. Culyer, *The NICE cost-effectiveness threshold - What it is and what that means.* Pharmacoeconomics, 2008. **26**(9): p. 733-744.
8. Committee, H.o.C.H., *NICE: First report of the Health Committee 2007-2008. HC27-I.* London: Stationery Office, 2008.
9. NICE, *Guide to the methods of technological appraisal.* ref: N0514, 2004.
10. Devlin, N. and D. Parkin, *Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis.* Health Economics, 2004. **13**(5): p. 437-452.
11. Gafni, A. and S. Birch, *Guidelines for the adoption of new technologies - a prescription for uncontrolled growth in expenditures and how to avoid the problem.* Canadian Medical Association Journal, 1993. **148**(6): p. 913-917.
12. Williams, A., *What could be nicer than NICE?* OHE annual lecture (book available), 2004.
13. Drummond, M., et al., *Methods for the Economic Evaluation of Health Care Programmes. 3rd ed.* . Oxford University Press, 2005.
14. Birch, S. and A. Gafni, *Changing the problem to fit the solution: Johannesson and Weinstein's (mis) application of economics to real world problems.* J Health Econ, 1993. **12**(4): p. 469-76.
15. Johannesson, M. and M.C. Weinstein, *On the decision rules of cost-effectiveness analysis.* J Health Econ, 1993. **12**(4): p. 459-67.
16. Health, D.o., *A new value-based approach to the pricing of branded medicines: A consultation.* 2010.
17. NICE, *First report of the Health Committee 2007-2008. HC27-I.* London: Stationery Office, 2008.
18. Weinstein, M. and R. Zeckhauser, *Critical ratios and efficient allocation.* Journal of Public Economics, 1973. **2**: p. 147-157.
19. Stinnett, A.A. and A.D. Paltiel, *Mathematical programming for the efficient allocation of health care resources.* Journal of Health Economics, 1996. **15**(5): p. 641-653.
20. Epstein, D., et al., *Efficiency, equity and budgetary policies: informing decisions using mathematical programming.* Medical Decision Making, 2007. **27**: p. 128-37.
21. Abelson, P., *The value of life and health for public policy.* Economic Record, 2003. **79**: p. S2-S13.
22. Bobinac, A., et al., *Willingness to pay for a Quality-Adjusted Life-Year: The individual perspective.* Value in Health, 2010. **13**(8): p. 1046-1055.
23. Byrne, M.M., K. O'Malley, and M.E. Suarez-Almazor, *Willingness to pay per quality-adjusted life year in a study of knee osteoarthritis.* Medical Decision Making, 2005. **25**(6): p. 655-666.
24. Dolan, P., et al., *QALY maximisation and people's preferences: a methodological review of the literature.* Health Economics, 2004. **14**(2): p. 197-208.
25. Green, C. and K. Gerard, *Exploring the social value of health-care interventions: a stated preference discrete choice experiment.* Health Economics, 2009. **18**(8): p. 951-976.
26. Groot, W. and H.M. van den Brink, *The value of health.* Bmc Health Services Research, 2008. **8**.
27. Gyrd-Hansen, D., *Willingness to pay for a QALY.* Health Economics, 2003. **12**(12): p. 1049-1060.
28. Gyrd-Hansen, D., *Willingness to pay for a QALY - theoretical and methodological issues.* Pharmacoeconomics, 2005. **23**(5): p. 423-432.
29. Johnson, F.R. and M. Backhouse, *Eliciting stated preferences for health-technology adoption criteria using paired comparisons and recommendation judgments.* Value in Health, 2006. **9**(5): p. 303-311.

30. King, J.T., et al., *Willingness to pay for a quality-adjusted life year: Implications for societal health care resource allocation.* Medical Decision Making, 2005. **25**(6): p. 667-677.
31. Lieu, T.A., et al., *Willingness to pay for a QALY based on community member and patient preferences for temporary health states associated with Herpes Zoster.* Pharmacoeconomics, 2009. **27**(12): p. 1005-1016.
32. Mason, H., M. Jones-Lee, and C. Donaldson, *Modelling the monetary value of a qaly: a new approach based on uk data.* Health Economics, 2009. **18**(8): p. 933-950.
33. Luis Pinto-Prades, J., G. Loomes, and R. Brey, *Trying to estimate a monetary value for the QALY.* Journal of Health Economics, 2009. **28**(3): p. 553-562.
34. Shiroiwa, T., et al., *International survey on willingness-to-pay (wtp) for one additional qaly gained: what is the threshold of cost effectiveness?* Health Economics, 2010. **19**(4): p. 422-437.
35. Yaesoubi, R. and S.D. Roberts, *A game-theoretic framework for estimating a health purchaser's willingness-to-pay for health and for expansion.* Health Care Management Science, 2010. **13**(4): p. 358-377.
36. Polsky, D., *Does willingness to pay per quality-adjusted life year bring us closer to a useful decision rule for cost-effectiveness analysis?* Medical Decision Making, 2005. **25**(6): p. 605-606.
37. Smith, R.D. and J. Richardson, *Can we estimate the 'social' value of a QALY? Four core issues to resolve.* Health Policy, 2005. **74**(1): p. 77-84.
38. Hirth, R.A., et al., *Willingness to pay for a quality-adjusted life year: In search of a standard.* Medical Decision Making, 2000. **20**(3): p. 332-342.
39. Haninger, K. and J. Hammitt, *Willingness to pay for Quality-Adjusted Life Years: empirical inconsistency between cost-effectiveness analysis and economic welfare theory.* OECD, 2006.
40. Baker, R., et al., *Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project.* Health Technology Assessment, 2010. **14**(27): p. 1-+.
41. Rawlins, M.D. and A.J. Culyer, *National Institute for Clinical Excellence and its value judgments.* British Medical Journal, 2004. **329**(7459): p. 224-227.
42. Rawlins, M.D., D. Barnett, and A. Stevens, *Pharmacoeconomics: NICE's approach to decision-making.* British Journal of Clinical Pharmacology, 2010. **70**(3): p. 346-249.
43. Tappenden, P., et al., *A stated preference binary choice experiment to explore NICE decision making.* Pharmacoeconomics, 2007. **25**(8): p. 685-693.
44. NICE, *Appraising Life Extending End-of-Life Treatments.* London: NICE, 2009.
45. NICE, *Draft Guide to the Methods of Technology Appraisal.* London: NICE, 2012.
46. Committee, H.o.C.H., *NICE response to the first report of session 2007-2008. HC550.* London: Stationery Office, 2008.
47. Committee, H.o.C.H., *The government's response to the Health Select Committee's first report of session 2007-08 on the National Institute for Health and Clinical Excellence. Cm7331.* London: Stationery Office, 2008.
48. Braithwaite, R. and M. Roberts, *$50,000 per QALY: inertia, indifference, or irrationality?* presented at: Annual Meeting of the Society for Medical Decision Making, 2004.
49. Birch, S. and A. Gafni, *The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost-effectiveness threshold.* Journal of health services research & policy, 2006. **11**(1): p. 46-51.
50. Collier, J., *Parliamentary review asks NICE to do better still.* British Medical Journal, 2008. **336**(56-57).
51. Towse, A., *Should NICE's threshold range for cost per QALY be raised? Yes.* British Medical Journal, 2009. **338**.
52. Appleby, J., et al., *Searching for cost effectiveness thresholds in the NHS.* Health Policy, 2009. **91**(3): p. 239-245.
53. Hughes, D.A. and R.E. Ferner, *New drugs for old: disinvestment and NICE.* British Medical Journal, 2010. **340**.
54. Buxton, M., *How much are health-care systems prepared to pay to produce a QALY?* European Journal of Health Economic, 2005. **6**(4): p. 285-28.
55. Elshaug, A., et al., *Identifying existing health care services that do not provide value for money.* The Medical Journal of Australia, 2009. **190**(5): p. 269-73.
56. Laupacis, A., et al., *How attractive does a new technology have to be to warrant adoption and utilization - tentative guidelines for using clinical and economic evaluations.* Canadian Medical Association Journal, 1992. **146**(4): p. 473-481.
57. Birch, S. and A. Gafni, *Cost-effectiveness ratios - in a league of their own.* Health Policy, 1994. **28**(2): p. 133-141.

58. Drummond, M., G. Torrance, and J. Mason, *Cost-effectiveness league tables - more harm than good.* Social Science & Medicine, 1993. **37**(1): p. 33-40.
59. Appleby, J., N. Devlin, and D. Parkin, *NICE's cost effectiveness threshold - How high should it be?* British Medical Journal, 2007. **335**(7616): p. 358-359.
60. Martin, S., N. Rice, and P. Smith, *The link between health care spending and health outcomes: evidence from English programme budgeting data.* CHE Research Paper 24, 2007a.
61. Martin, S., N. Rice, and P. Smith, *Further evidence on the link between health care spending and health outcomes in England.* CHE Research Paper 32, 2007b.
62. Martin, S., N. Rice, and P. Smith, *The link between health care spending and health outcomes for the new English Primary Care Trusts.* CHE Research Paper 42, 2008b.
63. Martin, S., N. Rice, and P. Smith, *Does health care spending improve health outcomes?* Journal of Health Economics, 2008a: p. 826-842.
64. Martin, S. and P. Smith, *How good at commissioning health are English primary care trusts? A preliminary statistical analysis.* Report to the Health Foundation, 2009.
65. Martin, S., N. Rice, and P. Smith, *Panel data estimates of the link between health care spending and health outcomes for English Primary Care Trusts.* Mimeo, 2010.
66. Martin, S., N. Rice, and P. Smith, *Comparing costs and outcomes across programmes of health care.* Health Economics, 2012: p. 316-337.
67. Cochrane, A.L., A.S.S. Leger, and F. Moore, *Health service "input" and mortality "output" in developed countries (Reprinted from Journal of Epidemiology and Community Health vol 32, pg 200-205, 1968).* Journal of Epidemiology and Community Health, 1997. **51**(4): p. 344-348.
68. Young, F.W., *An explanation of the persistent doctor-mortality association.* Journal of Epidemiology and Community Health, 2001. **55**(2): p. 80-84.
69. St Leger, S., *The anomaly that finally went away?* Journal of Epidemiology and Community Health, 2001. **55**(2): p. 79-79.
70. Nolte, E. and M. McKee, *Does health care save lives?* The Nuffield Trust, London, 2004.
71. Gravelle, H.S.E. and M.E. Backhouse, *International cross-section analysis of the determination of mortality.* Social Science & Medicine, 1987. **25**(5): p. 427-441.
72. Cremieux, P.Y., P. Ouellette, and C. Pilon, *Health care spending as determinants of health outcomes.* Health Economics, 1999. **8**(7): p. 627-639.
73. Nixon, J. and P. Ulmann, *The relationship between health care expenditure and health outcomes. Evidence and caveats for a causal link.* The European journal of health economics : HEPAC : health economics in prevention and care, 2006. **7**(1): p. 7-18.
74. Bokhari, F.A., Y. Gai, and P. Gottret, *Government health expenditures and health outcomes.* Health Econ, 2007. **16**(3): p. 257-73.
75. Moreno-Serra, R. and P.C. Smith, *The Effects of Health Coverage on Population Outcomes: A Country-Level Panel Data analysis.* Results for Development Institute, Washington D.C., Working Paper, 2011.
76. Health, D.o., *NHS finance manual. December 2005 edition. See http://www.dh.gov.uk/assetRoot/04/13/18/26/04131826.pdf.* 2005a.
77. Smith, P.C., N. Rice, and R. Carr-Hill, *Capitation funding in the public sector.* Journal of the Royal Statistical Society Series a-Statistics in Society, 2001. **164**: p. 217-241.
78. Office, N.A., *Good governance report: review of programme budgeting.* London: NAO, 2008.
79. Appleby, J., Harrison, T., Foot, C., Smith, A. and Gilmour, S. , *Explaining variations in primary care trusts' spending on cancer services.* The King's Fund, London, 2011.
80. Lakhani, A., H. Olearnik, and D.e. Eayres, *Compendium of clinical and health indicators: data definitions and user guide for computer files.* London, NCHOD, 2006.
81. Baum, C.F., M.E. Schaffer, and S. Stillman, *ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. See http://ideas.repec.org/c/boc/bocode/s425401.html.* 2010.
82. Shea, J., *Instrumental relevance in multivariate linear models: a simple measure.* Review of Economics and Statistics, 1997. **79**: p. 348-352.
83. Stock, J.H. and M. Yogo, *Testing for weak instruments in linear IV regression.* NBER Technical Working Paper 284, 2002.
84. Ramsey, J.B., *Tests for specification errors in classical linear least-squares regression analysis.* Journal of the Royal Statistical Society Series B-Statistical Methodology, 1969. **31**(2): p. 350-&.

85. Pesaran, M.H. and L.W. Taylor, *Diagnostics for IV regressions.* Oxford Bulletin of Economics and Statistics, 1999. **61**(2): p. 255-+.
86. Durbin, J., *Errors in variables.* Review of the International Statistical Institute, 1954. **22**: p. 23-32.
87. Health, D.o., *PCT recurrent revenue allocations exposition book: 2009/10 and 2010/11.* Department of Health, London, 2009.
88. Health, D.o., *Recurrent resource allocations: 2006/07, 2007/08.* Department of Health, London, 2005c.
89. Health, D.o., *Personal communications.* 2012.
90. Curtis, L., *Unit costs of health and social care 2011.* PSSRU, University of Kent., 2011.
91. Conley, T.G., C.B. Hansen, and P.E. Rossi, *Plausibly exogenous.* Review of Economics and Statistics, 2012. **94**(1): p. 260-272.
92. Small, D.S., *Sensitivity analysis for instrumental variables regression with overidentifying restrictions.* Journal of the American Statistical Association, 2007. **102**(479): p. 1049-1058.
93. Dixon, J., et al., *A person based formula for allocating commissioning funds to general practices in England: development of a statistical model.* British Medical Journal, 2011. **343**.
94. Wailoo, A.D., S. Tosh, J, *The incorporation of health benefits in cost utility analysis using the eq-5d: report by the decision support unit.* School of Health and Related Research, University of Sheffield, 2010.
95. Dolan, P., et al., *A social tariff for EuroQol: results from a UK general population survey.* CHE discussion paper 138, University of York, 1995.
96. Currie, C.J., et al., *The routine collation of health outcomes data from hospital treated subjects in the Health Outcomes Data Repository (HODaR): descriptive analysis from the first 20,000 subjects.* Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 2005. **8**(5): p. 581-590.
97. Cohen, J.W., et al., *The Medical Expenditure Panel Survey: a national health information resource.* Inquiry, 1996. **33**(4): p. 373-89.
98. Claxton, K., et al., *Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development.* Health Technology Assessment, in press 2012.
99. Claxton, K., *The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies.* J Health Econ, 1999. **18**(3): p. 341-64.
100. Jackson, C., et al., *A Framework for Addressing Structural Uncertainty in Decision Models.* Medical Decision Making, 2011. **31**(4): p. 662-674.
101. Soares, M.O., et al., *Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration.* Stat Med, 2011. **30**(19): p. 2363-80.
102. Maynard, A. and A. Street, *Seven years of feast, seven years of famine: boom to bust in the NHS?* BMJ, 2006. **332**(7546): p. 906-908.
103. Trading, O.o.F., *The Pharmaceutical Price Regulation Scheme. An OFT market study. VBP Report 2007.* London: OFT, 2007.
104. Claxton, K., et al., *Discounting and decision making in the economic evaluation of health care technologies.* Health Economics, 2011. **20:2-15. DOI: 10.1002/hec.1612**.
105. Paulden, M. and K. Claxton, *Budget allocation and the revealed social rate of time preference for health.* Health Economics 2011. **24 MAR: DOI: 10.1002/hec.1730**.
106. NICE, *TA237: Ranibizumab for the treatment of diabetic macular oedema.* 2011.
107. NICE, *Macular oedema (diabetic) - ranibizumab (rapid review of TA237): appraisal consultation document.* 2012.
108. Novartis, *Single technology appraisal (STA) manufacturer submission: Lucentis ® (ranibizumab) for the treatment of visual impairment due to diabetic macular oedema (DMO).* 2010.
109. Broome, J., *Trying to value a life.* Journal of Public Economics, 1978. **9**: p. 91-100.
110. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure.* J Gen Intern Med, 2001. **16**(9): p. 606-13.
111. Spitzer, R.L., et al., *A brief measure for assessing generalized anxiety disorder: the GAD-7.* Arch Intern Med, 2006. **166**(10): p. 1092-7.
112. WHO, *Global Burden of Disease Study 2010.* The Lancet, 2012.

Appendix A

SYSTEMATIC REVIEW OF THE LITERATURE ON THE COST-EFFECTIVENESS THRESHOLD

Contents

A. Systematic review approach

A1. Introduction

The aim of the systematic review was to inform the development of the conceptual framework, as well as the design, implementation and interpretation of the empirical analyses. Rather than define a set of very specific questions to answer through the review, objective was to characterise the existing literature in terms of the questions addressed and approaches taken. However, it was hoped that insights would be provided on topics including:

- General conceptualisation of the cost effectiveness threshold
- How NICE's cost effectiveness threshold should be defined, characterised and operationalised
- Approaches to estimating cost effectiveness thresholds in general and the NICE threshold in particular

In the initial stages of this systematic review it became clear that the "traditional" method of conducting systematic searches of existing literature on the topic of the cost-effectiveness threshold would be insufficient to deal with the requirements of this particular study. Here we refer to the "traditional" method as the practice of finding key terms and Medical Subject Headings (MeSH) that most accurately capture the range of literature relevant to the topic, while attempting to include as few irrelevant studies as is possible (making use of programs such as Medline).

The main weaknesses of using such an approach for a systematic review of this topic is that it requires a pre-existing knowledge of the terms used and topics covered in the current literature. This process has always required a degree of expertise (and luck) as to the strategy taken, including both knowledge of the literature to find likely search terms and skill in the construction of the strategies. The implications of excluding a single key term are potentially equivalent to ignoring vast areas of the literature. In addition, the traditional approach relies on key terms existing that suitably encapsulate the relevant literature. Finding common terms used in literature with potential relevance to the cost effectiveness threshold was found to be a significant problem as many relevant topics were not specifically aimed at issues relating to the NICE cost-effective threshold (for example the Martin et al. Papers (1-3) which provide a precursor to this project). In addition, due to the wide range of coverage of topics such a "threshold" and "cost-effective", any attempts at a systematic review would be either excessively large or result in a clearly limited snap-shot of the existing literature.

As a result a pragmatic approach was taken to the identification of relevant papers, one of "pearl growing" which can be defined here as the use of existing collections of studies to identify additional relevant parts of the literature. The approach uses a pool of "initial pearls" to grow the literature both through references and citations until all relevant papers have been discovered. This approach therefore relies on the expertise of the authors of the exiting literature to populate the pool of studies rather than the searcher's potentially limited knowledge.

While this approach of "pearl growing" was significantly limited by the existing software available and has a time consuming element, it represents an approach that corrects for many of the failings of traditional searches for topics that share the characteristics of the cost-effectiveness threshold.

A2. Systematic Review Methods

The "pearl growing" method of systematic review can be characterised into five steps for the identification of relevant papers.

1. Identification and extraction of "initial pearls".
    - "Initial pearls" were indentified through consultation with researchers with experience of the cost-effectiveness threshold literature. Fourteen initial pearls were indentified through this process. These publications were chosen for their wide ranging coverage of the topic as well as their anticipated significance

2. Extraction of Citations and References from "initial pearls".
   - Citations: Web of Knowledge was selected to perform the citation searches. The reason for this selection was in part due to expert advice from an information specialist as well as brief and non-systematic investigations of citation results from a range of alternative software packages.
   - References: Web of Knowledge was also used for the collection of papers' references.
   - Both citations and references were exported into an EndNote library for the purpose of collection and further analysis (exclusion of repeats, title searching and review of the abstracts).

3. Identification of further "pearls" from cited and referenced papers.
   - Once citations and references of the "initial pearls" had been collected, they were subjected to a set of investigations to identify further "pearls".
   - Papers were excluded based on whether the titles or abstracts suggested the paper contained information on five topics of interest. These topics had been previously identified given the objectives of the project and from a review of the "initial pearls" and included papers were classified by whether they could inform,
     i) introduction to the cost-effectiveness threshold topic and policy context,
     ii) discussion and debate around the current value use of the threshold,
     iii) potential methods suggested to find a suitable threshold value,
     iv) specific values proposed,
     v) the use of individual and societal valuations of health gains to inform the value of the threshold

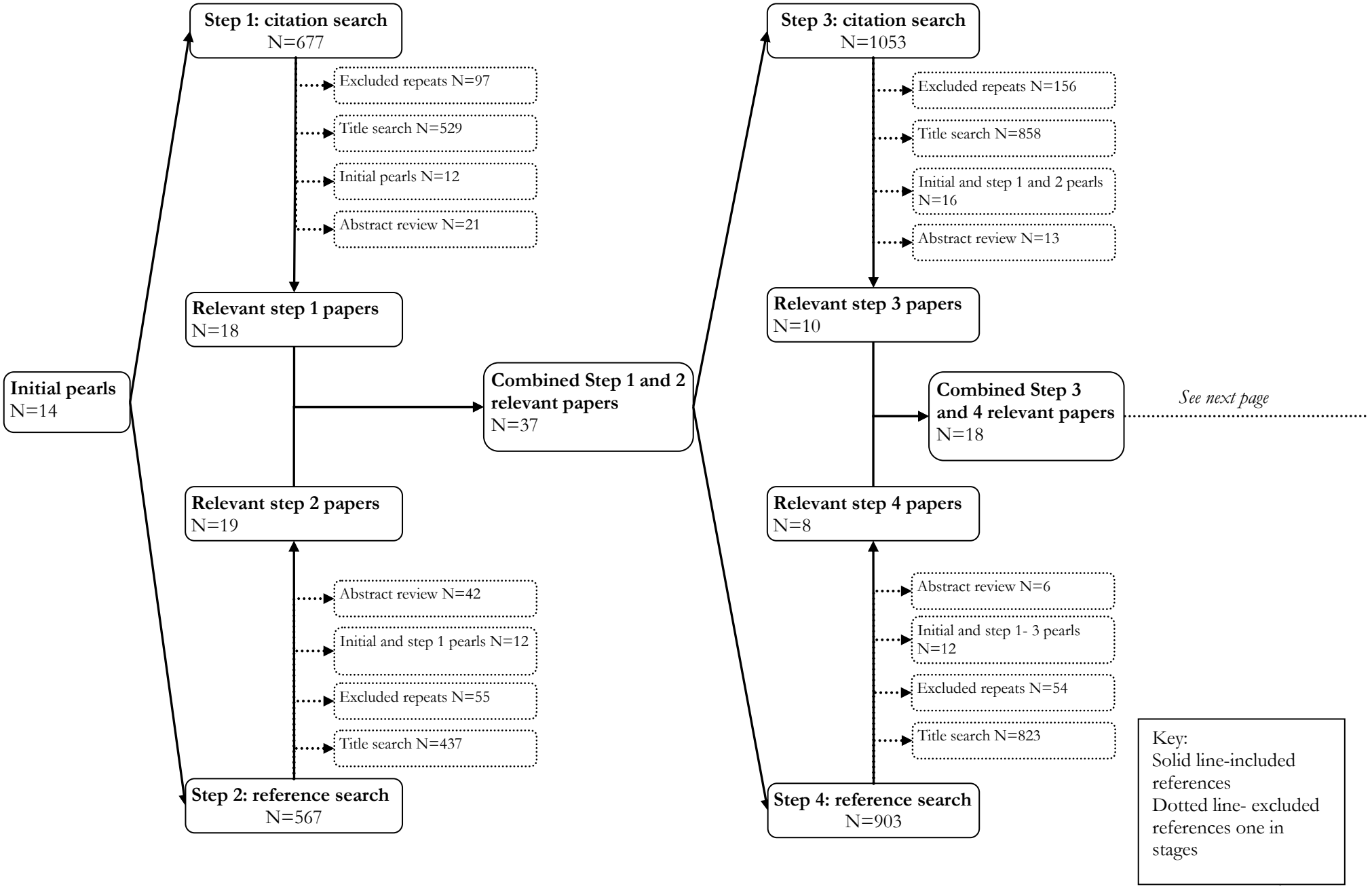
4. Repetition of citation and reference searches.
   - The process was then repeated for the "pearls" identified in step 3.
   - This process was repeated until no new "pearls" were discovered by additional iterations.

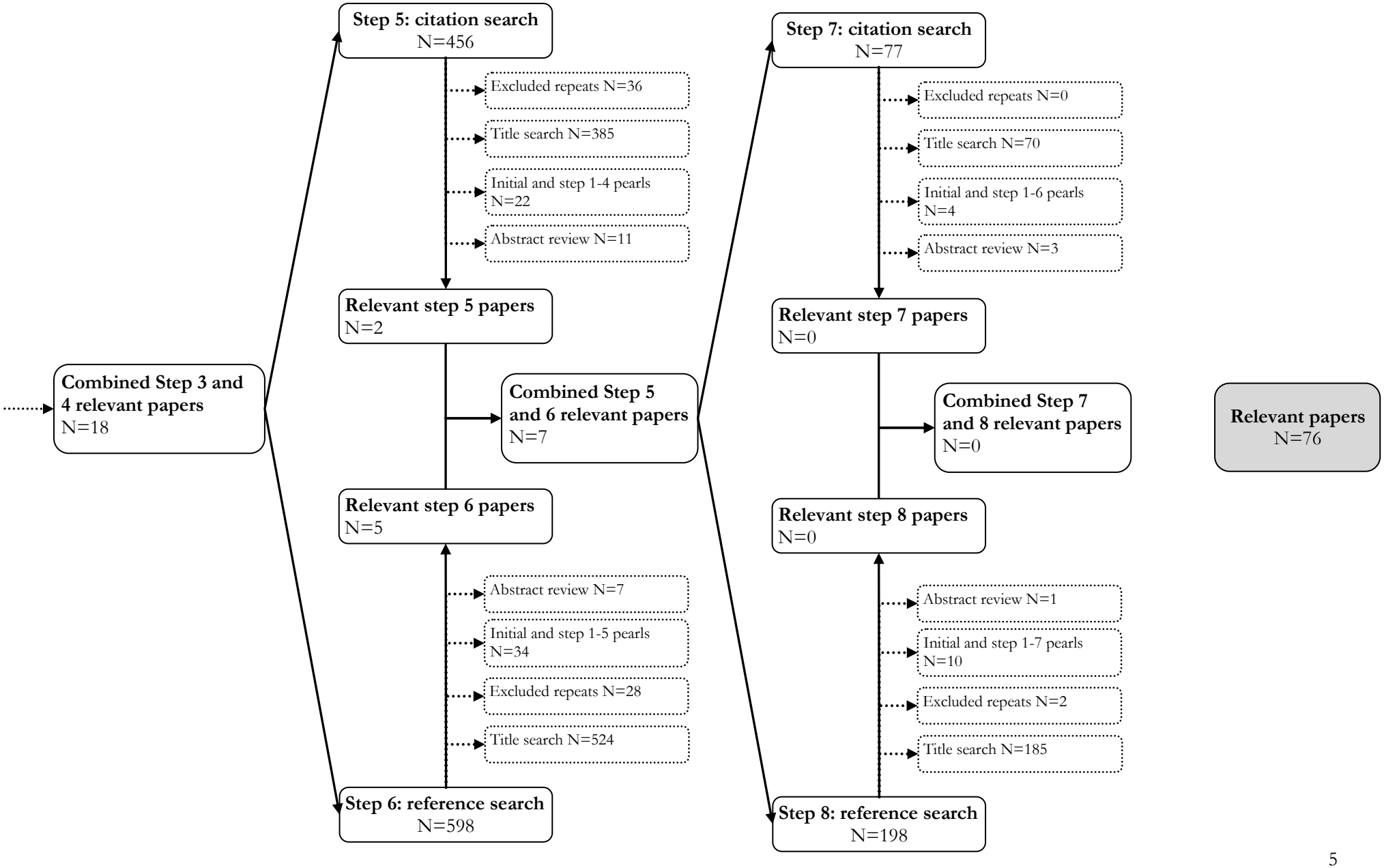
5. Manual search of references
   - To ensure as complete a search had been conducted as possible a retrospective manual search of all of the "pearls'" references was conducted. Any potentially relevant references not discovered previously (most likely due to a mix of user error and limitations with the software used) were added to the analysis at the relevant step and further pearl growing methods applied to them to ensure completeness of results.
   -

A3. Systematic review results

The "pearl growing" method of systematic review revealed 76 papers deemed relevant. The results from each stage of the process are reported in Figure A.1. The figure highlights that after four iterations no new relevant papers were identified by the systematic process.

Figure A.1: graph showing process results from pearl growing systematic review




Combined Step 3 and 4 relevant papers N=18
Step 5: citation search N=456
Excluded repeats N=36
Title search N=385
Initial and step 1-4 pearls N=22
Abstract review N=11
Relevant step 5 papers N=2
Combined Step 5 and 6 relevant papers N=7
Relevant step 6 papers N=5
Abstract review N=7
Initial and step 1-5 pearls N=34
Excluded repeats N=28
Title search N=524
Step 6: reference search N=598
Step 7: citation search N=77
Excluded repeats N=0
Title search N=70
Initial and step 1-6 pearls N=4
Abstract review N=3
Relevant step 7 papers N=0
Combined Step 7 and 8 relevant papers N=0
Relevant step 8 papers N=0
Abstract review N=1
Initial and step 1-7 pearls N=10
Excluded repeats N=2
Title search N=185
Step 8: reference search N=198
Relevant papers N=76

B. Review of literature

A4. Introduction and policy context

Due to the broad range of context which the relevant literature covers it is necessary to break down the literature review into several topics, these will be discussed independently. The 76 papers (see section E for all of these papers) identified by the systematic review were defined into five different categories:

1. literature covering the introduction to the cost-effectiveness threshold topic and policy context,
2. discussion and debate around the current value use of the threshold,
3. potential methods suggested to find a suitable threshold value,

These categories were chosen to reflect the broad range of relevant topics and areas of discussion covered by the cost-effectiveness threshold literature. It should be noted that the majority of the literature identified by the literature review fell into the first and last categories, with very few covering multiple categories sufficiently completely to be discussed in more than one section. The final category will only be discussed briefly as it can be seen as a separate, unrelated approach to the threshold required for purposes of decision making by the National Institute for Health and Clinical Excellence (NICE).

The majority of papers (34 of the 76 papers discovered) identified in the literature review could be characterised as introducing the idea of a cost-effectiveness threshold (these consist of the very early literature pre-dating NICE) or discussing the policy context through the years (4-37). This section will characterise the main areas of discussion in the literature and briefly describe the key parts of the literature development.

A4.1. Definition of the cost effectiveness threshold

An important place to start is the consideration of how the literature has defined the cost-effectiveness threshold. This is important to analyse in the review as not only is it worth ensuring that a good definition has been presented; but it also allows us to assess whether the existing literature uses a definition that is both consistent and accurate.

One of the earliest definitions of something resembling the modern interpretation of the cost-effective threshold comes from Weinstein and Zeckhauser(36). Their paper identifies a "critical ratio" between monetary costs and a measure of health gains. This critical ratio was argued to represent 'a cut-off point for allocation' of an activity in a budget constrained public sector entity ((36), *p.1*.).

A similar, more recent approach to define the threshold is that taken by Devlin (35) where the author considered a hypothetical budget constrained health care sector, with a perfectly informed decision maker who only considers the cost per QALY of health technologies. Assuming perfect information, the decision maker is able to rank all of the potential health care activities based on their cost per QALY. A decision maker will implement as many of the relatively low cost per QALY activities as possible until the budget is used up. Eventually a point will be reached where society is not willing to pay for a further marginal increase in QALYs and would rather the funding used on other consumption. The cost per QALY at which this cut-off occurs can be described as the cost-effectiveness threshold as it represents the switching point between an activity being funded and not. As the budget is assumed to be fully responsive, any new technologies with a cost per QALY below this threshold will be funded in the future.

A4.2. NICE and the cost effectiveness threshold

The use and valuation of a cost-effectiveness threshold by NICE has been controversial. Williams (37) highlighted three events that may be argued to have particularly muddied the water. Firstly, NICE did not set a threshold value by the government at the time of its inception in 1999. This meant that NICE was obliged to come up with a de novo estimate fairly rapidly. Through his set of discussions with NICE,

Williams stated that at the point of inception NICE came up with a value of 'roughly £30k per QALY, plus or minus £5k depending on the specific circumstances' *((37), p.7.)*

The second event which Williams refers to was NICE's initial resistance to acknowledging that any form of threshold value existed. Following analyses such as Towse and Pritchard (35) and Devlin(38) investigating previous NICE decisions and inferring an implicit threshold, NICE began to publish details of its approach to an ICER threshold. The major step was the 2004 Guide to the Methods of Technology Appraisal (30) that provided these details, although the definition of the £20,000 to £30,000 threshold range may be considered loose and open to interpretation. While the 2004 Guide was one of the first official references to the threshold, Sir Michael Rawlins did state at the 2001 NICE Annual General Meeting that the Institute would 'need to be very clear in its reasons for supporting technologies with cost-effectiveness ratios higher than £30,000 per QALY' (Littlejohns in (35)).

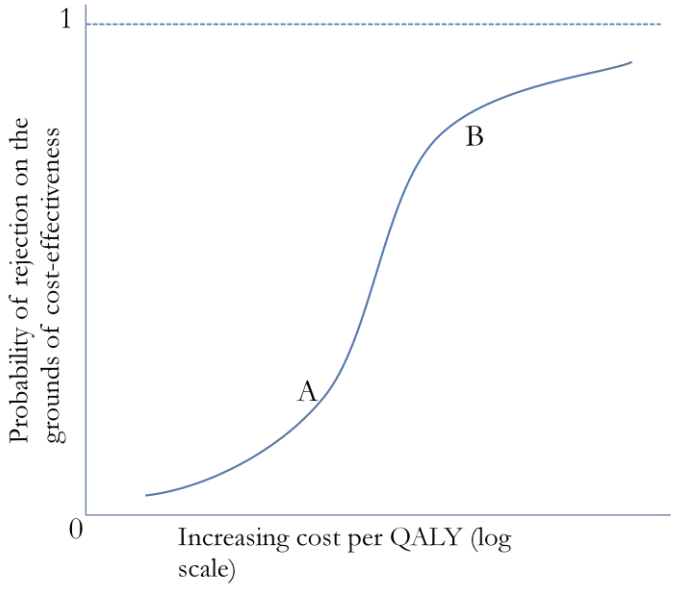
Williams' final event is the often quoted £20,000 to £30,000 threshold range having never been scientifically justified. Authors such as Rawlins and Culyer(39) have argued that there has never been an empirical basis for the values or any definitive meaning behind the range. They therefore argued that the threshold should not be the only tool for NICE to draw conclusions about new technologies.

A4.3. The threshold as a range

The idea of such a threshold range has been part of the literature for some time. Kaplan and Bush (24) considered the idea of a less abrupt approach than that suggested by Weinstein and Zeckhauser(36). Kaplan and Bush (24) investigated a set of early Medicare adoption decisions and presented broad criteria of acceptance based on a set of threshold ranges in terms of cost per additional well year. These were defined as <$20k/well year (cost effective), $20k-$100k (possibly controversial but justifiable), >$100k questionable when compared with other expenditure). However, the authors noted that a$100k cut-off was not relevant to the policy decisions at the time and that all results would need significant future investigation. Similarly Laupacis(26) presented five "grades of recommendation" for decisions about technological reimbursement in Canada.

The conclusions of both of these papers can be represented graphically by Figure A.2, which is also described or presented in much of the literature (see (39), Littlejohn in (35), (29), (38)). This graph represents the probability of rejection of a new technology as a function of technology's ICER. The graph clearly shows two points of inflection (A and B in Figure A.2), these two points represent an interpretation of the lower and upper bounds of a cost-effectiveness threshold range.

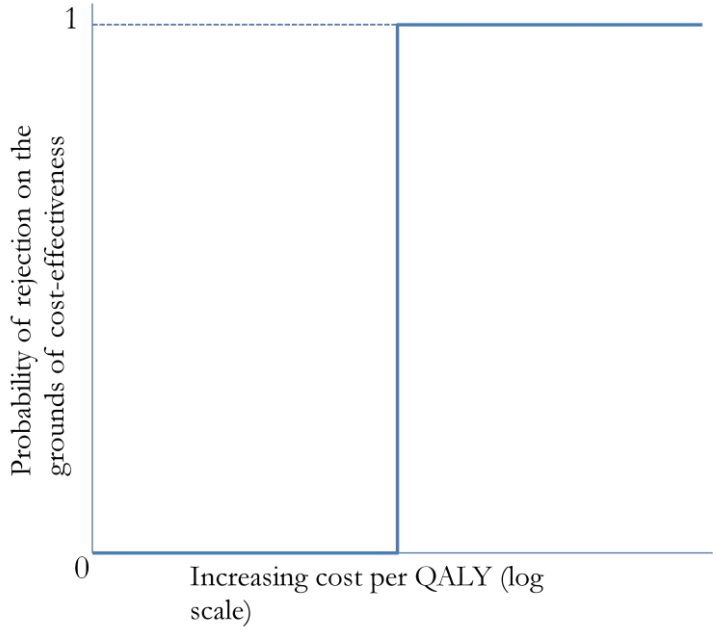
Figure A.2: Probability of rejection with a 'soft' cost-effectiveness threshold



The literature often makes use of the terms "soft" and "hard" when referring to the threshold. The term "soft" is often used in a similar way to the threshold "range" (alternatively Akehurst's "smudge" (Akehurst in (35)). While the underlying idea is the same, a "soft" threshold has also been used to refer to a single threshold. For example, McCabe et al. (29) argued that it is both feasible and probably desirable to use a single threshold rather than a range, as the threshold should represent the point beyond which factors other than cost effectiveness are considered. This approach would suggest that all new technologies with an ICER below the threshold should receive funding (regardless of their impact on other factors such as equity of health). It is, however, unclear from this paper what the implications are for technologies with an ICER beyond the single threshold value.

In contrast, a "hard" threshold represents the situation where the ICER valuation is the sole relevant variable in an adoption decision, as demonstrated in Figure A.3(15). It is an important point that if a "hard" threshold is set no other factors can be considered in the decision maker's consideration of a new technology. The difference between a "hard" and a "soft" threshold is therefore largely based on whether the ICER reflects all considerations. As, assuming the decision maker is optimising health, it should represent the most effective allocation of a health care budget but cannot account for any equity concerns (such as the severity of the condition, unmet need and orphan diseases) that are not included in the calculation of the ICER. Authors such as Dolan et al. (40) have demonstrated that a "hard" threshold may not be able to suitably reflect the non-linearity of social or political values of QALYs to factors such as quality and length of life and for those with worse health prospects or dependents.

Figure A.3: graph showing a "hard" cost-effectiveness threshold



A4.4. What does the threshold represent?

Two broad lines of thought have developed on what the threshold represents, social willingness to pay (WTP) and shadow pricing.(13, 29, 36, 37, 41, 42) The key difference between the two is the budget that should be considered by those accepting or rejecting health technologies. The social WTP approach (usually implicitly) assumes that the budget of the health care sector is flexible to the value of health gains determined by society. So in this case it is the value society places on the health benefits (for example in QALYs) generated by new health care programmes and technologies is estimated first, and then the health care budget is the sum of society's willingness to pay for all treatments. In other words, the threshold is set exogenously with no reference to a budget constraint.

In contrast, the shadow pricing approach takes the budget as given (at least beyond the control of those who determine the cost-effectiveness threshold) (13, 29). The threshold is, therefore, endogenous based on the services currently provided within the system. When a new programme or technology is accepted into the system and imposes an additional cost onto the budget, the only way to meet those costs is to remove or down-scale existing services which will incur *opportunity costs* in terms of population health. Hence the threshold represents the ICER of the least cost effective *existing* service covered by the budget. In principle, it is this service which is removed to fund a new programme or technology. In practice, a range of criteria is likely to be used to identify appropriate services for displacement to make room in the budget for new interventions.

In the UK the main source of debate about which of these concepts of the threshold is the correct one lies in NICE's remit. Authors such as Culyer (13) have discussed NICE's position as a "searcher" or a "setter" of the threshold. The distinction between these two roles is that a threshold "searcher" does not set a threshold with the motivation of maximising social welfare under the assumption of a flexible NHS budget, but instead investigates the threshold value that is appropriate given current NHS activities and the fixed budget as set down by Parliament.

Much of the literature on this topic is founded in the discussion of the correct constitutional role of NICE, the potential negative implications of setting a threshold and the feasibility of identifying displaced activities. In 2007, Culyer et al. (13) argued that it is not appropriate for NICE to be characterised as a threshold setter. The authors argued that the setting of a threshold would effectively imply that NICE sets the NHS budget. The setting of the NHS budget, they highlight, is the constitutional responsibility of

Parliament, not NICE. Hence the paper argues that NICE should concern themselves with being threshold "searchers", seeking to identify 'an optimal threshold ICER, at the ruling rate of expenditure, that is consistent with the aim of the health service to maximise population health' *((13), p.4)*.

In a similar vein Appleby et al. (43) concluded that the threshold used by NICE should be consistent with the decisions made by local commissioners within the NHS. This is important given that NICE provides little guidance to the NHS regarding interventions suitable for disinvestment to release the funding necessary to cover the new technologies it recommends. If the threshold is set too high NICE may well accept new technologies which are less cost effective than the services which local commissioners displace to fund those technologies. Conversely, if the threshold is set too low, NICE is likely to reject services that are cost effective relative to existing services delivered from the NHS budget. The authors conclude that, in the short term, NICE have to act as a threshold "searcher" to ensure continuity in the NHS.

Alternative arguments have been put forward which reject the idea of NICE as a threshold "searcher". Firstly, some authors (such as Gafni and Birch (17, 18)) have made the case that an implicit threshold has the potential to lead to spiralling inflation if new cost effective technologies are funded without sufficient disinvestment. However, McCabe et al.(29) argued that Culyer's characterisation of the NICE threshold could overcome this challenge if it were regularly reviewed so as to be flexible over time to changes in the NHS budget and the productivity of the sector, and if the threshold for new activities with a non-marginal budget impact was greater than those with a marginal impact. The issue of the inflationary pressure of a threshold is discussed further below.

Another concern raised about Culyer et al.'s characterisation of the NICE threshold is that of Towse et al. (44). They argue that a lack of knowledge of the true opportunity cost of new activities makes us unable to identify the value of those activities being displaced and, therefore, it is impossible for NICE to "search" for a threshold relating to activities displaced at the margin. The issue of the difficulty of identifying current activities at the margin in terms of cost-effectiveness will be dealt with later in this chapter.

A4.5. Factors considered by NICE other than the comparison of the ICER and threshold

As was discussed in the section the threshold range, the suitable threshold approach is dependent on the policy context around it, specifically if the comparison of the ICER with the threshold represents the only relevant piece of information that informs an adoption decision (a "hard" threshold) or if it is simply one of many factors considered ("soft" threshold). In the case of the UK, NICE has openly stated the ICER of a technology is not the sole consideration of the committee in its adoption decisions (30).

Both NICE and a number of other authors have provided overviews of the other factors that are considered by NICE in the adoption decision, these are provided in Table A.1.

Table A.1: table showing factors other than ICER considered by NICE

NICE (30, 31)	Rawlins et al. (45)	Tappenden et al. (34)	Devlin et al. (38)
Uncertainty of variables	Severity of illness	Uncertainty of the ICER	Uncertainty of the ICER
Availability of comparators	End of life treatment	Availability of comparators	Burden of disease
Clinical priorities (as set by Secretary of State)	Stakeholder opinion	Severity of illness	
Clinical need	Innovation	*age not significant	
Availability of resources	Population characteristics (disadvantaged and children)		
Innovation			
Disease characteristics and population size			
Wider social costs and benefits			
Length of benefit			

This table suggests that the threshold is only one consideration to decisions makers at NICE. However, , in principle, these other types of benefits could be added to health benefits and compared to potential treatments for displacement which also have wider social benefits. In other words, this wider set of considerations relating to the benefits of new technologies should arguably also be reflected in the threshold.[1]

A4.6. Multiple thresholds

Similarly some have argued for using different thresholds for different situations (29, 47). The two main cases for using different thresholds are the size of the budgetary impact, or depending on if the decision represents an investment in additional activities or a disinvestment in current activities.

The topic of different thresholds for different budgetary impacts of a proposed technology has received very little analytical attention from the literature. McCabe et al. (29) argue that technologies with a large budgetary impact should be evaluated against a lower threshold than those with a relatively small impact. The reason for this is a large budgetary impact will require a greater displacement of current activities (assuming a fixed overall budget); this may result in displacement of non-marginal activities which may be associated with a lower ICER than those at the margin.

Several authors have suggested the use of different threshold values depending on whether the decision represents an investment in additional activities or a disinvestment in current activities. O'Brien et al.'s 2002 (47) paper considers the difference in willingness to accept monetary compensation to forgo a health care program and willingness to pay for the same benefit and link it to the cost effectiveness threshold. This paper came from the perspective of the threshold representing social preferences rather than the shadow price of a fixed budget constraint and highlights that from a traditional 'welfarist' economics standpoint; a greater threshold value for disinvestment may be welfare maximising. Similarly both Devlin et al. (38) and Speight et al. (48) have suggested a threshold for disinvestment of currently

[1] This is the aim of the new Value Based Pricing approach currently under development by the Department of Health [10, 44]

performed activities could be lower than for new activities, however, neither present any methodology for calculating the weight of a disinvested activity.

This is in contrast with the view that cost effectiveness analysis guides the decisions of health systems with the objective of maximising some measure of health benefit subject to a budget constraint. Hughes et al. (21) has argued that differential threshold with respect to investment and disinvestment would result in sub-optimal levels of population health. This is because a new technology that would improve health may be rejected under a policy of having different thresholds for investment and disinvestment but not if the threshold values were the same. The authors argue that this failure to maximise population health represents an avoidable inefficiency not related to the aim of the health care sector to maximise health and thus making the case for a single threshold value for disinvestment and investment. This point can be seen as a further case for the shadow price approach as opposed to the social WTP perspective as it highlights that, given a fixed NHS budget, the social WTP approach will not lead to a maximisation of health.

A4.7. The need for an independent threshold panel

Related to the discussion over the correct role of NICE in determining a suitable cost effectiveness threshold for the NHS is the literature on the potential for an independent threshold panel. Such a panel has been characterised in a similar manner to the Monetary Policy Committee (the setters of Bank of England's interest rate who act independently of the Government of the United Kingdom), as an independent committee responsible for the setting and updating of the cost-effectiveness threshold used by NICE.

The papers covering this topic are consistent in their call for an independent threshold panel, with no papers identified arguing against it. The main case provided in the literature for an independent setter is the removal of political influence; Claxton et al. (10) argue that political influence may drive the threshold up as politicians seek to use the threshold as a means to encourage investment by pharmaceutical companies. Williams (37) suggests that NICE is biased in the setting of a threshold, as its political connections mean a higher threshold makes it is more popular with the "sellers" (the author defines sellers as not only the pharmaceutical industry but health care professionals and patient groups) by allowing more technologies to be approved. Similarly papers by Appleby et al. (43) and Raftery et al. (49) call for the creation of an independent threshold setter. 2008 Health Select Committee (50) recommended that a body independent of NICE should be established to set and review the threshold. However, it is unclear if such a body would also be independent of political influence or just of the NICE structure.

A4.8. Arguments against the use of a cost effectiveness threshold

A number of authors have argued against the use of a threshold. As mentioned earlier authors such as Gafni and Birch (17, 18) have suggested that the threshold approach risks leading to spiralling increases in inflationary pressures on health care spending, and present an alternative approach based on the use of league tables of cost-effectiveness. The reason, they argue, is that there is no guarantee that the activities displaced are less cost-effective than those new technologies imposing cost on the health system budget. This observation is coupled with the expectation of authors such as Cohen et al. (11) that pharmaceutical firms will inevitably price their drugs so as to ensure the ICER of their proposed new technology is sufficiently close to the threshold to ensure adoption and thereby gain maximum producer surplus. This observation implies that providers such as the NHS may be forced to pay above market costs of new technologies by revealing their maximum willingness to pay, in the form of the threshold. In addition the point raised in McCabe et al. (29) that the threshold should be adjusted regularly over time to ensure its efficiency seeks to address both of these arguments.

Other authors such as Eichler et al. (15) have raised and debated the issues around the theoretical base for the cost-effectiveness threshold, namely the assumption of perfect divisibility of healthcare programs, constant returns to scale and constant marginal opportunity costs(15, 17, 19, 51, 52).

Bridges et al. (53) argues that a unique threshold value imposes impractical assumptions in the case of the US health care sector, and fails to account for supply and demand side variations in the market. As an alternative the authors propose a series of thresholds that reflect regional, dynamic, budgeting and general methodological differences. They conclude that the case for abandoning a fixed threshold outweighs those for keeping one in the US and that any threshold should vary across payer, over time, in the true budget impact of interventions and in the measurement of the effectiveness of interventions. This argument has clear links to the argument for shadow pricing of the threshold rather than the social WTP approach, as the shadow price approach is based on the view that the threshold is determined by budget and current efficiency which can be seen to differ over time and across payers. The unresolved issue here is the degree to which different sub-groups (e.g. by region or budget) require different threshold values.

A4.9. Identification of activities under the threshold

An important part of the literature is the discussion around the identification of activities with an ICER greater than the proposed threshold. The importance of this discussion stems from the requirement of new activities to displace current activities that are at the margin of what is cost-effective. If it is not possible to identify these activities separately from others then threshold analysis is methodologically flawed, as the funding of a new activity may impact on an activity with an ICER above the proposed threshold.

Most literature on this topic focuses on the importance of identifying activities to be displaced rather than the process and feasibility of doing so. For example, Hughes et al. (21) and McCabe et al. (29) highlight the implications of inconsistent displacement on geographic variations in health care provision and that the lack of consistency in the displacement process undercuts the use of a single cost effectiveness threshold for the evaluation of new technologies. Similarly Buxton et al. (54) suggests that, in order to fully appreciate the opportunity cost of the implementation of a new technology, we must have a clear knowledge of those activities displaced at the cost effectiveness margin.

Few authors have sought to develop methods to identify the activities that should be displaced to free-up budget for new more cost-effective activities. Elshaug et al. (16) outlines a set of criteria for the identification of existing, potentially non-cost-effective practices which could then be further assessed to assess their cost-effectiveness using health technology assessment. The criteria suggested include factors such as: new evidence on safety; efficacy or cost-effectiveness, geographic variation that have become apparent since technology adoption, clinical heterogeneity in the clinical procedure, and technological development.

A5. The current value of the threshold

Since it became evident that decision making bodies such as NICE are using (more or less explicit) cost-effectiveness thresholds, there has been a significant level of debate over its appropriate value (35, 37-39, 43-45, 47-49, 53-58). In this section we will present three areas of the debate:

- The lack of empirical basis to the current value
- Arguments over the value being generally too high or too low
- If and how the threshold should change over time

A5.1. Lack of empirical base to the current value

Since NICE made it clear that it uses an explicit threshold (30) there has been little hiding the lack of evidential justification behind the £20,000 to £30,000 range. Indeed the Health Select Committee (50) heard (during their enquiry into NICE in 2008) that the NICE threshold has no basis in hard science.

Similarly Appleby (43) noted that "the uncomfortable truth is that NICE's threshold has no basis in either theory or evidence."

Similarly the US value of $50,000 per QALY, which is often cited as the cost-effectiveness threshold relevant to resource allocation decisions in that country, is often attacked for its lack of empirical founding (20, 23, 32, 55). Some have suggested that the US figure is rooted in the cost-effectiveness of Hospital Renal Dialysis (20), although why this makes it suitable for use more generally is unclear.

A5.2. The threshold changing over time

Another concern of current NICE practice is the apparent lack of change in the threshold value used since the body's inception. Many authors have argued that factors such as the NHS budget, price inflation, technological developments in the NHS and the discount rate applied to economic evaluations (20, 24, 25, 59) have all changed since the first use of the cost-effectiveness threshold. As such, the threshold should have changed to reflect this fact. Braithwaite et al. (55) sought to demonstrate the impact of budget and technological growth on the optimal threshold. By creating a computer simulation of the US Medicare system, the authors were able to demonstrate the impact of these factors. While there is no doubt in the literature that the NICE threshold should potentially change over time[2] no papers have been identified which model the impact of any changes on the threshold.

Both Ubel et al. (57) and Raftery (49) discuss the principles behind the directional change the threshold should take over time. Ubel et al. (57) have argued that the optimal threshold value needs to fall over time assuming medical innovation continues at roughly its current rate. Raftery (49) has noted that, in real terms, the threshold has been falling since 1999 as, in order to stay constant in real terms, it should have increased given inflation (up 40% in the time period) and increased NHS spending (up 90%). The authors argue that this decline in the threshold should have been observed in the value used by NICE in decision making. They describe the suggestion of a rise in the threshold being linked to the observed growth of the NHS budget over the last decade as "audacious" (49). It is unclear to what extent the authors disagree with this interpretation of NHS efficiency as a relevant factor affecting the optimal threshold.

A5.3. Threshold value generally too high or low

The majority of the debate over the current use of the threshold in the UK (and elsewhere) has been centred on whether the current value is too low or too high. The papers that will be discussed in this section focus on the general discussion of necessary directional change in the value rather than the presentation of a specific value; the latter is discussed in more detail in the following section on the proposed values of the threshold in the literature.

Vernon et al. (58) presented an analysis of the implications of the threshold being above or below its optimum value in terms of signals to the companies involved in research and development of new medical products. The authors concluded that if the threshold is set too low (below the economic value of the health benefit) it will result in research and development investment levels that are too low relative to their economic value (at the margin). The reason for this lies in a lack of returns to investments for the pharmaceutical companies. However, in the isolated case of the threshold relevant to the NHS (a small proportion of the world pharmaceutical market), the impact of changes to the threshold on the international pharmaceutical market equilibrium is unknown but likely to be small.

Similarly, thresholds set too high (above the economic value of the health benefit) will result in inefficiently high levels of research and development spending, such that the health care provider is funding projects that do not have a sufficient impact on social welfare.

---

[2] In fact in the 2004 NICE Methods Guide [30] noted that "the threshold will change over time as the budget for healthcare changes" *(p. 33)*. However, there is no clear reference to this change in the 2008 Methods Guide [31].

The literature that argued the threshold is too high in the UK can be broadly characterised into three key papers. Alan Williams (37) made the case that, intuitively, the threshold should not be significantly greater than the GDP per capita (roughly £18,000 in the UK in 2004). He made the case that, while it may be possible to provide a lot of the population with health care when the threshold is above the GDP per capita, it is not possible to provide health care for much of the population without imposing great hardship on those expected to foot the bill (the tax payer or government debt).

Secondly, Raftery (49) argued that, while the UK threshold has been historically too high, it does not need reducing as the real value has decreased since 1999 due to inflationary pressure and increases in the NHS budget. He also suggests that recent policies implemented by NICE, such as greater weight being given to the benefits of treatment accruing to patients at the end of their life, need to be offset by reductions in the threshold for all other treatments for expenditure to remain within the NHS budget. Finally, Raftery cites the opportunity cost analysis of trastuzumab (4) which showed that more cost-effective oncology services were being sacrificed to fund trastuzumab in breast cancer. This result suggests directly that, in some cases at least, the threshold value is too high.

Work by Martin et al. (1, 3) investigated the cost per life year saved in a selection of the 23 programme budgeting categories used in the NHS; these results are presented in Table A.2. It is important to note that these results are presented as the cost per life year gained rather than the cost per QALY of the least cost-effective current activity. The authors and others have used these results to argue that the threshold used by NICE may be too high (28). Similarly, Collier's (12) report of the Health Select Committee suggests that the threshold used by NICE is higher than that used by PCTs.

Table A.2: table showing cost per life year gained results of Martin et al.(1-3) papers

Programme budgeting category	cost per life year gained	
	2005/6 data	2004/5 data
Cancer	£13,137	£13,931
Circulation problems	£8,426	£7,979
Respiratory problems	£7,397	N/A
Gastro-intestinal problems	£18,999	N/A
Diabetes	£26,453	N/A

In contrast, a range of authors have argued that the current NICE threshold is too low. Both Speight et al. (48) and Towse (44) argued that the inclusion of wider social costs/benefits and full consideration of social willingness to pay for additional health gains show that the threshold should be significantly larger. Both cite recent NICE work by Mason et al. (56, 60) which suggested the threshold should be between £30,000 and £75,000 per QALY based on attempts to model a willingness to pay based value of a QALY based on observations of the value of avoiding a statistical fatality. Similarly in the US Ubel et al. (57) have argued that, if inflation and willingness to pay valuations are taken into account, the relevant threshold in the US should be closer to $200,000 per QALY that the regularly cited $50,000.

Those analyses which conclude the UK and US thresholds should be significantly higher have, at the core of their argument, the assumption that the respective health care budget is fully capable of responding to society's willingness to pay for additional health gains.

6. Potential methods for threshold estimation

There are broadly three approaches that can be taken to determine the threshold value (43, 54): social WTP, shadow pricing of the budget constraint and non analytical approaches such as expert elicitation. This project is concerned with the latter approach to estimating the cost effectiveness threshold. This is entirely consistent with the remit of the NHS in general and NICE in particular - they do not set the NHS budget but have to allocate those finite resources appropriately.

A6.1. Papers seeking to elicit social WTP and non-analytical approaches

The majority of the literature that has presented a proposed value for the threshold (in the UK, US and elsewhere) has done so using valuation methods based on willingness to pay for an additional health benefit (40, 59-76). However other approaches have been suggested. For example the World Health Organisation's (WHO) 2002 report (77) suggested that interventions costing less than three times GDP per capita for each DALY averted represent good value for money.

Lee et al. (78) sought to update the US "dialysis standard" often claimed to be the base of the US Medicare threshold (20). The authors present a valuation of $129,090 per QALY based on current dialysis practice in the US. Finally in an appendix to their edited book, Towse et al. (35) provide an interesting set of results drawn from a set of participants to the associated workshop (the majority of which were health economists). The participants were asked to anonymously record their view on what threshold NICE should apply. Eighteen responses were recorded with the average of all responses being £29,000 per QALY.

A6.2 Papers considering the shadow price of the budget constraint

The systematic review only identified four different papers by three different authors that suitably fell into the category of shadow pricing of the budget constraint.

Williams (37) suggested investigating the cost effectiveness of NHS interventions that represent the majority of the budget (he speculated that some 300 interventions accounted for about 90% of the cost incurred by the NHS). The purpose of this would be to identify current NHS activities that might not be cost-effective. He acknowledged the implausibility of conducting full technological appraisals on such a large number of interventions (estimating this would take 10 years, at which point it would be necessary to re-evaluate the initial appraisals), and thus suggested relying on expert opinion and existing patient data to speed up the process.

While Williams' recommendations related to identifying current interventions with a high cost per QALY as the basis for disinvestment, there is the potential to take this approach further and use it for a method to determine the cost effectiveness threshold even down to the level of a local decision maker. This was attempted by Appleby et al. (79) who conducted a feasibility experiment into the estimation of the appropriate NHS threshold by examining decision making in the NHS at a local level. The authors propose a structured model considering new technology's cost per weighted QALY gain in a table of all existing services. In an attempt to test the feasibility of this model they conducted interviews with senior NHS staff as well as investigating information on public health to construct a list of healthcare services introduced or discontinued in 2006/7. The authors found that it was feasible to identify decisions and to make the important step of estimating their cost-effectiveness; however, they noted that any attempts to fully evaluate sufficient decisions as to estimate a threshold would require a detailed understanding of the understanding of the decision structure at a local level as well as a significant number of observations.

The other key papers seeking to develop and implement methods for estimating the NHS threshold were those of Martin et al. (1-3). They aimed to establish a link between health care spending and health outcomes in the NHS after having adjusted for the need of the patient population. They made use data around the observed mortality at PCT level in the NHS alongside data expenditure data on health care across 23 programmes of care based on ICD010 disease categories. As has been mentioned earlier in this chapter these papers present the cost per life year of a range of programme budgeting categories, however, the key result of these papers is that it is possible to make use of existing data to determine such valuations for current NHS interventions. The authors concluded that while their results are highly limited and do not present a single cost per QALY estimate for the optimal threshold they can "inform the decisions of NICE on whether their current threshold for accepting new technologies is set at an appropriate level" *(p.37)*. These studies are the precursor of analyses presented in this report, and further details can be found in Appendix B and in Chapter 3 of the main report.

In the area of the efficient allocation of healthcare it is also important to note the contribution of the earlier mathematical papers such as Stinnett and Paltiel (41) who outlined mathematical techniques to approach the problem through the use of a mixed integer programming approach. While there approach differs from the interpretation of the threshold as used in this study it represented an important step in the evaluation of the methodology of seeking to solve the optimisation problem apparent in healthcare.

C. Conclusion

This systematic review of the literature surrounding the cost-effectiveness threshold has highlighted the significant range and diversity of the literature. Despite the international and mature nature of the literature there are significant differences in the suggested methods to represent a cost-effectiveness threshold. The main areas of debate relevant to this report have revolved around the role of NICE as a "searcher" or "setter" of the threshold (13, 29). While some authors have implicitly argued for NICE to fulfil a role of a threshold "setter" by suggesting method of elicitation of social WTP valuations of a QALY, death or life year (40, 59-76) the literature of most relevance to this research has sought to consider estimation methods consistent with its role as a "searcher"(1-3, 43).

References

1. Martin S, Rice N, Smith P. The link between health care spending and health outcomes: evidence from English programme budgeting data. CHE Research Paper 24. 2007a.
2. Martin S, Rice N, Smith P. Further evidence on the link between health care spending and health outcomes in England. CHE Research Paper 32. 2007b.
3. Martin S, Rice N, Smith P. The link between health care spending and health outcomes for the new English Primary Care Trusts. CHE Research Paper 42. 2008b.
4. Barrett A, Roques T, Small M, Smith R. How much will Herceptin really cost? Br Med J. 2006;333:1118-20.
5. Birch S, Gafni A. Cost-effectiveness ratios - in a league of their own. Health Policy. 1994;28(2):133-41.
6. Birch S, Gafni A. The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost-effectiveness threshold. Journal of health services research & policy. 2006;11(1):46-51.
7. Brock DW. How much is more life worth? Hastings Center Report. 2006;36(3):17-9.
8. Brouwer W, van Exel J, Baker R, Donaldson C. The new myth - The social value of the QALY. Pharmacoeconomics. 2008;26(1):1-4.
9. Chambers JD, Neumann PJ, Buxton MJ. Does medicare have an implicit cost-effectiveness threshold? Med Decis Mak. 2010;30(4):E14-E27.
10. Claxton K, Lindsay AB, Buxton MJ, Culyer AJ, McCabe C, Walker S, et al. Value based pricing for NHS drugs: an opportunity not to be missed? Br Med J. 2008;336(7638):251-4.
11. Cohen J, Looney W. Re: how much is life worth: cetuximab, non-small cell lung cancer, and the $440 billion question. Journal of the National Cancer Institute. 2010;102(15).
12. Collier J. Parliamentary review asks NICE to do better still. Br Med J. 2008;336(56-57).
13. Culyer A, McCabe C, Briggs A, Claxton K, Buxton M, Akehurst R, et al. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. J Health Serv Res Policy. 2007;12(1):56-8.
14. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables - more harm than good. Soc Sci Med. 1993;37(1):33-40.
15. Eichler HG, Kong SX, Gerth WC, Mavros P, Jonsson B. Use of cost-effectiveness analysis in health-care resource allocation decision-making: How are cost-effectiveness thresholds expected to emerge? Value in Health. 2004;7(5):518-28.
16. Elshaug A, Moss J, Littlejohns P, al. e. Identifying existing health care services that do not provide value for money. The Medical Journal of Australia. 2009;190(5):269-73.
17. Gafni A, Birch S. Guidelines for the adoption of new technologies - a prescription for uncontrolled growth in expenditures and how to avoid the problem. Can Med Assoc J. 1993;148(6):913-7.
18. Gafni A, Birch S. Incremental cost-effectiveness ratios (ICERs): The silence of the lambda. Soc Sci Med. 2006;62(9):2091-100.
19. Garber A, Phelps C. Economic foundations of cost-effectiveness analysis. Journal of Health Economics. 1997;16:1-31.
20. Grosse SD. Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. Expert Rev Pharmacoecon Outcomes Res. 2008;8(2):165-78.
21. Hughes DA, Ferner RE. New drugs for old: disinvestment and NICE. Br Med J. 2010;340.
22. Johannesson M, Meltzer D. Some reflections on cost-effectiveness analysis. Health Econ. 1998;7(1):1-7.
23. Johnson FR. Einstein on willingness to pay per QALY: Is there a better way? Med Decis Mak. 2005;25(6):607-8.
24. Kaplan R, Bush J. Health-related quality of life measurement for evaluation research and policy analysis. Health Psychology. 1982;1(1):61-80.
25. Laufer F. Thresholds in cost-effectiveness analysis - More of the story. Value in Health. 2005;8(1):86-7.
26. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization - tentative guidelines for using clinical and economic evaluations. Can Med Assoc J. 1992;146(4):473-81.

27. Mauskopf J, Rutten F, Schonfeld W. Cost-effectiveness league tables - Valuable guidance for decision makers? Pharmacoeconomics. 2003;21(14):991-1000.
28. Maynard A, Bloor K. The future role of NICE. Br Med J. 2010;341.
29. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold - What it is and what that means. Pharmacoeconomics. 2008;26(9):733-44.
30. NICE. Guide to the methods of technological appraisal. ref: N0514. 2004.
31. NICE. Guide to the methods of technological appraisal. ref: N1618. 2008.
32. Polsky D. Does willingness to pay per quality-adjusted life year bring us closer to a useful decision rule for cost-effectiveness analysis? Med Decis Mak. 2005;25(6):605-6.
33. Rascati KL. The $64,000 question- What is a quality-adjusted life-year worth? Clinical Therapeutics. 2006;28(7):1042-3.
34. Tappenden P, Brazier J, Ratcliffe J, Chilcott J. A stated preference binary choice experiment to explore NICE decision making. Pharmacoeconomics. 2007;25(8):685-93.
35. Towse A, Pritchard C, Devlin N. Cost-effectiveness thresholds: economic and ethical issues. Office of Health Economics: Kings Fund. 2002.
36. Weinstein M, Zeckhauser R. Critical ratios and efficient allocation. Journal of Public Economics. 1973;2:147-57.
37. Williams A. What could be nicer than NICE? OHE annual lecture (book available). 2004.
38. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ. 2004;13(5):437-52.
39. Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. Br Med J. 2004;329(7459):224-7.
40. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. Health Econ. 2004;14(2):197-208.
41. Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. Journal of Health Economics. 1996;15(5):641-53.
42. Epstein D, Chalabi Z, Claxton K, Sculpher MJ. Efficiency, equity and budgetary policies: informing decisions using mathematical programming. Med Decis Mak. 2007;27:128-37.
43. Appleby J, Devlin N, Parkin D. NICE's cost effectiveness threshold - How high should it be? Br Med J. 2007;335(7616):358-9.
44. Towse A. Should NICE's threshold range for cost per QALY be raised? Yes. Br Med J. 2009;338.
45. Rawlins MD, Barnett D, Stevens A. Pharmacoeconomics: NICE's approach to decision-making. British Journal of Clinical Pharmacology. 2010;70(3):346-249.
46. Health Do. A new value-based approach to the pricing of branded medicines: A consultation. 2010.
47. O'Brien BJ, Gertsen K, Willan AR, Faulkner LA. Is there a kink in consumers' threshold value for cost-effectiveness in health care? Health Econ. 2002;11(2):175-80.
48. Speight J, Reaney M. Wouldn't it be NICE to consider patients' views when rationing health care? Br Med J. 2009;338.
49. Raftery J. Should NICE's threshold range for cost per QALY be raised? No. Br Med J. 2009;338.
50. NICE. First report of the Health Committee 2007-2008. HC27-I. London: Stationery Office. 2008.
51. Sendi P, Briggs A. Affordability and costeffectiveness: decision-making on the costeffectiveness plane. Health Econ. 2001;10:675-80.
52. Sendi P, Al M. Revisiting the decision rule of costeffectiveness analysis under certainty and uncertainty. Soc Sci Med. 2003;57:969-74.
53. Bridges JFP, Onukwugha E, Mullins CD. Healthcare rationing by proxy cost-effectiveness analysis and the misuse of the $50 000 threshold in the US. Pharmacoeconomics. 2010;28(3):175-84.
54. Buxton M. How much are health-care systems prepared to pay to produce a QALY? European Journal of Health Economic. 2005;6(4):285-28.
55. Braithwaite R, Roberts M. $50,000 per QALY: inertia, indifference, or irrationality? presented at: Annual Meeting of the Society for Medical Decision Making. 2004.
56. Mason AR, Drummond MF. Public funding of new cancer drugs: Is NICE getting nastier? European Journal of Cancer. 2009;45(7):1188-92.

57. Ubel PA, Hirth RA, Chernew ME, Fendrick AM. What is the price of life and why doesn't it increase at the rate of inflation? Arch Intern Med. 2003;163(14):1637-41.
58. Vernon JA, Goldberg R, Golec J. Economic evaluation and cost-effectiveness thresholds signals to firms and implications for R&D investment and innovation. Pharmacoeconomics. 2009;27(10):797-806.
59. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: In search of a standard. Med Decis Mak. 2000;20(3):332-42.
60. Mason H, Jones-Lee M, Donaldson C. Modelling the monetary value of a qaly: a new approach based on uk data. Health Econ. 2009;18(8):933-50.
61. Abelson P. The value of life and health for public policy. Economic Record. 2003;79:S2-S13.
62. Baker R, Bateman I, Donaldson C, Jones-Lee M, Lancsar E, Loomes G, et al. Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. Health Technology Assessment. 2010;14(27):1-+.
63. Bobinac A, van Exel NJA, Rutten FFH, Brouwer WBF. Willingness to pay for a Quality-Adjusted Life-Year: The individual perspective. Value in Health. 2010;13(8):1046-55.
64. Byrne MM, O'Malley K, Suarez-Almazor ME. Willingness to pay per quality-adjusted life year in a study of knee osteoarthritis. Med Decis Mak. 2005;25(6):655-66.
65. Green C, Gerard K. Exploring the social value of health-care interventions: a stated preference discrete choice experiment. Health Econ. 2009;18(8):951-76.
66. Groot W, van den Brink HM. The value of health. Bmc Health Services Research. 2008;8.
67. Gyrd-Hansen D. Willingness to pay for a QALY. Health Econ. 2003;12(12):1049-60.
68. Gyrd-Hansen D. Willingness to pay for a QALY - theoretical and methodological issues. Pharmacoeconomics. 2005;23(5):423-32.
69. Haninger K, Hammitt J. Willingness to pay for Quality-Adjusted Life Years: empirical inconsistency between cost-effectiveness analysis and economic welfare theory. OECD. 2006.
70. Johnson FR, Backhouse M. Eliciting stated preferences for health-technology adoption criteria using paired comparisons and recommendation judgments. Value in Health. 2006;9(5):303-11.
71. King JT, Tsevat J, Lave JR, Roberts MS. Willingness to pay for a quality-adjusted life year: Implications for societal health care resource allocation. Med Decis Mak. 2005;25(6):667-77.
72. Lieu TA, Ray GT, Ortega-Sanchez IR, Kleinman K, Rusinak D, Prosser LA. Willingness to pay for a QALY based on community member and patient preferences for temporary health states associated with Herpes Zoster. Pharmacoeconomics. 2009;27(12):1005-16.
73. Luis Pinto-Prades J, Loomes G, Brey R. Trying to estimate a monetary value for the QALY. Journal of Health Economics. 2009;28(3):553-62.
74. Shiroiwa T, Sung Y-K, Fukuda T, Lang H-C, Bae S-C, Tsutani K. International survey on willingness-to-pay (wtp) for one additional qaly gained: what is the threshold of cost effectiveness? Health Econ. 2010;19(4):422-37.
75. Smith RD, Richardson J. Can we estimate the 'social' value of a QALY? Four core issues to resolve. Health Policy. 2005;74(1):77-84.
76. Yaesoubi R, Roberts SD. A game-theoretic framework for estimating a health purchaser's willingness-to-pay for health and for expansion. Health Care Management Science. 2010;13(4):358-77.
77. WHO. The World Health report: reducing risks, promoting healthy life. 2002.
78. Lee CP, Chertow GM, Zenios SA. An empiric estimate of the value of life: updating the renal dialysis cost-effectiveness standard. Value in Health. 2009;12(1):80-7.
79. Appleby J, Devlin N, Parkin D, Buxton M, Chalkidou K. Searching for cost effectiveness thresholds in the NHS. Health Policy. 2009;91(3):239-45.

Papers discovered by the literature review

Note: Not all of these papers are referenced in this Appendix and some references used were not discovered through the systematic review

Initial Pearls

1. Appleby, J., N. Devlin, and D. Parkin, *NICE's cost effectiveness threshold - How high should it be?* BRITISH MEDICAL JOURNAL, 2007. **335**(7616): p. 358-359.
2. Appleby, J., et al., *Searching for cost effectiveness thresholds in the NHS*. HEALTH POLICY, 2009. **91**(3): p. 239-45.
3. Bridges, J.F.P., E. Onukwugha, and C.D. Mullins, *Healthcare Rationing by Proxy Cost-Effectiveness Analysis and the Misuse of the $50 000 Threshold in the US*. PHARMACOECONOMICS, 2010. **28**(3): p. 175-184.
4. Culyer, A., et al., *Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence.* JOURNAL OF HEALTH SERVICES RESEARCH & POLICY, 2007. **12**(1): p. 56-8.
5. Devlin, N. and D. Parkin, *Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis.* HEALTH ECONOMICS, 2004. **13**(5): p. 437-452.
6. Gafni, A. and S. Birch, *Incremental cost-effectiveness ratios (ICERs): The silence of the lambda.* Social Science & Medicine, 2006. **62**(9): p. 2091-2100.
7. McCabe, C., K. Claxton, and A.J. Culyer, *The NICE cost-effectiveness threshold - What it is and what that means.* PHARMACOECONOMICS, 2008. **26**(9): p. 733-744.
8. Raftery, J., *Should NICE's threshold range for cost per QALY be raised? No.* BRITISH MEDICAL JOURNAL, 2009. **338**.
9. Towse, A., *Should NICE's threshold range for cost per QALY be raised? Yes.* BRITISH MEDICAL JOURNAL, 2009. **338**.
10. Braithwaite et al. What Does the Value of Modern Medicine Say About the $50,000 per Quality-Adjusted Life-Year Decision Rule? Medical Care, 2008, 46 (4) pp349
11. Grosse, S, Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. Expert Review of Pharmacoeconomics & Outcomes Research, 2008, 8 (2) pp 165
12. Rawlins, M.D. and A.J. Culyer, National Institute for Clinical Excellence and its value judgments. British Medical Journal, 2004. 329(7459): p. 224-227.
13. Chambers, Neumann and Buxton Does Medicare Have an Implicit Cost-Effectiveness Threshold? Medical Decision Making, 2010, 30 (4) pp E14-27
14. Chambers, J.D., P.J. Neumann, and M.J. Buxton, Does Medicare Have an Implicit Cost-Effectiveness Threshold? Medical Decision Making, 2010. 30(4): p. E14-E27.

Step 1 Results

1. Brouwer, W., et al., *The new myth - The social value of the QALY*. Pharmacoeconomics, 2008. **26**(1): p. 1-4.
2. Claxton, K., et al., *Value based pricing for NHS drugs: an opportunity not to be missed?* British Medical Journal, 2008. **336**(7638): p. 251-254.
3. Cohen, J. and W. Looney, *Re: How Much Is Life Worth: Cetuximab, Non-Small Cell Lung Cancer, and the $440 Billion Question.* Journal of the National Cancer Institute, 2010. **102**(15).
4. Eichler, H.G., et al., *Use of cost-effectiveness analysis in health-care resource allocation decision-making: How are cost-effectiveness thresholds expected to emerge?* Value in Health, 2004. **7**(5): p. 518-528.
5. Green, C. and K. Gerard, *EXPLORING THE SOCIAL VALUE OF HEALTH-CARE INTERVENTIONS: A STATED PREFERENCE DISCRETE CHOICE EXPERIMENT*. Health Economics, 2009. **18**(8): p. 951-976.

6. Groot, W. and H.M. van den Brink, *The value of health.* Bmc Health Services Research, 2008. **8**.
7. Hughes, D.A. and R.E. Ferner, *New drugs for old: disinvestment and NICE.* British Medical Journal, 2010. **340**.
8. Lieu, T.A., et al., *Willingness to Pay for a QALY Based on Community Member and Patient Preferences for Temporary Health States Associated with Herpes Zoster.* Pharmacoeconomics, 2009. **27**(12): p. 1005-1016.
9. Mason, A.R. and M.F. Drummond, *Public funding of new cancer drugs: Is NICE getting nastier?* European Journal of Cancer, 2009. **45**(7): p. 1188-1192.
10. Mason, H., M. Jones-Lee, and C. Donaldson, *MODELLING THE MONETARY VALUE OF A QALY: A NEW APPROACH BASED ON UK DATA.* Health Economics, 2009. **18**(8): p. 933-950.
11. Maynard, A and Bloor, K *The future role of NICE.* BMJ. 341:c6286, November 13, 2010.
12. Rascati, K.L., *The $64,000 question- What is a quality-adjusted life-year worth?* Clinical Therapeutics, 2006. **28**(7): p. 1042-1043.
13. Rawlins, M., D. Barnett, and A. Stevens, *Pharmacoeconomics: NICE's approach to decision-making.* British Journal of Clinical Pharmacology, 2010. **70**(3): p. 346-349.
14. Shiroiwa, T., et al., *INTERNATIONAL SURVEY ON WILLINGNESS-TO-PAY (WTP) FOR ONE ADDITIONAL QALY GAINED: WHAT IS THE THRESHOLD OF COST EFFECTIVENESS?* Health Economics, 2010. **19**(4): p. 422-437.
15. Speight, J. and M. Reaney, *Wouldn't it be NICE to consider patients' views when rationing health care?* British Medical Journal, 2009. **338**.
16. Tappenden, P., et al., *A stated preference binary choice experiment to explore NICE decision making.* Pharmacoeconomics, 2007. **25**(8): p. 685-693.
17. Weinstein, M.C., *How much are Americans willing to pay for a quality-adjusted life year?* Medical Care, 2008. **46**(4): p. 343-345.
18. Yaesoubi, R. and S.D. Roberts, *A game-theoretic framework for estimating a health purchaser's willingness-to-pay for health and for expansion.* Health Care Management Science, 2010. **13**(4): p. 358-377.

Step 2 Results

1. Appleby J, Devlin N, Parkin D, et al. Searching for local NHS cost effectiveness thresholds: a feasibility study. NICE confer Manchester; 2007 Dec 5-6 [online] Available from URL: http://www.nice2007.co.uk/ApplebyDevlin.pdf
2. Birch S, Gafni A. The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost–effectiveness threshold. J. Health Serv. Res. Policy11,46–51 (2006).
3. Braithwaite RS, Roberts MS. $50,000 per QALY: inertia, indifference, or irrationality? Presented at: Annual Meeting of the Society for Medical Decision Making. Atlanta, GA, USA, 17–20 October, 2004.
4. Drummond M, Torrance G, Mason J. Cost–effectiveness league tables: more harm than good? Soc. Sci. Med.37,33–40 (1993).
5. Gerard K, Mooney G. QALY league tables: handle with care. Health Econ 1993; 2 (1): 59-64
6. Gyrd-Hansen D. Willingness to pay for a QALY: theoretical and methodological issues. Pharmacoeconomics 2005; 23 (5): 423-32
7. Hammitt JK. The $64,000 question: what are we willing to pay for a QALY. ISPOR Connections11(1),7–9 (2005).
8. Hirth RA, Chernew ME, Miller E, et al. Willingness to pay for a quality-adjusted life year: in search of a standard. Med Decis Making 2000; 20 (3): 332-42
9. King JT Jr, Tsevat J, Lave JR, Roberts MS. Willingness to pay for a quality-adjusted life year: implications for societal health care resource allocation. Med. Decis. Making25,667–677 (2005). [CrossRef] [Medline]
10. Lee C, Chertow G, Zenios S. An empiric estimate of the value of life: updating the renal dialysis cost-effectiveness standard. Value Health 2009; 12 (1): 80-7
11. Martin S, Rice N, Smith P. Further evidence on the link between health care spending and health outcomes in England [CHE 28. National Institute for Health and Clinical Excellence. NICE discussion paper 32]. York: University of York, 2007
12. Martin S, Rice N, Smith PC. The link between health care spending and health outcomes for the new English primary care trusts. CHE Research Paper 42, University of York, York; 2008.

13. Mauskopf J, Rutten F, Schonfeld W. Cost–effectiveness league tables: valuable guidance for decision makers? Pharmacoeconomics21,991–1000 (2003).
14. Smith RD, Richardson J. Can we estimate the 'social' value of a QALY? Four core issues to resolve. Health Policy74,77–84 (2005). [CrossRef] [Medline]
15. Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. J Health Econ 1996; 15 (5): 641-53
16. Towse A, Pritchard C, Devlin N, eds. Cost effectiveness thresholds: economic and ethical issues. London: Office of Health Economics, Kings Fund, 2002.
17. Ubel PA, Hirth RA, Chernew ME, et al. What is the price of life and why doesn't it increase at the rate of inflation? Arch Int Med 2002; 163: 1637-41
18. Williams A. What could be nicer than NICE? London: Office for Health Economics, 2004. OHE Annual Lecture 2004
19. Winkelymayer WC, Weinstein MC, Mittelman MA, Glynn RJ, Pliskin JS. Health economic evaluations: the special case of end-stage renal disease treatment. Med Decis Making 2002; 22:417-30.

Step 3 Results

1. Baker, R., et al., Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. Health Technology Assessment, 2010. 14(27)
2. Bobinac, A., et al., Willingness to Pay for a Quality-Adjusted Life-Year: The Individual Perspective. Value in Health, 2010. 13(8): p. 1046-1055.
3. Brock, D.W., How much is more life worth? Hastings Center Report, 2006. 36(3): p. 17-19.
4. Byrne, M.M., K. O'Malley, and M.E. Suarez-Almazor, Willingness to pay per quality-adjusted life year in a study of knee osteoarthritis. Medical Decision Making, 2005. 25(6): p. 655-666.
5. Griffin, S., K. Claxton, and M. Sculpher, Decision analysis for resource allocation in health care. Journal of Health Services Research & Policy, 2008. 13: p. 23-30.
6. Gyrd-Hansen, D., Willingness to pay for a QALY. Health Economics, 2003. 12(12): p. 1049-1060.
7. Harrison, S., A POLICY AGENDA FOR HEALTH-CARE RATIONING. British Medical Bulletin, 1995. 51(4): p. 885-899.
8. Laufer, F., Thresholds in cost-effectiveness analysis - More of the story. Value in Health, 2005. 8(1): p. 86-87.
9. Pinto-Prades, J.L., G. Loomes, and R. Brey, Trying to estimate a monetary value for the QALY. Journal of Health Economics, 2009. 28(3): p. 553-562.
10. Vernon, J.A., R. Goldberg, and J. Golec, Economic Evaluation and Cost-Effectiveness Thresholds Signals to Firms and Implications for R&D Investment and Innovation. Pharmacoeconomics, 2009. 27(10): p. 797-806

Step 4 Results

1. Abelson, P., *The value of life and health for public policy.* Economic Record, 2003. **79**: p. S2-S13.
2. Fryback, D.G. and W.F. Lawrence, *Dollars may not buy as many QALYs as we think: A problem with defining quality-of-life adjustments.* Medical Decision Making, 1997. **17**(3): p. 276-284.
3. Gafni, A. and S. Birch, *GUIDELINES FOR THE ADOPTION OF NEW TECHNOLOGIES - A PRESCRIPTION FOR UNCONTROLLED GROWTH IN EXPENDITURES AND HOW TO AVOID THE PROBLEM.* Canadian Medical Association Journal, 1993. **148**(6): p. 913-917.
4. Johnson, F.R., *Einstein on willingness to pay per QALY: Is there a better way?* Medical Decision Making, 2005. **25**(6): p. 607-608.
5. Laupacis, A., et al., *HOW ATTRACTIVE DOES A NEW TECHNOLOGY HAVE TO BE TO WARRANT ADOPTION AND UTILIZATION - TENTATIVE GUIDELINES FOR USING CLINICAL AND ECONOMIC EVALUATIONS.* Canadian Medical Association Journal, 1992. **146**(4): p. 473-481.
6. Polsky, D., *Does willingness to pay per quality-adjusted life year bring us closer to a useful decision rule for cost-effectiveness analysis?* Medical Decision Making, 2005. **25**(6): p. 605-606.

7. Martin, S., N. Rice, and P. Smith, The link between health care spending and health outcomes: evidence from English programme budgeting data. CHE Research Paper 24, 2007a.
8. Chambers, J.D., P.J. Neumann, and M.J. Buxton, Does Medicare Have an Implicit Cost-Effectiveness Threshold? Medical Decision Making, 2010. 30(4): p. E14-E27.

Step 5 Results

1. Birch, S. and A. Gafni, COST-EFFECTIVENESS RATIOS - IN A LEAGUE OF THEIR OWN. Health Policy, 1994. 28(2): p. 133-141.
2. Johnson, F.R. and M. Backhouse, Eliciting stated preferences for health-technology adoption criteria using paired comparisons and recommendation judgments. Value in Health, 2006. 9(5): p. 303-311.

Step 6 Results

1. Dolan, P., et al., QALY maximisation and people's preferences: a methodological review of the literature. Health Economics, 2004. 14(2): p. 197-208.
2. Baker, R., et al., Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. Health Technology Assessment, 2010. 14(27)
3. O'Brien, B.J., et al., Is there a kink in consumers' threshold value for cost-effectiveness in health care? Health Economics, 2002. 11(2): p. 175-180.
4. Buxton, M., How Much Are Health-Care Systems Prepared to Pay to Produce a QALY? European Journal of Health Economic, 2005. 6(4): p. 285-28.
5. Mason, A.R. and M.F. Drummond, Public funding of new cancer drugs: Is NICE getting nastier? European Journal of Cancer, 2009. 45(7): p. 1188-1192.

Appendix B

THE LINK BETWEEN NHS SPENDING AND MORTALITY: ESTIMATING THE COST OF A LIFE YEAR IN ENGLAND

Contents

THE LINK BETWEEN NHS SPENDING AND MORTALITY: ESTIMATING THE COST OF A LIFE YEAR IN ENGLAND[1]

Prologue

This report presents, in a linear fashion, details of the econometric work undertaken to estimate the link between NHS spending and mortality. It also presents details of how the econometric work is used to calculate the cost of a life year. This report is designed to serve as a reference document in support of the main project report, which highlights the major findings from the project. As a supporting document this report provides far more detail than most interested parties will require. Nevertheless, those who seek more detail than that contained in the main project report may find the material here useful.

[1] This study builds on previous work that was undertaken as part of the Quest for Quality and Improved Performance, a five-year initiative of the Health Foundation.

A BACKGROUND, MODEL, DATA, AND ESTIMATION APPROACH

B1. Introduction

In a recent White Paper the new British Conservative government emphasized the importance of clinical outcomes. It notes that, in future, success will be measured, not through the achievement of process targets, such as short waiting times, but against outcomes such as cancer and stroke survival rates [1]. Although the NHS budget is ring-fenced against the on-going public sector deficit reduction programme, its budget is still likely to be under considerable pressure, and attention is likely to focus on the extent to which any additional health care expenditure yields genuine patient benefits in the form of improved health outcomes.

However, one of the most fundamental yet unresolved issues in health policy is the extent to which additional health care expenditure yields patient benefits, in the form of improved health outcomes. The work of health technology agencies, such as the English National Institute for Health and Clinical Excellence (NICE), has greatly improved our understanding at the micro-level of the costs and benefits of individual therapeutic technologies. However, there remains a dearth of evidence at the macro-level on the benefits of increased health system expenditure.

Recently a series of studies has taken advantage of the availability of two new datasets to examine the relationship between NHS expenditure and mortality rates for various disease categories[2-5]. One dataset contains mortality rates for various disease categories at the level of geographically defined local health authorities, known as Primary Care Trusts (PCTs). The other dataset presents NHS expenditure by PCT on 23 broad programmes of care. This dataset embraces most items of publicly funded expenditure, including inpatient, outpatient and community care, and pharmaceutical prescriptions.

Like previous studies, we employ a model that assumes that each PCT receives an annual financial lump sum budget from the national ministry and allocates its resources across the 23 programmes of care to maximize the health benefits associated with that expenditure. Estimation of this model using the expenditure and mortality data facilitates two related studies: first, a study of how changes in the NHS budget impact on expenditure in each care programme; and second, a study of the link between expenditure in a programme and the health outcomes achieved, notably in the form of disease specific mortality rates. The latter study also permits the calculation of the cost of an additional life year for individual programmes of expenditure.

The work presented here draws heavily upon previous studies. These were constrained in a number of ways and, in this analysis; we build on and improve these previous studies in four major ways:

- first, due to data limitations previous studies related expenditure in time period *t* to mortality in periods *t*, *t-1*, and *t-2* combined. In doing this, such studies assumed that PCTs had reached some sort of equilibrium in the expenditure choices they make and the outcomes they secure. This is probably not an unreasonable assumption given the relatively slow pace at which both types of variable change but, with more recent mortality data now available, here we relate expenditure in time period *t* to mortality in periods *t*, *t+1*, and *t+2* combined (see section B8.5).

- second, previous studies have tended to focus on a very limited number of care programmes (e.g., for cancer, circulatory disease, gastro-intestinal problems and respiratory problems). Here we present plausible outcome models for a larger number of budgeting categories.

- third, previous estimates of the cost of a life year have been for individual programmes of care. Here we present estimates of the cost of a life year for an enlarged number of programmes and, importantly, with the aid of assumptions about the productivity of programmes without a meaningful mortality-based outcome indicator, we extend our individual programme estimates to incorporate expenditure across all programmes of care.

- finally, although previous results and our current models 'pass' the appropriate statistical tests, we subject our latest results to a substantial sensitivity analysis.

The structure of this report is as follows. Section B2 presents a brief review of previous empirical studies in this domain, which have often yielded conflicting results. A straightforward theoretical model of the budgetary problem faced by a PCT manager seeking to allocate limited funds between competing programmes of care is presented in section B3. The programme budgeting and health outcome (mortality) data are described in sections B4 and B5 respectively. Section B6 outlines our estimation methods and some of the issues surrounding them.

In section B7 we commence our empirical work by estimating well specified econometric models that outline (a) the budgetary expenditure choices and (b) the health outcomes achieved by PCTs using expenditure data for 2005/6 and mortality data for 2002/3/4. Section B8 presents results using expenditure data for 2006/7 and mortality data for 2004/5/6. It also presents results using the same expenditure data but updating the mortality data to 2006/7/8. Several pieces of sensitivity analysis are also included in section B8, but the major piece of sensitivity analysis – examining the impact of relaxing the instrument validity restriction – is reported in section B9.

In section B10 we re-estimate our model using updated expenditure and mortality data. In particular, we use the programme budgeting expenditure for 2007/8 and mortality data for 2007/2008/2009 to re-estimate our outcome and expenditure equations. In section B11 we update the dataset again, and this time we employ programme budgeting expenditure data for 2008/9 and mortality data for 2008/2009/2010. We also compare the elasticities and cost of a life year estimates that we have obtained using expenditure and mortality data for different years.

Finally, section B12 presents a summary of our findings and some concluding remarks.

B2. Previous studies

There is a large literature on the determinants of international variations in health care spending in which income levels often play a central role [6]. However, whether more expenditure generates better outcomes – for example, in terms of reduced mortality – remains a matter of debate. For example, Fisher and Welch [7] note various ways in which more health care might harm patients and they cite various studies supporting their arguments. In a comprehensive review, Nolte and McKee [8] discuss many studies that examine the impact of health care and other explanatory variables on some measure of health care outcome. Nolte and McKee point out that researchers usually combine a production function approach with the application of regression analysis. For example, in an early cross-sectional study of 18 developed countries, Cochrane et al.[9] use regression analysis to examine the statistical relationship between mortality rates on the one hand and per capita GNP and per capita consumption of inputs such as health care provision on the other. They find that the indicators of health care provision were generally not associated with outcomes in the form of mortality rates. Thereafter, the failure to identify strong and consistent relationships between health care expenditure and health outcomes (after controlling for other factors) has become a consistent theme in the literature, whilst, in contrast, socioeconomic factors are often found to be good determinants of health outcomes[8, 10, 11].

This failure to detect a significant positive relationship between expenditure and health outcome might reflect the difficulties associated with any such study rather than the absence of such a relationship. For example, Gravelle and Backhouse [12] examine some of the methodological difficulties associated with empirical investigation of the determinants of mortality rates. These include simultaneous equation bias and the associated endogeneity problem (that the level of health care input might reflect the level of health outcome achieved in the past), and that a lag may occur between expenditure and outcomes (studies typically assume that expenditure has an immediate effect on mortality). To avoid the difficulties imposed by data heterogeneity inherent in international analyses, the study by Cremieux et al[13] examines the relationship between expenditure and outcomes across ten Canadian provinces over the fifteen-year period 1978-1992. They find that lower healthcare spending is associated with a significant increase in infant mortality and a decrease in life expectancy.

Although challenging the received empirical wisdom, one difficulty with the Cremieux et al[13] study is that the estimated regression equation consists of a mixture of potentially endogenous variables (such as the number of physicians, health spending, alcohol and tobacco consumption, expenditure on meat and fat) and exogenous variables (such as income and population density). The authors' chosen estimation technique (GLS) does not allow for this endogeneity and consequently the coefficients on the endogenous variables may be biased [12]. Or's [14] study of the determinants of variations in mortality rates across 21 OECD countries between 1970 and 1995 may suffer from the same weakness. She finds that the contribution of the number of doctors to reducing mortality in OECD countries is substantial but her estimation technique assumes that the number of doctors is exogenous to the health system.

Nixon and Ulmann[15] provide a detailed review of 16 studies that have examined the relationship between health care inputs and health outcomes, using macro-level data. They also undertake their own study using data for 15 EU countries over the period 1980-1995. They employ three health outcomes measures – life expectancy at birth for males and females, and the infant mortality rate – and a dozen or more explanatory variables including: per capita health expenditure, number of physicians (per 10,000 head of population), number of hospital beds (per 1,000 head of population), the average length of stay in hospital, the in-patient admission rate, alcohol and tobacco consumption, nutritional characteristics, and environmental pollution indicators. Nixon and Ulmann conclude that although health expenditure and the number of physicians have made a significant contribution to improvements in infant mortality, '...health care expenditure has made a relatively marginal contribution to the improvements in life expectancy in the EU countries over the period of the analysis'. Again, however, the study does not allow for the possibility that some of the explanatory variables may be endogenous.

Although loosely based on the notion of a health production function, the traditional empirical study described above has rarely been informed by an explicit theoretical model. This is understandable, as the processes giving rise to the observed health outcome are likely to be very complex, and any theoretical

model might become rather unwieldy. However, this absence of a model has usually led to an atheoretical search for measures of health inputs demonstrating a statistically 'significant' association with health outcomes. In contrast, in this study we inform our empirical modelling with a theoretical framework. We believe that this may lead to a more convincing and better specified model of health outcomes than that used in many previous studies, and this model is outlined in the next section.

B3. Theoretical model

Our modelling framework assumes that each PCT i receives an annual financial lump sum budget y_i from the national ministry, and that annual total expenditure cannot exceed this amount. The PCT must then decide how to allocate its budget across the J programmes of care (J=23 in this case). For each programme of care there is a 'health production function' $f_j(.)$ that indicates the link between local spending x_{ij} on programme j and health outcomes in that programme h_{ij}. Health outcomes might be measured in a variety of ways, but the most obvious is to consider some measure of improvement in life expectancy, possibly adjusted for quality of life, in the form of a quality adjusted life year.

The nature of the specific health production function confronted by a PCT will depend on two types of local factors: the clinical needs of the local population relevant to the programme of care (which we denote n_{ij}) and broader local environmental factors z_{ij} relevant to delivering the programme of care (such as input prices, geographical factors, or other uncontrollable influences on outcomes). Both clinical and environmental factors may be multidimensional in nature. Increased expenditure then yields improvements in health outcomes, as expressed for example in improved local mortality rates, but at a diminishing rate. That is:

$$h_{ij} = f_j(x_{ij}, n_{ij}, z_{ij}); \partial f_j / \partial x > 0; \partial^2 f_j / \partial x^2 < 0 \qquad (3.1)$$

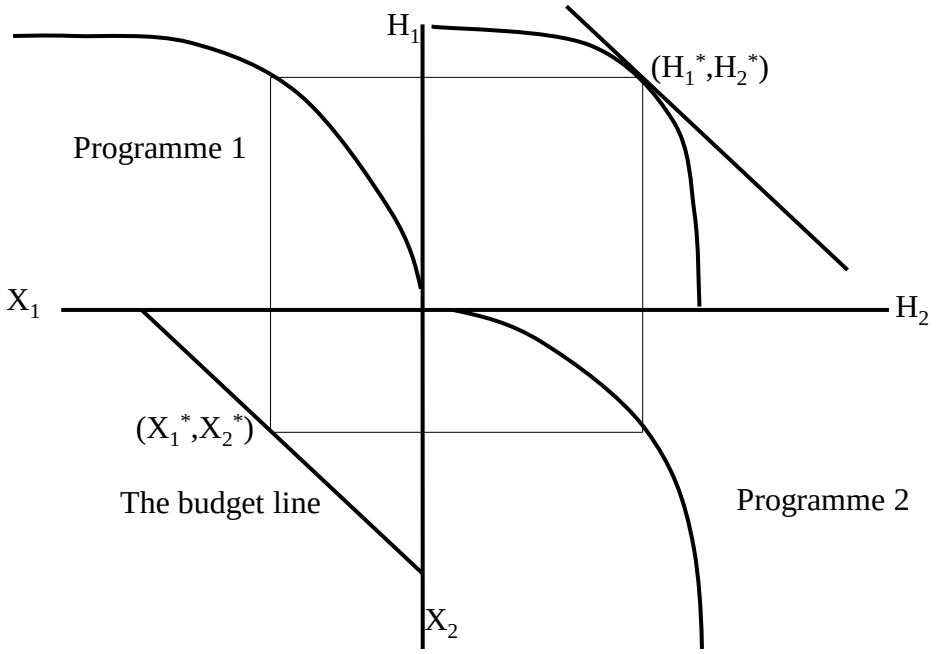
We assume there is a PCT social welfare function $W(.)$ that embodies health outcomes across the J programmes of care. Assuming no interaction between programmes of care, each PCT allocates its budget so as to maximise total welfare subject to the local budget constraint and the health production function for each programme of care:

$$\begin{aligned} &\max && W(h_{i1}, h_{i2}, \ldots, h_{iJ}) \\ &\text{subject to} && \sum_j x_{ij} \le y_i \\ & && h_{ij} = f_j(x_{ij}, n_{ij}, z_{ij}); \quad \mathrm{j} = 1, \ldots \mathrm{J} \end{aligned} \qquad (3.2)$$

It can of course quite plausibly be argued that decision-makers do not discriminate between health outcomes in different programmes of care, and that W (.) is merely the sum of such outcomes. However, there is no need for that assumption in our formulation.

Each PCT allocates expenditure across the 23 programmes of care so that the marginal benefit of the last pound spent in each programme of care is the same. This is represented diagrammatically in Figure B3.1, which illustrates the trade-off between just two programmes of care. The top left hand quadrant indicates the health production function for programme 1, whilst the bottom right hand quadrant indicates the health production function for programme 2, albeit in transposed form. The bottom left hand quadrant indicates the budget constraint: the expenditure choice must lie on the budget line. This means that for each feasible pair of expenditure choices (points on the budget constraint line), a pair of health outcomes in the two programmes emerges, which is traced out as the health production possibility frontier in the top right quadrant. The PCT will choose the point on this frontier that maximizes welfare. In this example, we have indicated a simple health maximizing approach (the maximum health summing across the two programmes), leading to optimal health outcomes (H_1^*, H_2^*) and expenditure (X_1^*, X_2^*).

Figure B3.1: graph showing optimal trade-off between two programmes of care



Solving the constrained maximisation problem yields the result that the optimal level of expenditure in each category, x_{ij}^*, is a function of the need for health care in each category ($n_{i1}, n_{i2},..., n_{ij}$), environmental variables affecting the production of health outcomes in each category ($z_{i1}, z_{i2},..., z_{ij}$), and PCT income ($y_i$). Thus

$$x_{ij}^* = g_j(n_{i1},\ldots n_{iJ}, z_{i1},\ldots z_{iJ}, y_i); \quad j = 1,...,J \tag{3.3}$$

Thus, for each programme of care there exists an expenditure equation (3.3) explaining expenditure choice of PCTs and a health outcome equation (3.1) that models the associated health outcomes achieved.

Our model is static in the sense that the health production function (3.1) assumes that all health benefits occur contemporaneously with expenditure. We acknowledge that for some programmes of care benefits might occur one or more years after expenditure has occurred. This is particularly likely to be the case for those programmes aimed at encouraging healthy lifestyles, where some benefits may occur decades after the actual programme expenditure. For other programmes, such as maternity/reproductive conditions and neonate conditions, benefits may be largely contemporaneous with expenditure. Furthermore, we do not model the decision maker's time preferences.

For our empirical modelling, however, we are constrained by the data we have available, which are largely cross-sectional in nature. Due to data limitations, previous studies have had to relate expenditure in period *t* to mortality data in periods *t*, *t-1*, and *t-2* combined so that the mortality data precedes the expenditure data. This is not ideal. Implicitly previous studies have had to assume that the data represent a quasi long-run equilibrium position, and that relative expenditure levels and health outcomes within each PCT have been reasonably stable over a period of time. As we shall see, this appears to be a reasonable assumption because we obtain similar results when we estimate our models using expenditure for period *t* with either mortality data for periods *t*, *t-1*, and *t-2* combined (section B8.4) or with mortality data for periods *t*, *t+1*, and *t+2* combined (section B8.5).

Having outlined our model, the next section discusses the datasets used to estimate this model.

B4. NHS programme budgeting in England

The English National Health Service (NHS) is the archetypal centrally planned and publicly funded health care system. Its revenue derives almost entirely from national taxation, and access to the system is generally free to the patient. Primary care is an important element of the system, and general practitioners act as gatekeepers to secondary care and pharmaceuticals. The system is organized geographically, with responsibility for the local administration of the NHS devolved to local health authorities known as Primary Care Trusts (PCTs).[2] For the purposes of this study, there were 303 PCTs with an average population of about 160,000 people until October 2006. In October 2006 the 303 PCTs became 152 PCTs. Some PCT boundaries remained unchanged while other PCTs were merged with one or more neighbours to form a new, larger, PCT. In a few cases the geographic area covered by an existing PCT was split between two or more new PCTs. These 152 PCTs have an average population of about 330,000 people.[3] PCTs are allocated fixed annual budgets by the national ministry, within which they are expected to meet expenditure on most aspects of health care, including inpatient, outpatient and community care, primary care and pharmaceutical prescriptions.

B4.1 The rationale behind the construction of programme budget data

Traditionally, PCTs and their predecessors have reported expenditure on the basis of inputs (for example, total expenditure on pay and non-pay items). However, NHS policy makers have for some time realized that this approach does not create clinically meaningful financial data or help in the design and evaluation of programmes of patient care. The Department of Health therefore initiated a 'Programme Budgeting' project. This has sought to create an accounting system that is more aligned with the distinct outputs and health outcomes of the health care system. Since April 2003, in addition to its conventional accounting data, each PCT has prepared expenditure data disaggregated according to 23 programmes of health care. These programmes are defined by reference to the International Classification of Diseases (ICD) Version 10 codes at the four digit level, and most programme budget categories reflect ICD 10 chapter headings (e.g., cancer and tumours, circulation problems, renal problems, neonates, problems associated with the skin, problems associated with vision, problems associated with hearing, etc). In some cases, the 23 categories are broken down into further sub-areas to achieve a closer match with the various National Service Frameworks (NSFs): for example, the large mental health category is broken down into 'substance abuse', 'dementia', and 'other'.

Programme budgeting seeks to allocate all types of PCT expenditure to the various programme budget categories, including secondary care, community care and prescribing. However, the system acknowledges that a medical model of care may not always be appropriate, and two specific non-clinical groups -- 'Healthy Individuals' and 'Social Care Needs' -- have been created. These are intended to capture the costs of disease prevention programmes and the costs of services that support individuals with social rather than health care needs. In addition, in some cases it is not possible to assign activity by medical condition, preventative activity, or social care need and, in these cases, expenditure is assigned to a residual category (PBC 23) entitled 'Other'. The most important element of this residual programme is expenditure on general practitioner services (PBC 23a). In principle, it should be possible to allocate each GP consultation to a particular care programme. However, at the moment the available data information systems do not permit such an allocation and so all primary care expenditure is allocated to this residual programme. The use of this residual category ensures that all expenditure is assigned to a programme of care [16].

The aim of the programme budget classifications is to identify the entire volume of health care resources assigned to broad areas of illness according to the primary diagnosis associated with an intervention. It serves a number of purposes, most notably to assist in the local planning of health care. But for this

[2] Strictly speaking, these local health authorities are Primary Care Organisations (PCOs) but the vast majority of these are 'Trusts' and we retain this terminology throughout.

[3] In April 2010 two PCTs (East & North Hertfordshire (5P3) and West Hertfordshire (5P4)) merged to form a single organisation (Hertfordshire PCT (5QV)) so that, since this date, there have been 151 PCTs. At the same time Blackburn and Darwen PCT (5CC) became Blackburn and Darwen Teaching Care Trust Plus (TAP). In April 2011 Solihull Care Trust (TAM) became a PCT (5QW).

study its crucial merit is that it opens up the possibility of examining the statistical relationship between local programme spending and the associated disease-specific outcome.

B4.2 The collection of programme budgeting data

Programme budgeting information is collected centrally by the Department of Health as part of the annual accounts process. Each PCT is required to submit an annual programme budgeting return to the Department which shows how their total expenditure is allocated across the 23 programme budgeting categories.

Various forms of data collection and analysis are required to map PCT expenditure onto acute, community and other services to the 23 programme budget categories. From the PCT perspective, however, the construction of each PCT's return largely involves collating information provided by other bodies and drawing on other information already in the PCT's own annual accounts. Thus General/Personal Medical Service expenditure, which is already reported in PCT accounts, relates to direct primary care and is mapped in its entirety to programme budget category 23a (Other: GMS/PMS); General Ophthalmic Service expenditure (again from PCT accounts) maps directly to programme budget category 8 (eye/vision problems); and General Dental Service expenditure maps directly to programme budget category 12 (dental problems). Prescribing and pharmaceutical services expenditure is allocated to programme budget categories on the basis of an annual apportionment report provided by the Prescription Pricing Authority for each PCT as part of the annual accounts process. This apportionment report allocates each PCT's annual FHS prescribing expenditure across the 23 programme budget categories. The balance of any primary healthcare purchased by the PCT is allocated /apportioned across the 23 programme budget categories on the basis of local records, with any remaining expenditure allocated/apportioned in line with the distributions already made across the budget categories.

It is the responsibility of all NHS providers – which includes PCTs, NHS Trusts, and Foundation Hospitals – to allocate admitted patient care expenditure across the programme budgeting categories, specific to each PCT that utilises its services. These allocations are constructed using 'finished consultant episodes' (from the mandatory administrative Hospital Episode Statistics data set returned by each provider) each of which is assigned to a Healthcare Resource Group (HRG), an English version of DRGs. National grouping software automatically assigns each HRG to one of the 23 programme budgeting categories and attaches the provider's average reference cost for the relevant HRG to each record. For each PCT this information generates a split of inpatient care expenditure by programme budget category for each of its secondary healthcare providers.

There are numerous difficulties faced when attempting to allocate non-admitted patient care activity (that is, outpatients, community services, direct access, A&E etc) to programme budget categories. The difficulties are primarily due to the absence of clear diagnostic codes. The 'primary reason for care' (equivalent to a diagnosis code) is not information that is routinely collected for community patients. Because of this, the approach prescribed is for service providers to produce a generic allocation analysis/report, for all PCTs making use of their services, for all non-admitted patient care costs across the 23 programme budget categories. Once derived, this generic allocation analysis/report is made available to PCTs at the same time as the unique (PCT specific) inpatient care information described above. Unlike the first apportionment report relating to admitted patient care, the non-admitted patient care apportionment report will not be unique to the PCT, but will represent the provider's overall experience. PCTs are expected to use this data to inform the apportionment of their own spend on non-admitted patient care across the 23 programme budget categories.

The Department of Health recognises that this approach – the provision of a PCT specific breakdown of admitted patient care costs and a generic allocation of all PCTs non-admitted patient care spend by providers – is likely to generate a crude method for apportioning non-admitted patient care costs. PCTs and their providers are therefore encouraged to put in place other arrangements that allow a more sophisticated analysis of non-admitted patient care expenditure. Such arrangements may well rely on an activity sampling approach [16].

Mental Health providers may not need to complete and forward detailed admitted and non-admitted patient care apportionment reports to PCTs. The nature of the services they provide may be such that the entire spend with them relates exclusively to the Mental Health programme (budget category 5). Ambulance Trusts are required to provide non-admitted patient care information to those PCTs for whom they provide services. Where it is not possible to split the activity by PCT, a generic non-admitted patient care report is produced for all purchasers [16].

The Department of Health has been criticised for the rather simplistic way in which it has apportioned certain costs among categories, and there are obvious issues with the allocation of costs associated with patients who have multiple disorders. However, the programme budgeting project is very much work-in-progress and the Department is investigating ways to improve the accuracy with which costs are allocated across programmes (for example, the Department is investigating the possibility of allocating training expenditures to specific programmes rather than to the generic medical training programme PBC 23b).[4]

B4.3 Programme budgeting expenditure, 2003/4 - 2008/9

National (all PCT) expenditure per head and the growth in this expenditure are shown for each programme budget category for 2003/04 to 2008/09 in Table B4.1. Comparable data for each programme budget sub-category is shown in Table BA.1 in the annex. Year on year comparisons of expenditure in each group are complicated by the fact that the algorithms used to allocate activity to PBCs are regularly revised. For example, for 2006/7 two major changes were made to the methods employed to construct the programme budgeting data. First, expert medical opinion was employed to re-evaluate the existing mapping from inpatient diagnosis codes to programme budget category. This led to the re-assignment of just over 10% of all diagnosis (ICD10) codes from one programme budgeting category to another.[5] [6] Second, activity to be costed used the newly introduced version 4 of the Healthcare Resource Group (HRG) software which, among other things, changed the methodology for calculating non-admitted patient care costs. HRG4 reflected advances in clinical practice and was designed to generate a much more accurate costing of complex cases. Other developments, such as the transfer of responsibility for dental funding from local dental boards to PCTs, also complicate the interpretation of comparisons through time (for example, per capita dental expenditure by PCTs increased from £13.55 in 2004/5 to £51.93 in 2006/7).

The expenditure figures for the first year (2003/4) are calculated on a slightly different basis to those for the other years (2004/5-2008/9). In particular, the figures for 2003/4 are on a 'net expenditure' basis while the figures for 2004/5-2008/9 are on an 'own population' basis. The 'own population' figure starts with net expenditure; it adds any expenditure funded from sources outside of the NHS; and then deducts any expenditure on other PCTs' populations incurred through lead/host commissioning arrangements. In 2006/7 and across all PBCs, expenditure per head on an own population basis was 2.3% greater than expenditure on a net population basis.

In 2004/5 total PCT expenditure per person was £1,200. The category attracting the most expenditure was the 'other' category (programme budget category 23) with per capita expenditure of almost £158 (13.2%). This category included primary care expenditure, workforce training expenditure, and a range of

[4] Some commentators have suggested that some of the within programme variation in expenditure observed across PCTs reflects different accounting conventions or unknown local factors. One way of reducing the impact of such unobserved heterogeneity is to construct a longitudinal data set with expenditure and mortality for each PCT for several years. With the availability of several years of data for both expenditure and mortality, we wanted to estimate a panel data model. However, most of the instruments employed here are based on the 2001 Census and thus estimation of a panel model will not be possible until these too become time variant; this should occur later this year with release of the 2011 Census data at PCT level. The same difficulty arises with the estimation of an incremental model.

[5] This figure ignores intra-category changes (for example, where an ICD10 code is re-allocated from category 1A to 1B) and only counts cross-category changes (for example, where the code is switched from category 1 to category 2).

[6] This expert review also led to the introduction of 40 additional sub-categories including 10 sub-categories for the cancer and tumour programme.

other miscellaneous expenditure items. Of these components, primary care expenditure was by far the largest element at £127 per head.

In 2004/5 there were two other categories with a budget share of over 10%: mental health (budget category 5) attracted 12.2% of expenditure (£147 per person), and circulation problems (budget category 10) recorded 10.2% of expenditure (£122 per person). Seven programme budget categories – cancers and tumours (£76), gastro-intestinal problems (£73), trauma and injuries (£72), musculo-skeletal problems (£72), respiratory problems (£63), genito-urinary problems (£62), and maternity and reproductive conditions (£55) – had expenditure shares of between 4.6% and 6.3%. Finally, the 13 remaining categories – from hearing problems (£6) to learning disability (£43) – each account for between 0.5% and 3.6% of total expenditure.

By 2008/9 total PCT expenditure per person had increased to £1,531 (up 28% from 2004/5). The residual 'other' category (programme budget category 23) still accounted for the largest share of expenditure (14.9%) with per capita expenditure of almost £228, of which £145 was accounted for by primary care expenditure. Mental health (budget category 5) still accounted for just over 12% of expenditure, but the expenditure share recorded by circulation problems (budget category 10) had fallen from 10.2% to 8.5%. Other categories recording a fall in budget share of more than one half of one percentage point included: the gastro-intestinal system (down from 6.1% to 5.1%), the musculo-skeletal system (down from 6% to 5.2%), trauma and injuries (down from 6% to 4.2%), and maternity (down from 4.6% to 3.9%).

Categories recording an increase in budget share of more than one half of one percentage point included neurological problems (up from 2.9% to 4.4%) and dental problems (up from 1.1% to 4.1%).

Some of these changes will partly reflect revisions to the algorithms used to allocate expenditure to particular PBCs. For example in 2006/7 expenditure per person on musculo-skeletal problems fell by 11% and expenditure on trauma and injuries fell by 25%. In the same year, expenditure on neurological problems increased by 35%. This suggests that some types of activity, which were previously allocated to musculo-skeletal problems and/or trauma and injuries, were re-allocated to neurological problems. Similarly, up to and including 2006/7, expenditure that was not directly attributable to a particular programme category was apportioned using admitted patient care percentages.[7] In other words, if x% of total admitted patient care expenditure was allocated to PBC 1, then x% of all expenditure that was not directly attributable to a particular programme category was also allocated to PBC 1. With effect from 2007/8, however, NHS organisations were asked to select an appropriate basis for the apportionment of this non-programme specific expenditure and that, where no reasonable basis existed, such expenditure was to be allocated to the 'Other – Miscellaneous' (PBC 23X) category.

These two changes to the algorithm used to allocate expenditure to particular PBCs illustrate that year-on-year comparisons of expenditure need to be interpreted with care.

Obviously, expenditure per head on any given programme varies from one PCT to another and Table B4.2 presents some statistics that indicate the degree of variation in expenditure levels across PCTs by programme budget category. The first four columns of Table B4.2 present descriptive statistics for PCT expenditure per person by PBC. These reveal that, for example, PCT per capita expenditure in the cancer programme averaged £96.30 across all PCTs, with the minimum spend being £62.90 and the maximum being £155.70.

Some PCTs will be spending more than other PCTs simply because they face higher input costs. The second set of four columns in Table B4.2 present descriptive statistics for PCT per capita expenditure

[7]Expenditure on, for example, community care, A&E, ambulance services, and outpatients can be difficult to attribute a particular PBC. Critical care, rehabilitation, and specialised commissioning across care settings will also be difficult to attribute to a particular programme.

that has been adjusted for the unavoidable geographical variation in costs (input prices) faced by PCTs.[8] However, if anything this adjustment appears to increase the variation in expenditure across PCTs; for example, the range of per capita expenditure on cancer increases from between £62.90 and £155.70 (unadjusted) to between £59.10 and £163.10 (adjusted for local health care input prices).

Another cause of the variation in expenditure levels will be the fact that the need for health care will vary from one PCT to another. For example, areas with a relatively large proportion of elderly residents, or PCTs operating in relatively deprived locations, can be expected to experience relatively high levels of spending. The Department of Health has a well-developed methodology for estimating the relative health care needs of PCTs, which it uses as the basis for allocating health care funds to PCTs[17]. Recent 'needs' formulae have been derived from an adjustment for the demographic profile of the PCT and a series of econometric analyses of the link between health care expenditure and other socio-economic factors at a small area level within England[18].

The final set of four columns in Table B4.2 present descriptive statistics for PCT per capita expenditure that has been adjusted for both the unavoidable geographical variation in costs and the local need for health care faced by PCTs.[9] For virtually every PBC, this adjustment reduces the variation in expenditure across PCTs; for example, the standard deviation of PCT per capita expenditure falls from £19.70 to £15.30 for the cancer programme. Although this adjustment reduces the variation in expenditure levels across PCTs, this decline is quite modest and there are still substantial differences in expenditure even after allowing for differences in local cost and need. For example, expenditure per head in the circulation problems category varies between £78 and £328 using cost adjusted expenditure data, but falls between £76 and £327 using cost and need adjusted population data.

This variation in expenditure across PCTs has led some commentators to question the reliability of the programme budgeting data. In a good governance report, the National Audit Office [19] sought to '…examine the quality, timeliness and suitability of Programme Budgeting data to support [their] audit of the Department of Health Resource Account and determine whether the systems and processes in place to provide the data are accurate.' The NAO undertook a survey of Trusts, PCTs and SHAs. The NAO noted that a number of PCTs expressed concern about the accuracy of data supplied to them by their service providers and noted that this was believed to be because most Trusts did not use or find the data they supply to PCTs of any use to themselves. Overall, the NAO's main conclusion was that while the processes for collecting the budgeting data were well defined in most areas, there remained scope for improvement to the robustness of some of the data (such as the non-admitted patient care data).

Appleby, Harrison, Foot, Smith and Gilmour [20] also considered the issue of data reliability in their study of variations in PCT spending on cancer services. They noted a rather dramatic variation in spending across PCTs for any given year, and that a relatively large number of PCTs report relatively large year-on-year changes in cancer expenditure. However, and as the authors point out, it is difficult to define what might be either an implausible level of expenditure or an implausibly large change in expenditure. Moreover, the interpretation of a large change in expenditure is complicated by the fact that the Department of Health makes regular changes (improvements) to the algorithm used to allocated activity to programme budget categories (as detailed above).

As a case study of the reliability of the programme budgeting data, Appleby et al [20] report the results of West Kent PCT's use of an alternative approach to producing the programme budgeting data for cancer and tumours. This alternative approach identified similar levels of expenditure to the traditional method at the aggregate level, but there were differences between the two approaches at the sub-programme level (that is, for expenditure on specific cancer sites and in the residual 'other cancers' category).

---

[8] This cost adjustment reflects the fact that health economy input prices vary considerably across the country and, for some inputs, are up to 40% higher in London and the south east of England than elsewhere. We have used a weighted average of the three Market Forces Factor Indices (MFFs) for HCHS, for prescribing, and for GMS/PMS to adjust the raw expenditure figures in Table 4.2 for local input prices (see Department of Health, 2009).

[9] This needs adjustment incorporates the AREA resource allocation formula for HCHS (see Department of Health, 2005c).

As with any dataset, there are likely to be recording and other errors associated with the programme budgeting data. However, there is no evidence on the magnitude of such errors and we have no reason to believe that such errors are likely to bias our estimates in one particular direction (for example, we have no reason to believe that measurement errors are systematically related to other relevant factors such as mortality rates). In this study, our focus is on whole programme expenditure and thus we avoid the data reliability issues inherent in any analysis of the sub-programme expenditure data.[10] Moreover, although we present estimates of the cost of a life year for individual programmes, our primary focus is on the cost of a life year across *all* programmes combined. The advantage of this is that the impact of a PCT reporting, for example, too little expenditure in one category might be offset by reporting too much expenditure in another.

While we note that the allocation of expenditure might not be consistent across PCTs there is no systematic evidence that the magnitude of any inconsistency is sufficiently large to cause concern. Accordingly, for each disease category, the observed variation in expenditure per person – holding constant input prices and the need for health care - offers the opportunity to examine whether PCTs that spend more on health care achieve a better outcome and, if so, at what cost. Empirical estimates of the strength of this relationship for both individual and all programmes of care are presented later in this report.

[10] The ACCA/Audit Commission (2011) looked at the reliability of the programme budgeting data for the diabetes sub-group within the endocrine and metabolic problems category. The ACCA/Audit Commission noted that programme budgeting data includes inpatient and prescribing expenditure, which are thought to be relatively reliably allocated to PBCs and to be consistently costed across PCTs, and outpatient and community service expenditure, which are thought to be less reliably allocated to PBCs and to be less consistently costed across PCTs. The ACCA/Audit Commission compared the variation in expenditure for inpatient and prescribing expenditure with that for total programme budget expenditure and found that the latter was far greater than the former. However, the interpretation of this result is not straightforward: as the ACCA/Audit Commission noted, it is difficult to know whether differences in programme budget spend are attributable to variation in service provision and efficiency, or simply to different approaches to cost allocation.

Table B4.1: table showing national (all PCT) expenditure per head (£) and growth in expenditure (%) by PBC group, 2003/4 - 2008/9

PBC #	PBC description	Spend (£) per head 2003/4	Spend (£)per head 2004/5	Spend (£)per head 2005/6	Spend (£)per head 2006/7	Spend (£)per head 2007/8	Spend (£)per head 2008/9	Growth (%) 2004/5	Growth (%) 2005/6	Growth (%) 2006/7	Growth (%) 2007/8	Growth (%) 2008/9	Share of total spend (%) 2004/5	Share of total spend (%) 2008/9
1	Infectious diseases	17.95	20.22	23.61	20.88	22.08	23.46	13	17	-12	6	6	1.7%	1.5%
2	Cancers and tumours	64.95	75.54	83.24	81.67	90.21	94.55	16	10	-2	10	5	6.3%	6.2%
3	Blood disorders	14.08	17.00	17.48	16.58	19.44	19.50	21	3	-5	17	0	1.4%	1.3%
4	Endocrine, nutritional	28.96	31.86	37.26	36.70	39.39	43.38	10	17	-1	7	10	2.7%	2.8%
5	Mental health	133.31	146.83	158.95	166.53	180.90	191.21	10	8	5	9	6	12.2%	12.5%
6	Learning disability	37.93	43.37	46.54	48.36	54.20	56.11	14	7	4	12	4	3.6%	3.7%
7	Neurological	29.83	35.09	41.06	55.27	62.43	67.64	18	17	35	13	8	2.9%	4.4%
8	Vision problems	24.61	27.65	28.24	26.97	30.69	32.95	12	2	-4	14	7	2.3%	2.2%
9	Hearing problems	5.73	6.32	6.27	6.21	8.07	8.16	10	-1	-1	30	1	0.5%	0.5%
10	Circulatory disease	110.12	122.37	124.28	122.06	124.77	129.94	11	2	-2	2	4	10.2%	8.5%
11	Respiratory system	54.60	62.71	69.56	65.07	67.68	77.97	15	11	-6	4	15	5.2%	5.1%
12	Dental problems	10.78	13.55	24.91	51.93	59.45	62.44	26	84	108	14	5	1.1%	4.1%
13	Gastro intestinal system	63.56	73.22	81.30	73.30	75.05	77.89	15	11	-10	2	4	6.1%	5.1%
14	Skin problems	20.98	24.90	26.84	28.31	30.41	32.34	19	8	5	7	6	2.1%	2.1%
15	Musculo Skeletal system	61.36	71.72	74.74	66.75	75.91	79.68	17	4	-11	14	5	6.0%	5.2%
16	Trauma and Injuries	62.31	72.13	76.41	57.29	57.56	63.54	16	6	-25	0	10	6.0%	4.2%
17	Genito Urinary system	55.32	62.38	67.38	68.98	67.83	73.78	13	8	2	-2	9	5.2%	4.8%
18	Maternity	52.28	55.04	60.42	57.64	57.09	60.44	5	10	-5	-1	6	4.6%	3.9%
19	Neonate conditions	11.72	13.93	13.42	13.17	15.15	17.23	19	-4	-2	15	14	1.2%	1.1%
20	Poisoning	9.68	12.32	14.25	14.59	15.84	18.31	27	16	2	9	16	1.0%	1.2%
21	Healthy individuals	20.29	22.77	26.18	26.85	31.44	35.74	12	15	3	17	14	1.9%	2.3%
22	Social care needs	24.81	30.93	33.59	30.29	35.29	36.58	25	9	-10	17	4	2.6%	2.4%
23	Other (includes GMS/PMS)	136.94	157.75	171.82	209.70	232.02	227.71	15	9	22	11	-2	13.2%	14.9%
1 to 23	All PBCs	1052.12	1199.60	1307.76	1345.10	1452.91	1530.59	14	9	3	8	5		

Notes: (i) The population figures for 2003/4, 2004/5 and 2005/6 are identical (the total for England is 49,175,998).
(ii) The corresponding figure for 2006/7 is 50,476,231, for 2007/8 it is 50,695,989, and for 2008/9 it is 51,220,531.
(iii) The spend per head figures are calculated by summing expenditure across all PCTs and dividing by the national population.
(iv) All figures are at current prices.

Table B4.2: table showing PCT expenditure per head by PBC, 2008/9: (a) unadjusted; (b) adjusted for local costs; and (c) adjusted for local costs and local need

		Spend per head (unadjusted), £				Spend per head (cost adjusted), £				Spend per head (cost and need adjusted), £			
	Programme budget category	Mean	StdDev	Min	Max	Mean	StdDev	Min	Max	Mean	StdDev	Min	Max
1	Infectious diseases	26.5	24.6	8.6	151.8	25.7	21.7	8.6	136.7	25.0	21.4	9.5	139.5
2	Cancers and tumours	96.3	16.9	62.9	155.7	96.7	19.7	59.1	163.1	94.2	15.3	55.2	154.0
3	Blood disorders	20.3	7.0	7.7	49.4	20.2	6.5	8.0	49.1	19.7	6.0	8.2	44.2
4	Endocrine, nutritional	44.6	8.8	28.9	74.8	44.7	9.5	27.4	77.0	43.3	6.1	29.9	61.5
5	Mental health	201.4	60.0	118.9	474.1	200.3	54.0	122.8	422.8	194.0	41.9	132.3	362.0
6	Learning disability	56.8	18.8	7.7	125.9	57.0	19.4	6.8	123.6	55.7	18.8	6.7	136.6
7	Neurological	68.5	13.8	41.1	133.8	68.8	15.6	38.4	137.5	66.9	12.1	41.5	125.2
8	Vision problems	33.2	6.7	16.7	57.7	33.4	7.5	14.8	59.2	32.5	6.1	15.6	48.3
9	Hearing problems	8.6	3.7	0.9	24.0	8.7	3.9	0.9	25.5	8.3	3.3	0.8	22.0
10	Circulatory disease	131.6	26.7	88.0	317.3	132.2	30.5	78.2	327.6	128.5	24.4	75.7	326.9
11	Respiratory system	80.5	17.4	48.0	141.2	80.9	19.8	42.7	145.3	78.1	12.4	48.2	126.0
12	Dental problems	64.8	13.4	28.0	111.9	64.9	14.1	24.9	115.8	63.0	10.7	28.1	97.1
13	Gastro intestinal system	80.0	14.5	46.7	119.6	80.4	16.8	41.5	124.6	78.0	11.3	41.6	114.4
14	Skin problems	33.1	8.0	18.1	66.4	33.3	8.6	16.5	69.1	32.2	6.3	16.0	57.7
15	Musculo Skeletal system	79.9	17.6	43.3	127.3	80.4	19.9	39.6	132.5	78.2	16.6	41.0	116.4
16	Trauma and Injuries	63.2	16.7	12.5	139.3	63.4	17.4	11.5	125.0	61.8	15.6	10.4	103.6
17	Genito Urinary system	75.7	13.7	49.9	112.3	75.6	13.6	48.4	108.9	73.7	10.1	50.6	105.5
18	Maternity	63.3	16.7	24.6	124.4	63.1	15.8	21.9	117.9	61.4	12.8	24.4	96.5
19	Neonate conditions	18.4	7.3	6.4	46.4	18.2	6.8	6.6	43.7	17.8	6.6	5.8	47.8
20	Poisoning	18.6	4.2	10.8	31.2	18.7	4.7	9.6	32.3	18.2	3.9	10.1	33.1
21	Healthy individuals	38.4	18.1	9.7	125.0	38.4	17.8	8.9	115.6	36.7	14.5	9.4	104.5
22	Social care needs	40.8	56.6	0.1	415.2	41.2	59.2	0.1	432.9	39.7	55.0	0.0	411.5
23	Other (includes GMS/PMS)	230.8	44.5	138.2	396.1	230.2	42.4	140.7	356.5	226.8	45.8	134.1	346.0
All	All PBCs	1,575.6	196.7	1,225.7	2,079.9	1,576.3	217.3	1,183.0	2,173.1	1,534.0	86.2	1,390.1	1,987.0

Note: the above statistics relate to 152 PCTs and the mean expenditure figures will differ slightly from the national ones in Table B4.1 because the statistics across PCTs are not weighted for the size of each PCT's population.

B5. Health outcome and other data

B5.1 Health outcome data

Most studies of the relationship between expenditure and outcome have used some measure of mortality as an indicator of the latter. We too employ mortality as our outcome measure for two reasons: first, it is a relevant (but admittedly not comprehensive) measure of the outcome of health care expenditure; and second, it is available for more disease areas than any other outcome measure at PCT level.

Although mortality is available (by PCT) for several disease areas, it is not available for just over one-half of all programmes not least because it is simply not relevant for these programmes (e.g., for learning disabilities, vision problems, hearing problems, dental problems, and skin problems). Moreover, even where a mortality measure is available, the ICD10 coverage of the mortality data often falls short of the coverage of the expenditure data. For some programmes, therefore, we have combined the published mortality rates for two or more disease areas in an attempt to match the ICD10 coverage of the mortality data with that of the expenditure data.

Table B5.1 shows how we have attempted to marry the mortality data (column c) and the expenditure data (column a). However, and as Table B5.1 shows, the ICD10 coverage of the component mortality rates for some PBCs still falls short that of the expenditure data and the extent of this shortfall is illustrated by the ratio reported in the final column of Table B5.1. For example, the cancers and tumours programme covers all expenditure associated with ICD10 codes C00-C97 and D00-D49 but the PCT-based mortality data only relates to ICD10 codes C00-C97. At the national (all England) level, figures are available which show that, in 2008, there were 62,072 deaths of those aged under 75 years from codes C00-C97 and that there were 63,076 deaths from codes C00-C97 and D00-D49 combined. In other words, the PCT level mortality data reflects 98.4% of all deaths associated with the expenditure codes. Initially, we did not adjust our cost of life (year) estimates for this mismatch but, as we will see in section B8.6, an adjustment has been made for this mismatch in the final calculation of the cost of a life (year) associated with expenditure for 2006/7. The same adjustment has also been applied to the cost of a life (year) estimates associated with expenditure for 2007/8 and for 2008/9.

Of course, we acknowledge that mortality is a more relevant outcome indicator for some programmes (e.g., for circulatory problems) than for others (e.g., for epilepsy) and, for this reason, we would expect better results in some programmes than others. We also acknowledge that this focus on mortality ignores the impact of expenditure aimed at chronic care and at palliative care. Nevertheless, our focus on mortality is purely practical: it is both a widely available measure and it is clearly a relevant outcome indicator. Moreover, the approach adopted here is extendable in principle to other non-mortality based outcome indicators. We illustrate such an application in section B8.8 where we use EQ-5D utility scores pre- and post- an operative procedure from the PROMs programme to generate a non-mortality-based outcome indicator, and we use this indicator to estimate our outcome model.

Table B5.1: table showing ICD10 coverage of the expenditure and outcome measures

	ICD 10 coverage of programme budgeting category	Number of deaths, <75years, 2008, England (ONS, VS3) corresponding to column aICD10 codes	ICD 10 coverage of best match PCT based mortality rate(s)	Number of deaths, <75years, 2008, England (ONS, VS3) corresponding to column c ICD10 codes	Ratio (d/b)
	column a	column b	column c	column d	column e
PBC 1	Infectious diseases (large parts of A00-B99)*	1,968	Infectious diseases (A00-B99)	1,968	1.000
PBC 2	Cancers and tumours (C00-C97, D00-D49)	63,076	All cancers (C00-C97)	62,072	0.984
PBC 3	Blood disorders (D500-D899)	393	No relevant mortality rate by PCT available	n/a	n/a
PBC 4	Endocrine, nutritional and metabolic problems (E000-E899)	2,368	Diabetes (E10-E14)	1,501	0.634
PBC 5	Mental health (F00-F69, Z55, Z56)	n/a	No relevant mortality rate available	n/a	n/a
PBC 6	Learning disability (F700-F739, F780-F849, F88-F90, Q90, Q91)	n/a	No relevant mortality rate available	n/a	n/a
PBC 7	Neurological system (G000-G999, Q000-Q079, R200-R999)	5,238	Epilepsy* (G40-G41)	713	0.136
PBC 8	Eye and vision problems (H000-H599, Q100-Q159)	n/a	No relevant mortality rate available	n/a	n/a

ICD 10 coverage of programme budgeting category	Number of deaths, <75years, 2008, England (ONS, VS3) corresponding to column a ICD10 codes	ICD 10 coverage of best match mortality rate(s)		Number of deaths, <75years, 2008, England (ONS, VS3) corresponding to column c ICD10 codes	Ratio (d/b)
column a	column b	column c		column d	column e
PBC 9 Hearing problems (H600-H999, Q160-Q179)	n/a	No relevant mortality rate available		n/a	n/a
PBC10 Circulation problems (I00-I99, Q20-Q28)	39,923	Circulatory diseases (I00-I99)		39,590	0.992
PBC11 Respiratory problems (A150-A169,* A190-A199, J000-J989, Q300-Q349, R000-R099)	14,417	Asthma (J45-J46)		382	
		Bronchitis, emphysema, other COPD (J40-J44)	7,174	11,147	0.773
		Pneumonia (J12-J18)	3,591		
PBC12 Dental problems (K000-K099)	n/a	No relevant mortality rate available		n/a	n/a
PBC13 Gastro-intestinal problems (I840-I859, K091-K929, Q380-Q459, R100-R198)	10,656	Liver disease* (K70, K73-K74)		5,195	
				6,082	0.571
		Ulcers (K25-K27)	887		
PBC14 Skin problems (L000-L999, Q351-Q379, Q800-Q859)	367	No relevant mortality rate available		n/a	n/a
PBC15 Musculo-skeletal problems (M00-M99, Q18, Q650-Q799)	933	No relevant mortality rate available		n/a	n/a

ICD 10 coverage of programme budgeting category	Number of deaths, <75years, 2008, England (ONS, VS3) corresponding to column a	ICD 10 coverage of best match mortality rate(s) ICD10 codes		Number of deaths, <75years, 2008, England (ONS, VS3) corresponding to column c	Ratio (d/b) ICD10 codes
column a	column b	column c		column d	column e
PBC16 Trauma, burns and injuries (S000-S999, T000-T357, T79, T90-T98)	5,809*	Fracture of thighbone* (S72)	174	] 1,014	0.175
		Skull, cranial injury* (S02, S06, T90)	840		
PBC17 Genito-urinary problems (A50-A64, N00-N99, Q500-Q649, R30-R39, R86-R87)	1,565	Chronic renal failure* (N18)	269		0.172
PBC18 Maternity and reproductive problems (N96-N98, O000-O999, Z300-Z391)	41	No relevant mortality rate available	n/a	┐ | | 2,193	8.213
PBC19 Neonate conditions (P000-P299, P350-P399, P500-P619, P700-P839, P900-P969)	226	Infant mortality rate* per 1,000 live births, aged under 28 days (all ICD10 codes)	2,152	| ┘	but see note below
PBC20 Poisoning (Q86, R78, R82, T360-T888)	n/a	No relevant mortality rate available	n/a		n/a
PBC21 Healthy individuals	n/a	No relevant mortality rate available	n/a		n/a
PBC22 Social care needs	n/a	No relevant mortality rate available	n/a		n/a
PBC23 Other areas	n/a	No relevant mortality rate available	n/a		n/a

Notes: (i) the listed ICD10 coverage of the programme budgeting expenditure data includes the major ICD10 codes covered.

(ii) the ICD10 coverage of PBC 1 includes large elements of codes A00-B99 but a substantial minority of these codes map to the respiratory (PBC 11) and gastro-intestinal (PBC 13) programmes. We do not have the detailed deaths data to remove them from the total for A00-B99 and then to add them to the respiratory and gastro-intestinal programmes. Instead, we acknowledge that the number of deaths attributed to PBC 1 will be overstated (and that the adjustment ratio in column e will be too low), and that the number of deaths attributed to PBCs 11 and 13 will be understated (and that their adjustment ratios in column e will be too high) but this is the best that can be achieved given the available data.

(iii) the ICD10 coverage of the all England mortality data does not always match precisely that of the expenditure data or the PCT level mortality data; again, we have done the best that can be achieved given the available data. In particular: the national epilepsy mortality data relates to ICD10 G40 (687 deaths) but the PCT level data relates to G40 and G41 (annual average over 2007/8/9 is 713 deaths) ; the national renal failure mortality data relates to ICD10 N17-N19 (415 deaths) but the PCT level data relates to N18 (annual average over 2007/8/9 is 269 deaths); the national liver disease mortality data relates to ICD10 K70-K77 (6,020 deaths) but the PCT level data relates to K70, K73-K74 (annual average over 2007/8/9 is 5,195 deaths); and there is no good ICD10 match for femur and skull fracture deaths using national VS3 data (the PCT level data relates to S72, S02, S06, T90: annual average over 2007/8/9 is 1,014 deaths). For these four cases we use the annual average number of deaths over 2007/8/9 from the PCT-level data as the numerator when calculating the coverage adjustment factor (column e).

(iv) the number of deaths in England for those aged under 75 years for the trauma, burns and injuries programme (column b) relates to 2004 and is for the secondary cause of death (Martin, Rice and Smith, 2012).

(v) the mortality rate for neonate conditions relates to deaths aged under 28 days for all ICD10 codes but the expenditure data relates only to 'P' ICD 10 codes. Hence the large adjustment factor of 8.213 because the coverage of the expenditure data is much smaller than that of the mortality data. However, at the very end of the project it became clear that although the number of deaths data for those aged under 75 years includes those dying at all ages under 75 years (including those at under 1 year), the disease specific years of life lost totals for those aged under 75 years excludes those dying at under 1 year of age and actually refers to those dying at ages 1 to 74 (the argument is that infant deaths are mostly a result of causes specific to the age and have different causes to disease specific deaths later in life). We therefore have two adjustment factors for the maternity and neonates programme: first, an adjustment factor for the number of deaths derived on the same basis as the adjustment factors for other programmes; and second an adjustment factor for the YLL that reflects both the YLL in the maternity and neonates programme, as well as the YLL associated with deaths that would have been attributed to other programmes had the individual died over 1 year of age. (NB The total number of deaths in England in 2008 of those aged under 1 year is 3,184 and if we divide 2,193 by (3,184+41) we obtain the YLL coverage adjustment factor (=0.679) for maternity and neonates.

(vi) the PCT level mortality rates are available from the NHS Information Centre website.

Previous studies using the programme budgeting data have employed two alternative mortality based outcome indicators: the under 75 years of age standardised mortality rate (SMR) and the under 75 years standardised years of life lost rate (SYLLR). The SMR gives equal weight to all deaths irrespective of the age at which they occur but the SYLLR gives greater weight to deaths that occur at earlier ages.

We employed both the SMR and the SYLLR when undertaking some preliminary sensitivity analysis (i.e., in section B8.2 when considering, for example, which measure of need to use), but elsewhere we have focussed solely on a measure of the avoidable years of life lost (YLL).[11] This is calculated by summing over ages 1 to 74 years the number of deaths at each age multiplied by the number of years of life remaining up to age 75 years. The crude YLL rate is simply the number of years of life lost divided by the resident population aged under 75 years. Like conventional mortality rates, the crude YLL rate can be age standardised to eliminate the effects of differences in population age structures between areas, and this (age) standardised YLL rate is the health outcome variable generally employed in this study (Lakhani et al., 2006, p379).

Descriptive statistics for the SYLLRs employed in this study are shown in Table B5.2. For example, for all deaths over the three year period from 2006 to 2008, the annual SYLLR across all PCTs for those aged under 75 years averaged 467 years of life lost per 10,000 population, but this rate varied considerably across PCTs, ranging between 288 and 749 years of life lost per 10,000 population. Similarly large variations in the mortality rate across PCTs are evident for other disease groups.[12]

[11] One exception to this is the mortality rate for the trauma and injuries programme where initially only SMRs were available.

[12] The NHS IC reports mortality rates using deaths pooled over a three year period because the relatively small number of annual deaths in some disease categories might lead to large year-on-year fluctuations in death rates at PCT level.

Table B5.2: table showing descriptive statistics for the mortality variables

Variable	Obs	Mean	StdDev	Min	Max
all causes of death, SYLLR, 2002/3/4	303	489.2	94.2	320.3	889.5
all causes of death, SYLLR, 2004/5/6	152	483.4	83.9	318.1	742.5
all causes of death, SYLLR, 2006/7/8	152	467.3	83.7	287.8	748.9
all causes of death, SYLLR, 2007/8/9	151	457.1	81.8	297.2	731.6
all causes of death, SYLLR, 2008/9/10	151	446.4	78.6	290.8	736.9
cancer, SYLLR, 2002/3/4	303	161.9	20.8	115.6	263.4
cancer, SYLLR, 2004/5/6	152	158.4	18.3	103.4	218.8
cancer, SYLLR, 2006/7/8	152	154.2	19.0	90.5	212.2
cancer, SYLLR, 2007/8/9	151	151.0	18.5	98.3	201.9
cancer, SYLLR, 2008/9/10	151	147.9	17.5	100.2	193.9
circulatory disease, SYLLR, 2002/3/4	303	114.4	31.3	57.7	225.7
circulatory disease, SYLLR, 2004/5/6	152	108.6	25.2	65.2	177.8
circulatory disease, SYLLR, 2006/7/8	152	99.0	23.7	54.4	156.7
circulatory disease, SYLLR, 2007/8/9	151	94.4	22.6	51.4	149.9
circulatory disease, SYLLR, 2008/9/10	151	91.1	21.7	50.9	154.8
asthma, SYLLR, 2002/3/4	303	2.7	2.0	0.0	12.2
asthma, SYLLR, 2004/5/6	152	2.4	1.3	0.1	6.3
asthma, SYLLR, 2006/7/8	152	2.0	1.1	0.0	5.0
asthma, SYLLR, 2007/8/9	151	1.9	1.1	0.0	5.7
asthma, SYLLR, 2008/9/10	151	1.7	1.1	0.0	4.6
bronchitis, emphysema, other COPD, SYLLR, 2002/3/4	303	12.5	5.7	2.6	35.5
bronchitis, emphysema & other COPD, SYLLR, 2004/5/6	152	12.0	4.8	3.7	26.1
bronchitis, emphysema & other COPD, SYLLR, 2006/7/8	152	12.0	4.8	4.0	24.4
bronchitis, emphysema & other COPD, SYLLR, 2007/8/9	151	11.8	4.7	4.1	24.8
bronchitis, emphysema & other COPD, SYLLR, 2008/9/10	151	11.6	4.9	4.2	26.6
pneumonia, SYLLR, 2002/3/4	303	9.1	4.1	1.4	24.6
pneumonia, SYLLR, 2004/5/6	152	9.7	3.7	3.6	21.9
pneumonia, SYLLR, 2006/7/8	152	9.7	3.9	3.6	32.4
pneumonia, SYLLR, 2007/8/9	151	9.8	4.0	3.9	34.4
pneumonia, SYLLR, 2008/9/10	151	9.3	4.0	2.8	36.1
tuberculosis, SYLLR, 2002/3/4	n/a				
tuberculosis, SYLLR, 2004/5/6	152	0.8	1.1	0.0	5.2
tuberculosis, SYLLR, 2006/7/8	152	0.8	1.0	0.0	7.6
tuberculosis, SYLLR, 2007/8/9	n/a				
tuberculosis, SYLLR, 2008/9/10	n/a				
respiratory problems, SYLLR, 2002/3/4 (exc TB)	303	24.3	9.7	5.4	64.2
respiratory problems, SYLLR, 2004/5/6 (inc TB)	152	24.9	8.9	9.7	51.7
respiratory problems, SYLLR, 2006/7/8 (inc TB)	152	24.6	8.5	11.3	56.4
respiratory problems, SYLLR, 2007/8/9 (exc TB)	151	23.4	8.1	8.5	57.4
respiratory problems, SYLLR, 2008/9/10 (exc TB)	151	22.6	8.5	8.5	65.0
liver disease, SYLLR, 2002/3/4	303	20.1	10.0	3.6	70.9
liver disease, SYLLR, 2004/5/6	152	22.9	9.9	8.2	75.0
liver disease, SYLLR, 2006/7/8	152	23.9	10.8	7.0	81.7

liver disease, SYLLR, 2007/8/9	151	23.7	10.6	9.4	81.1
liver disease, SYLLR, 2008/9/10	151	23.5	9.9	8.4	77.4
gastric, duodenal & peptic ulcers, SYLLR, 2002/3/4	303	2.6	1.6	0.0	10.2
gastric, duodenal & peptic ulcers, SYLLR, 2004/5/6	152	2.7	1.5	0.1	11.6
gastric, duodenal & peptic ulcers, SYLLR, 2006/7/8	152	2.4	1.3	0.5	8.5
gastric, duodenal & peptic ulcers, SYLLR, 2007/8/9	151	2.4	1.3	0.4	7.0
gastric, duodenal & peptic ulcers, SYLLR, 2008/9/10	151	2.3	1.4	0.4	7.6
gastro-intestinal problems, SYLLR, 2002/3/4	303	22.7	11.0	4.7	77.8
gastro-intestinal problems, SYLLR, 2004/5/6	152	25.6	10.7	9.3	80.3
gastro-intestinal problems, SYLLR, 2006/7/8	152	26.3	11.5	8.1	87.6
gastro-intestinal problems, SYLLR, 2007/8/9	151	26.1	11.1	10.7	86.3
gastro-intestinal problems, SYLLR, 2008/9/10	151	25.8	10.5	9.2	82.5
infectious diseases, SYLLR, 2002/3/4	303	7.0	4.2	0.1	28.1
infectious diseases, SYLLR, 2004/5/6	152	8.1	4.3	2.4	24.9
infectious diseases, SYLLR, 2006/7/8	152	8.3	4.4	0.6	26.1
infectious diseases, SYLLR, 2007/8/9	151	8.2	4.2	2.1	25.1
infectious diseases, SYLLR, 2008/9/10	151	7.7	4.0	1.6	22.6
diabetes, SYLLR, 2002/3/4	303	4.7	2.3	0.0	13.4
diabetes, SYLLR, 2004/5/6	152	4.5	2.1	1.3	15.3
diabetes, SYLLR, 2006/7/8	152	4.3	2.0	0.5	14.6
diabetes, SYLLR, 2007/8/9	151	4.0	1.8	0.3	11.2
diabetes, SYLLR, 2008/9/10	151	4.0	1.7	0.4	10.0
epilepsy, SYLLR, 2002/3/4	303	5.2	2.7	0.3	16.1
epilepsy, SYLLR, 2004/5/6	152	5.3	2.1	0.5	13.1
epilepsy, SYLLR, 2006/7/8	152	5.1	2.1	0.9	12.7
epilepsy, SYLLR, 2007/8/9	151	4.9	1.9	1.3	14.5
epilepsy, SYLLR, 2008/9/10	151	4.8	2.0	1.1	13.7
renal failure, SYLLR, 2002/3/4	303	0.9	0.9	0.0	6.0
renal failure, SYLLR, 2004/5/6	152	0.9	0.7	0.0	4.0
renal failure, SYLLR, 2006/7/8	152	0.8	0.7	0.0	5.5
renal failure, SYLLR, 2007/8/9	151	0.7	0.6	0.0	4.3
renal failure, SYLLR, 2008/9/10	151	0.6	0.6	0.0	3.0
fracture of femur (S72), SMR, 2002/3/4 (ages 65 to 84)	303	8.9	6.9	0.0	39.3
fracture of femur (S72), SMR, 2004/5/6 (ages 65 to 84)	152	10.1	6.6	0.0	30.6
fracture of femur (S72), SMR, 2006/7/8 (ages under 75)	152	0.4	0.3	0.0	1.4
fracture of femur (S72), SYLLR, 2007/8/9 (ages under 75)	151	0.3	0.3	0.0	1.7
fracture of femur (S72), SYLLR, 2008/9/10 (ages under 75)	151	0.3	0.3	0.0	2.1
skull fracture/injury, SMR, 2002/3/4 (ages under 75)	303	2.8	1.2	0.4	7.6
skull fracture/injury, SMR, 2004/5/6 (ages under 75)	152	1.9	0.8	0.4	4.4
skull fracture/injury, SMR, 2006/7/8 (ages under 75)	152	1.8	0.7	0.5	4.2
skull fracture/injury, SYLLR, 2007/8/9 (ages under 75)	151	1.7	0.7	0.2	4.2
skull fracture/injury, SYLLR, 2008/9/10 (ages under 75)	151	1.6	0.6	0.1	3.0
trauma, SMR, 2002/3/4 (w/average of femur and skull fractures)	303	4.8	2.4	0.3	15.3
trauma, SMR, 2004/5/6 (sum of femur and skull fracture rates)	152	12.0	6.8	1.9	32.8
trauma, SMR, 2006/7/8 (sum of femur and skull fracture rates)	152	2.1	0.8	0.6	4.7
trauma, SMR, 2007/8/9 (sum of femur and skull fracture rates)	151	2.1	0.8	0.2	4.6

trauma, SMR, 2008/9/10 (sum of femur and skull fracture rates)	151	1.9	0.8	0.1	4.4
infant mortality rate, <28 days per 1,000 live births, 2002/3/4	303	3.4	1.3	0.9	7.8
infant mortality rate, <28 days per 1,000 live births, 2004/5/6	130	3.4	0.9	1.2	6.2
infant mortality rate, <28 days per 1,000 live births, 2006/7/8	152	3.3	1.0	1.4	6.4
infant mortality rate, <28 days per 1,000 live births, 2007/8/9	151	3.2	1.0	1.2	6.9
infant mortality rate, <28 days per 1,000 live births, 2008/9/10	151	3.2	1.0	1.2	6.9

Note: the SYLLRs are directly age-standardised rates and are expressed as rates per 10,000 European Standard population. Source: NHS Information Centre website.

B5.2 Other variables

We employ an instrumental variable (IV) estimation technique to estimate our outcome and expenditure equations because (i) own programme expenditure is likely to be endogenous in the outcome equation and (ii) other programme need is likely to be endogenous in the own programme expenditure equation. IV estimation is described in section B6.2 but basically it involves replacing the endogenous variable in the equation of interest with its predicted value from an OLS regression which regresses the endogenous variable on a set of instrumental variables. These instruments should be good predictors of the endogenous variable (i.e., they should be relevant and strong predictors) but should be appropriately excluded from the equation of interest (i.e., they should be valid instruments).

We have a number of potential instruments available, mostly derived from 2001 Population Census. In our earlier studies we found that a small sub-set of these instruments proved sufficient to generate plausible results and these included:

- the proportion of the population providing unpaid care
- the proportion of households that are one pensioner households
- the index of multiple deprivation
- the proportion of the population in the white ethnic group.

We also had available a further set of potential instruments and, where our more limited set of instruments failed to generate plausible results, we extended our instrument search to include this wider set of variables. This extended set of instruments included:

- the proportion of residents born outside the European Union
- the proportion of the population of working age (16-74) with a limiting long term illness
- the proportion of the population aged 16-74 with no qualifications
- the proportion of the population aged 16-74 that are full-time students
- the proportion of households without a car
- the proportion of households that are owner occupied
- the proportion of households that are rented from a LA or HA
- the proportion of households that are rented from private landlords
- the proportion of households that are lone parent households with dependent children
- the proportion of the population aged 16-74 that are permanently sick
- the proportion of those aged 16-74 that are long-term unemployed
- the proportion of those aged 16-74 in employment that are working in agriculture
- the proportion of those aged 16-74 in managerial and professional occupations.

Details of the construction of all instruments are shown in Table BA.2 in the Annex.

Our instruments reflect factors, such as socio-economic deprivation and the availability of informal care in the community, which might indirectly impact upon mortality rates and/or health care expenditure levels. As we shall see, although our instruments 'pass' the appropriate statistical tests, some commentators claim that such tests may have 'low power' to detect the presence of invalid instruments. Consequently in section B9 we examine how sensitive our results are to the presence of invalid instruments.

Table B5.3 reports descriptive statistics for the socio-economic and needs variables as available for the regression analysis of programme budgeting (PB) expenditure data for 2007/8 and for 2008/9 (these statistics are for the variables in absolute form). For example, on average, lone pensioner households comprise 14% of all households, the 'white ethnic' group accounts for 89% of the population, and 10% of the population provide unpaid care.

In addition to the instrumental variables, Table B5.3 also report descriptive statistics for various other variables available for the regression analysis including the of Department of Health's 'need for health care' index (this incorporates the CARAN formula for HCHS and reflects need across all health care services), its need for HIV services index, and its need for maternity services index. The latter two indices are used to either supplement or replace the all service measure of need when estimating our models. The 'need for health care' index averages about 1 but varies substantially, with some PCTs having a needs index more than 25% below the national average and others facing a need for health care more than 30% above the national average.

Table B5.3 also reports descriptive statistics for some disease prevalence rates (e.g., for diabetes and for epilepsy) and, again, these are used to either supplement or replace the all service measure of need when estimating our models

Finally, the MFF index shows that input prices in the most expensive PCT are almost 20% above those in the least expensive PCT.

Table B5.3: table showing descriptive statistics for the instrumental and other variables

Description	Obs	Mean	Std. Dev.	Min	Max
Proportion of residents born outside the European Union	151	0.0794	0.0876	0.0088	0.3817
Proportion of population in white ethnic group	151	0.8927	0.1299	0.3942	0.9926
Proportion of population of working age (16-74) with LLT illness	151	0.1182	0.0250	0.0709	0.1798
Proportion of population providing unpaid care	151	0.0990	0.0118	0.0662	0.1221
Proportion of population providing unpaid care (<20 hrs week)	151	0.0667	0.0079	0.0461	0.0817
Proportion of population providing unpaid care (20-49 hrs week)	151	0.0113	0.0025	0.0065	0.0195
Proportion of population providing unpaid care (>50 hrs week)	151	0.0210	0.0051	0.0093	0.0353
Proportion of population aged 16-74 with no qualifications	151	0.2960	0.0642	0.1301	0.4555
Proportion of population aged 16-74 that are full-time students	151	0.0720	0.0270	0.0425	0.1626
Proportion of households without a car	151	0.2932	0.1046	0.1325	0.5761
Proportion of owner occupied households	151	0.6692	0.1128	0.2891	0.8205
Proportion of households in rented social (LA/HA) housing	151	0.2071	0.0918	0.0817	0.5356
Proportion of households in rented private housing	151	0.0924	0.0449	0.0349	0.2961
Proportion of lone pensioner households	151	0.1434	0.0184	0.0979	0.1942
Proportion of one parent households	151	0.0684	0.0180	0.0401	0.1207
Proportion of population aged 16-74 that are permanently sick	151	0.0574	0.0213	0.0242	0.1215
Proportion of population aged 16-74 are long-term unemployed	151	0.0113	0.0052	0.0036	0.0287
Proportion of 16-74 in employment that are in agriculture	151	0.0117	0.0119	0.0016	0.0668
Proportion of those aged 16-74 that are in professional occupations	151	0.2672	0.0688	0.1470	0.4958
Index of Multiple Deprivation 2007	151	23.8098	9.1168	8.0857	48.2627
Need index (incorporates CARAN formula)	151	1.0253	0.1334	0.7311	1.3479
MFF index for HCHS and prescribing	151	1.0021	0.0559	0.9410	1.1243
Diabetes prevalence rate 2007/8 (%, over 17 years)	151	5.4872	0.7982	3.22	8.51
Epilepsy prevalence rate 2007/8 (%, over 18 years)	151	0.7884	0.1489	0.41	1.09
HIV need index	151	1.1848	1.4984	0.1648	8.3332
Chronic kidney disease 2007/8 (%, over 18 years)	151	4.1687	1.2711	1.35	8.41
Maternity need index	151	1.0345	0.2106	0.6845	1.8129
Raw (unadjusted) population 2007/8	151	335,735	196,501	90,142	1,264,298

Note: these statistics are unweighted across PCTs and reflect the values for these variables as available for the regression analysis of PB expenditure data for 2007/8 and for 2008/9.
Sources: Population Census 2001, Department of Health (2009), NHS Information Centre website.

B6. Estimation issues and strategy

B6.1 Introduction

The theoretical framework suggests the specification and estimation of a system of equations, with an expenditure and health outcome equation for each of the 23 programmes of care. However, this approach makes infeasible data demands, requiring variables to identify expenditure, need, environmental factors and health outcomes in each of the 23 programmes of care. Moreover, mortality rates are available for less than half of the 23 programmes. Rather than estimate a system of equations, we proceed on a programme-by-programme basis, estimating health outcome and expenditure equations for those programmes for which mortality data is available.

In line with the theoretical framework presented in section B3, we specify the following expenditure (6.1) and health outcome (6.2) models for each of the *J* programmes of care (J=23)

$$x_i = a_1 + \sum b_{1j} \cdot n_{ij} + dy_i + e_{1i} \qquad j=1,\ldots,23 \qquad (6.1)$$

$$h_i = a_2 + b_2 n_i + f x_i + e_{2i} \qquad (6.2)$$

where x_i is the expenditure in PCT $_i$ in the selected programme
n_{ij} is the need for care in PCT $_i$ in programme $_j$
y_i is the total budget for PCT $_i$
h_i is the health gain in PCT $_i$ in the selected programme
n_i is the need for care in PCT $_i$ in the selected programme.

Ideally we should employ a programme specific indicator of the level of need for each care programme but these are not readily available. When estimating both the outcome and expenditure models we therefore proxy the own programme health care need using the 'needs' component of the Department of Health's resource allocation formula.[13] This needs element is specifically designed to adjust PCT allocations for local health care needs and accordingly, ceteris paribus, we would expect a positive relationship between expenditure x_i and need n_i for each programme of care. We would also expect a positive relationship between need n_i and adverse health outcomes h_i.[14]

The expenditure model includes both the own programme health care need (which is proxied using the 'needs' component of the Department of Health's resource allocation formula) and the need for health care in all other programmes. When estimating the expenditure model previous studies have proxied the need for health care in other (competing) programmes using the mortality rate in those other programmes. The precise definition of the programmes included in the 'other programme' mortality rate has varied a little, but here all of our preferred results from 2006/7 onwards use the 'all cause mortality rate excluding the mortality rate in the programme of interest' as the proxy for need in other programmes.[15]

B6.2 IV estimation

We do not use OLS to estimate equations (6.1) and (6.2) because both are likely to contain an endogenous regressor. Expenditure in the outcome equation (6.2) and other programme need in the expenditure equation (6.1) are both likely to be endogenous and, in the presence of an endogenous

[13] However, we do experiment with replacing and supplementing this all service measure of need with more programme specific measures where these are available (e.g., using the diabetes and epilepsy prevalence rates).
[14] Whilst need is a function of mortality/morbidity in the resource allocation formula, the relationship is not sufficiently strong enough for us to be concerned about the endogeneity of the need in any individual care programme.
[15] When estimating expenditure equations using PB data for 2005/6 for cancer and circulatory disease we persevere (for continuity with previous studies) with the use of the circulatory disease SYLLR as the proxy for other programme need in the cancer programme, and we use the cancer SYLLR as the proxy for other programme need in the circulatory disease programme (see Martin, Rice and Smith, 2008a & 2012).

regressor, OLS is both a biased and an inconsistent estimator. Instead, we use instrumental variable (IV) estimation and implement two-stage least squares (2SLS) using the -ivreg2- routine in Stata v11[21]. Unlike OLS, IV is a consistent estimator in the presence of an endogenous regressor and, although in finite samples the IV estimator will be biased, the belief is that (providing certain assumptions are met) this bias will be less than that associated with OLS.

For the health outcome equation, IV estimation can be viewed as finding variables (instruments) that are good predictors of programme expenditure but which are appropriately excluded from the equation of interest (that is, from equation 6.2). The assumption is that the instruments and exogenous variables from the equation of interest impact upon the health outcome through their impact on expenditure only, and that they do not have a direct effect on the outcome.[16] If, on the other hand, an instrument reflects unobserved factors that affect both expenditure and mortality directly, then the IV estimator becomes both biased and inconsistent. Such an instrument is said to be 'invalid' because it belongs in the equation of interest in its own right.

We have a number of potential instruments available, mostly derived from 2001 Population Census, and these are described in section B5.2. In our earlier studies we found that a small sub-set (four) of these instruments often proved sufficient to generate plausible results and we commenced our empirical work with these. If plausible results were not obtainable with some combination of these four instruments, we employed an extended instrument set. Further details of the identification of suitable instruments for each model can be found in section B7.3.

The available instruments reflect factors, such as socio-economic deprivation and the availability of informal care in the community, which might indirectly impact upon mortality rates and/or health care expenditure levels. The set of instruments associated with each estimated equation was selected on both technical and pragmatic grounds. From a pragmatic point of view, we require a parsimonious set of instruments that satisfy the necessary technical criteria. These are, firstly, that they have face validity, that is, that they are plausible determinants of the endogenous variable being instrumented, and secondly, that the instruments are both relevant and valid. The relevance of an instrument set refers to its ability to predict the endogenous variable of concern, whereas validity refers to the requirement that instruments should be uncorrelated with the error term in the equation of interest. The set of instruments was modified if, for example, the Hansen-Sargan test suggested that the set under test was not valid.

Should the instrument set be strong, relevant and valid, 2SLS will produce consistent estimates of the parameters of the reduced form models. We subject the instrument sets to tests for validity using the Sargan-Hansen test of overidentifying restrictions. The joint null hypothesis is that the instruments are valid instruments, i.e., they are uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation. A rejection of the null hypothesis casts doubt on the validity of the instruments. We test for instrument relevance using Shea's[22] partial R-squared measure; this reflects the correlation between the excluded instruments and the endogenous regressor. However, even if valid and relevant, non-zero but small correlations between the instruments and the endogenous regressors can lead to the problem of weak instruments. This can be the case even where correlations are shown to be significant at conventional levels of testing and sample sizes are large[23]. The IV estimator becomes a biased estimator if the instruments are weakly correlated with the endogenous regressors, and the extent of the bias can be specified relative to the bias of the OLS estimator.

For the case of a single regressor, Staiger and Stock[24] suggest applying the criterion that if the first-stage F-statistic, testing the null hypothesis that the instrument set does not significantly predict the endogenous regressor, is less than 10 then the instruments can be thought to be weak. Stock and Yogo[25] extend these ideas to the case where there can be multiple endogenous regressors and propose a test for the null that the instruments are weak and provide appropriate critical values. This is an extension of the Cragg and Donald [26] test for instrument relevance. For the case of a single endogenous

[16] The IV procedure involves the estimation of the second-stage expenditure equation as specified in equation 6.1 and the estimation of a first-stage expenditure equation associated with equation 6.2. The same variable might have different coefficients in these two equations because the equations will have different sets of covariates.

regressor, the Cragg-Donald statistic is simply the F-statistic of the test of the hypothesis that the instruments do not enter the first-stage regression. Stock and Yogo provide critical values of the F-statistic (and the Cragg-Donald statistic for multiple endogenous regressors) that tabulates the ratio of 2SLS bias to the bias of OLS. The weakness or otherwise of the instruments can then be assessed by the relative bias exceeding a given threshold (for example, 2SLS bias exceeding 5% of OLS bias).[17]

To ensure the robustness of our estimates to arbitrary heteroskedasticity, we estimate our models with Stata's -robust- option. The Cragg-Donald statistics are not valid in the presence of heteroskedascity. We therefore report the Kleibergen-Paap LM statistic (testing instrument relevance) and the Kleibergen-Paap F statistic (testing for weak instruments) which are valid in the presence of heteroskedascity. For further details of these tests see Baum, Schaffer and Stillman[21].

A general test of model specification is provided through the use of Ramsey's [27] reset test for OLS and an adapted version of the test for instrumental variables [28].[18] The tests are more properly thought of as tests of a linearity assumption in the mean function or a test of functional form restrictions and omitted variables[29] and can be useful as a general check of model specification.

Finally, we check that the presumed endogenous variable is in fact endogenous using the test proposed by Durbin [30]. If the null hypothesis of exogeneity cannot be rejected, then we also use the OLS estimator. And, although our instruments 'pass' the appropriate statistical tests, some commentators claim that such tests may have 'low power' to detect the presence of invalid instruments. Consequently in section B9 we examine how sensitive our results are to the relaxation of the assumption that the instruments are valid.

B6.3 Other estimation issues

In this research we build on previous studies that have used the PB data to estimate the outcome and expenditure models described in section B6.1. This previous research was undertaken over a period of years and a number of changes were made between these studies (these were sometimes forced on the researchers by, for example, data availability considerations). Here we persevere with the previous approach used to analyse the 2005/6 PB data [5], but we make some changes to the way in which the 2006/7 (and subsequent) PB data are analysed.

In the next section we start by re-visiting the results obtained by Martin, Rice and Smith [5] who used the 2005/6 PB data. In 2005/6 there were 303 PCTs but a series of mergers reduced this total to 152 in 2006/7. These mergers exacerbated greatly the difference in size between the PCTs and so from 2006/7 it makes less sense to give each PCT equal weight in any regression. This is discussed further in section B8.2 when we come to estimate our model using 2006/7 PB data.

Different PCTs face different costs when buying health care inputs. For example, some health economy input prices are up to 40% higher in London and the south east of England than elsewhere. In a previous study[3], we used the Market Forces Factor Index (MFF) that feeds into the Payment by Results tariffs for 2007/8 to adjust programme budgeting expenditure in 2006/7 for local input prices[31]. This index only reflects costs associated with the purchase of HCHS services but this was the only index available for the new (post October 2006) set of PCTs at the time of that study. Since then, a more comprehensive set of MFF indices for the 152 PCTs has been published [32]. In section B8.2 we investigate the use of alternative weighted averages of the HCHS, prescribing, and GMS/PMS MFF indices with weights reflecting the national share of expenditure across these three categories (these weights are 76.3%, 12.4%, and 11.3% respectively).[19] For 2005/6, however, we persevere with the MFF employed in the original Martin, Rice and Smith study [5], namely the HCHS MFF [33].

[17] For the case of a single endogenous regressor and three excluded instruments, Stock and Yogo (2002) critical values are as follows in term of the bias of 2SLS relative to bias of OLS as follows: relative bias 5% critical value = 13.9; relative bias 10%, critical value = 9.08; relative bias 20%, critical value = 6.46; relative bias 30%, critical value = 5.39.

[18] The OLS version of Ramsey's reset test was invoked using Stata's -ovtest- command, and the IV equivalent was invoked using -ivreset-.

[19] As all PCTs face the same prescribing costs, the prescribing MFF is 1 for all PCTs.

Estimation of the expenditure equation for any individual programme requires a proxy for the need for health care across all other programmes. Previous studies of PB expenditure in 2004/5, 2005/6 and 2006/7 have used the circulatory disease mortality rate as a proxy for the need for health care in other programmes in the cancer expenditure equation, and the cancer mortality rate as the proxy for need in other programmes in the circulatory disease expenditure equation[2, 3, 5]. As these are both programmes that attract considerable expenditure and record considerable mortality, it is not implausible that mortality and expenditure in one of the programmes will impact upon expenditure in the other. For other programmes (e.g., respiratory problems and gastro-intestinal problems) Martin, Rice and Smith[3, 5] used the all cause mortality rate as a proxy for the 'need in other programmes' variable when analysing expenditure in both 2005/6 and 2006/7. Here, however, we persevere with the previous approach when using 2005/6 PB data but, from 2006/7, in all programmes we proxy the need for health care in other (competing) programmes using the mortality rate in those other programmes (i.e., the all cause mortality rate minus the own programme mortality rate).

Finally, one data transformation that has been applied in all previous studies and is applied here too is to log transform all variables so that parameter estimates can be interpreted as elasticities. In other words, a regression coefficient of 0.5 implies that a 1% increase in the regressor is associated with a 0.5% increase in the dependent variable.

B. EMPIRICAL RESULTS

B7. Analysis of programme budgeting expenditure for 2005/6

This work builds on previous studies. Martin, Rice and Smith [2] reported outcome elasticities for two programmes (cancer and circulatory disease) using expenditure data for 2004/5 and pooled mortality data for 2002, 2003 and 2004. Martin, Rice and Smith [5] extended their preliminary analysis to include several other programmes and, in this extension, they used updated expenditure data (for 2005/6). However, the authors found it difficult to obtain sensible outcome models for some programmes of care. Here we commence our empirical work with an attempt to obtain plausible outcome models for those programmes that defeated Martin, Rice and Smith in their study. [5]

B 7.1 Construction of an alternative measure of need

Our preferred measure of the need for health care is calculated from the Department Health's programme budgeting (PB) dataset. This dataset includes programme budgeting expenditure for each care programme as well as the raw population and the 'unified weighted' population for each PCT. The unified weighted population incorporates adjustments to the raw population for both the need for health care as well unavoidable variations in local input costs. The latter are captured via an index which is known as the market forces factor (MFF). By removing the raw population and MFF adjustment from the unified weighted population we are left with the implied level of need, and this is the measure of need that was initially used in the estimation of the model.[5]

The Department of Health PB measure of need associated with expenditure for 2005/6 incorporates the AREA resource allocation formula. This has since been replaced with the CARAN formula and recent work by colleagues at York and the Nuffield Trust has investigated the possibility of constructing a person based resource allocation (PBRA) measure of need [34]. We therefore decided to investigate the possibility of applying PBRA methods to the construction of an alternative measure of need.

The construction of all of these measures of need involves two steps. The first step requires the estimation of the econometric relationship between the previous utilisation of services and the characteristics of the local areas as existed at the time of the utilisation (e.g., their demographic profile and other indicators of service need such as socio-economic measures of deprivation). The second step involves the use of this relationship to predict future health care use given predictions about future demographic characteristics and socio-economic measures of deprivation.

The major difference between the AREA and CARAN formulae and the PBRA formula is that the former largely use small area based indicators of socio-economic characteristics as indicators of the need for health care, whereas the latter largely obviates the requirement for these through the extensive use of individual based indicators of need. In particular, the PBRA formula employed here is based on an analysis of inpatient and outpatient cost data for 2007/8 for 10% of the entire population of England [34]. As regressors the PBRA utilisation model includes:

(a) 38 age/gender dummies;
(b) 150 ICD10 morbidity markers for each patient reflecting their use of inpatient services in the previous two years (that is, in 2005/6 and 2006/7 combined);
(c) 4 hospital encounter variables for each patient reflecting the intensity of their use of both outpatient and inpatient services in the previous two years (that is, in 2005/6 and 2006/7 combined);
(d) 10 small area based indicators of either local deprivation or health care supply characteristics; and
(e) 151 PCT dummies (reflecting variations in health care supply).

The coefficients from this modelling procedure are applied to patient registration data as at 1 April of the year for which the measure of need is required. Here we are studying expenditure in 2005/6 and so we applied the results of the modelling to patient registration data as at 1 April 2005. This requires the construction of a dataset containing the patient registration details of all 50 million patients registered with an English practice as at this date. To this we added the patient's age and gender as at April 2005.

We also added each patient's ICD10 morbidity markers and their encounter variables for 2003/4 and 2004/5 combined. Each patient's address (LSOA) is also added to the dataset and this is used to attach the small number of indicators reflecting the LSOA's socio-economic and health care supply characteristics.

Given this dataset, the calculation of PCT need (given supply) proceeds as follows. First, calculate the national average supply effect. This is the sum of the products of the national average values of the supply variables for the population as at 1 April 2005 and the relevant the regression coefficients.

Second, ignore supply and calculate PCT need. This involves calculating the PCT average values of the needs variables by age and gender group for the population as at 1 April 2005. Next, for each PCT, calculate need by age and gender as the sum of the products of the mean values of the needs variables and their respective regression coefficients. Then total PCT need is the sum of (need in each age/gender group multiplied by the number of patients in that age/gender group).

Finally, need given supply is calculated as total PCT need plus the number of patients multiplied by the national average supply effect. PCT need per person is simply total PCT need divided by the PCT population. Further details of how to use the results of the PBRA modelling to derive PCT weighted needs indices are presented in Dixon et al.[34]

B7.2 Re-estimation of models using a new measure of need

We re-estimated the outcome and expenditure models for the big four programmes as reported by Martin, Rice and Smith [5] using the new (PBRA based) measure of need.[20] In summary, the results for the cancer programme were acceptable but not quite as good as previously obtained, and the results for circulation problems, gastro-intestinal problems, and respiratory problems were poor (e.g., the signs on the expenditure and need variables in the outcome equation were counter-intuitive). These were unanticipated results and we were curious to know why our alternative measure of need performed less well than the more established measure.

We undertook a brief comparison of the two measures of need. Figure B7.1 provides a scatter plot of the PB and PBRA measures of need. There is a clear positive correlation between the two measures (correlation coefficient=0.6146), and the summary statistics in Table B7.1 suggest that they have similar ranges.

[20] The 'big four' programmes are the cancer, circulatory disease, respiratory problems, and gastro-intestinal problems programmes. They are 'big' programmes in terms of the number of deaths associated with each programme.

Figure B7.1: graph showing scatter plot of PB measure of need and PBRA measure of need

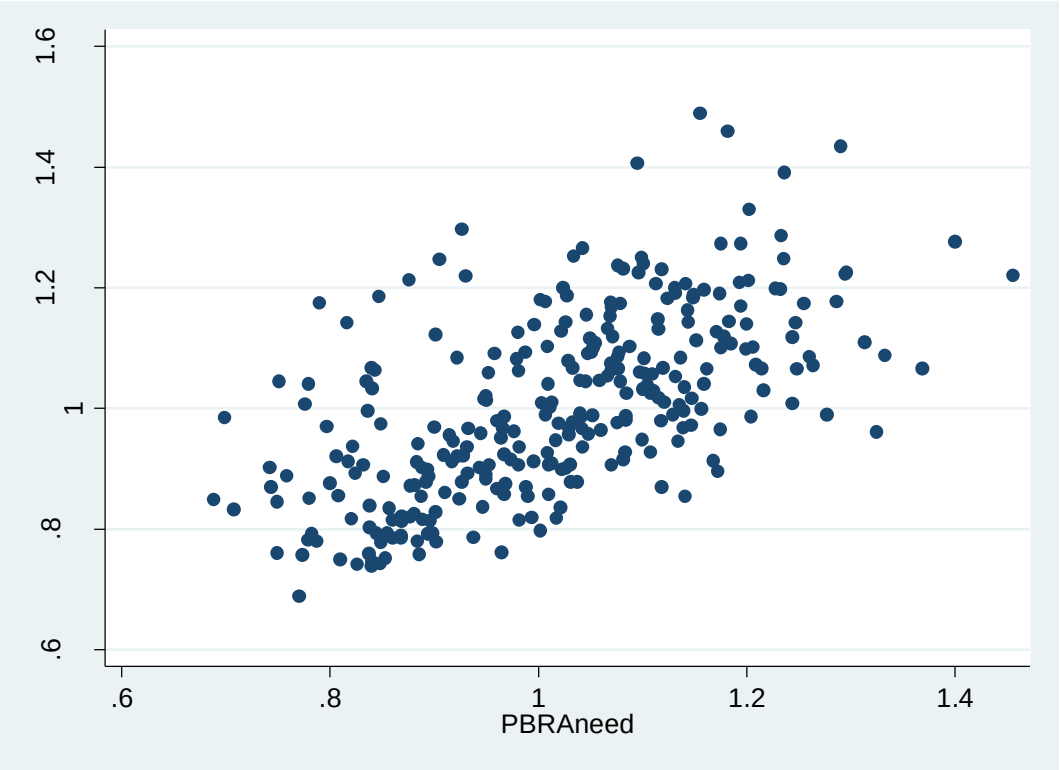


Table B7.1: table showing summary statistics for PB and PBRA based measures of need

Variable	Number of PCTs	Mean	Std. Dev.	Min	Max
PB need	295	1.0062	0.1511	0.6883	1.4889
PBRA need	303	1.0146	0.1448	0.6884	1.4554

Note that there are only 295 PCTs with a PB based measure of need because only 295 of the 303 PCTs were used to estimate our outcome and expenditure models (due to a lack of data for some PCTs).

Table B7.2 reports values for the PB and PBRA based measures of need for selected types of PCTs. These figures suggest that:

- the PBRA measure attributes more need to the least needy areas as defined by the PB measure (see Table B7.2a);
- the PBRA measure attributes more need to the coastal/retirement areas than does the PB measure (see Table B7.2b); and
- the PBRA measure attributes far less need to inner city areas than does the PB measure (see Table B7.2c).

Table B7.2: table showing values for the PB and PBRA based measures of need for selected types of PCT

(a) Examples of more affluent areas	PB need	PBRA need
Wokingham PCT	0.6883	0.7703
Blackwater Valley and Hart PCT	0.7376	0.8395
Bracknell Forest PCT	0.7410	0.8262
Royston, Buntingford and Bishop's Stortford PCT	0.7426	0.8484
Windsor, Ascot and Maidenhead PCT	0.7460	0.8404
Woking PCT	0.7499	0.8097
Chiltern and South Bucks PCT	0.7515	0.8530
Uttlesford PCT	0.7583	0.8861
North East Oxfordshire PCT	0.7588	0.8372
South Cambridgeshire PCT	0.7619	0.9647

(b) Examples of coastal/retirement areas	PB need	PBRA need
Suffolk Coastal PCT	0.9159	1.0815
Western Sussex PCT	0.9613	1.3248
North Somerset PCT	0.9651	1.1746
Poole PCT	0.9860	1.2045
South and East Dorset PCT	1.0079	1.2439
Fylde PCT	1.0304	1.2157
Southport and Formby PCT	1.0657	1.2141
North Norfolk PCT	1.0658	1.3684
Adur, Arun and Worthing PCT	1.0716	1.2641
East Devon PCT	1.0870	1.3325

(c) Examples of inner city areas	PB need	PBRA need
Brent PCT	0.9848	0.6991
Lambeth PCT	1.0454	0.7512
Islington PCT	1.1222	0.9014
Southwark PCT	1.1412	0.8163
Newham PCT	1.1746	0.7897
City and Hackney PCT	1.1849	0.8472
Bradford City PCT	1.2131	0.8757
Tower Hamlets PCT	1.2192	0.9299
Heart of Birmingham Teaching PCT	1.2466	0.9052
Central Manchester PCT	1.2965	0.9262
Central Liverpool PCT	1.4065	1.0948

Although these differences are at first perplexing, they become more understandable when it is noted that the PB and PBRA measures record the level of need across different baskets of services. The PB measure of need refers to all health care activity, that is, Hospital and Community Services (HCHS), prescribing, and GMS/PMS (primary care), but the PBRA model only incorporates hospital activity (and it excludes mental health and maternity from this).

The need for hospital based services is less related to deprivation than are other health care services. Hence the PBRA measure of need – because it only relates to hospital services – re-distributes need away from the more deprived PCTs and towards the more affluent ones. Moreover, expenditure on cancer

services is largely hospital based and hence a measure of need based on HCHS spend alone will be reasonably satisfactory for cancer (as indeed we found). However, such a measure of need will perform less well for other programmes (e.g., circulatory disease), where more of the expenditure is on prescribing and/or primary care.

To test these hypotheses we need to compare our PBRA measure of need with a PB measure of need that only relates to acute services (i.e., that excludes maternity and mental health, and all prescribing and GMS/PMS). The Department of Health's measure of need used for the 2005/6 allocations employs the AREA formula for HCHS. This formula does not permit a separation of acute and maternity need and so we cannot compare the PBRA measure of need for 2005/6 with the PB measure for 2005/6 for the same group of specialties (i.e., for acute services excluding maternity and mental health).

However, the CARAN formula, first implemented for the 2009/10 allocations, does distinguish between acute and maternity. But this formula has only been applied to the new (post October 2006, n=152) PCTs whereas our PBRA-based measure is for the old (pre October 2006, n=303) PCTs because we are modelling PB expenditure in 2005/6. However, not all of the old PCTs were involved in mergers in October 2006. Thus for about half of all PCTs, we can compare our PBRA based measure of need for 2005/6 with the CARAN-based measure of need for 2009/10 for the same set of HCHS services (i.e., for acute excluding maternity and mental health).

The correlation between PBRA need and CARAN acute need is much higher (correlation coefficient=0.8722) than that between the PBRA and PB need measures. And an inspection of the values taken by the various need indices (e.g., for acute, maternity, and mental health) for the inner city PCTs (where the PBRA and PB measures of need diverge the most) supports the hypothesis that it is the different service coverage of the PBRA and PB measures of need that explains why they are so poorly correlated (see Table B7.3).

For example, the PB index suggests that per capita need in Newham PCT is 17% *above* the national average but the PBRA index suggests that it is 21% *below* the national average. We believe that this difference is due to the fact that the PB index relates to all services whereas the PBRA index only relates to acute services. The separate figures for acute, maternity and mental health need from the CARAN formula confirm this hypothesis: CARAN acute need, like PBRA acute need, is well *below* the national average, but maternity and mental health need are well *above* it.

Table B7.3: table showing comparing PB, PBRA and CARAN need indexes for selected inner city PCTs

PCT	PB need (all services)	PBRA need (acute)	CARAN need		
			acute	maternity	mental health
City and Hackney PCT	1.1849	0.8472	0.8751	1.6783	1.5340
Tower Hamlets PCT	1.2192	0.9299	0.8451	1.4988	1.6663
Newham PCT	1.1746	0.7897	0.8683	1.8130	1.4486
Haringey PCT	1.0448	0.8347	0.8471	1.4023	1.2886
Brent PCT	0.9848	0.6991	0.8558	1.3420	1.2608
Camden PCT	1.0336	0.8402	0.7667	0.9163	1.3209
Islington PCT	1.1222	0.9014	0.8842	1.1399	1.4516
Lambeth PCT	1.0454	0.7512	0.8111	1.3916	1.3349
Southwark PCT	1.1412	0.8163	0.8445	1.3755	1.3905
Lewisham PCT	1.0402	0.7793	0.8549	1.4253	1.2236
Heart of Birmingham PCT	1.2466	0.9052	0.9078	1.5976	1.5621

B7.3 Re-estimation of poorly performing models with an extended instrument set

Martin, Rice and Smith [5] found it difficult to obtain sensible outcome models for some programmes of care. As we were unable to find an improved measure of need, we sought to improve the outcome and expenditure models reported in Martin, Rice and Smith[5] through the use of an extended set of regressors/instruments. Martin, Rice and Smith[5] had focussed on the use of four instruments but here we extend the modelling to include an additional 13 regressors/instruments (born outside EU, limiting long-term illness, no qualifications, full-time students, no car households, owner occupiers, privately rented, socially rented, lone parents, permanently sick, long-term unemployed, work in agriculture, work in professional occupation). Further details about these variables can be found in section B5.2 and precise details about how they were constructed can be found in Table BA.2 in the annex.

For each PBC, our modelling strategy with these additional regressors/instruments was the same:

(a) first, estimate an IV model using our preferred set of regressors (with need, budget, and other programme need for the own programme spend model, and with need and spend for the outcome model) and preferred set of instruments (proportion of households that are lone pensioner households, per cent of the population providing unpaid care, the IMD 2000, and the per cent of the population in the white ethnic group). Then adjust this set of instruments if necessary (e.g., remove from the instrument set or add an instrument to the regressor set if the Hansen-Sargan test indicates that this is appropriate). Estimate an OLS version of the IV model if the theoretically endogenous regressor is exogenous according to the relevant statistical test.

(b) second, if (a) fails to generate a reasonable model, add the same additional variables to both the regressor and instrument sets. Then eliminate insignificant regressors (least significant first, but always retaining e.g., the budget and other need variables in the expenditure model, and own programme spend in the outcome model). Then eliminate insignificant instruments until a reasonable model is obtained. Again, estimate an OLS version of the IV model if the theoretically endogenous regressor is exogenous according to the relevant statistical test.

B7.4 IV estimates of outcome and expenditure models

The above approach generates preferred outcome and expenditure models for each of the programmes with a mortality based outcome indicator. Outcome models are shown in Table B7.4 with expenditure models in Table B7.5. The corresponding first-stage regression results can be found in Tables BA.3 and BA.4 respectively in the annex.

The first four results in Table B7.4 show the outcome model for the big four programmes (i.e., for cancer, circulatory disease, respiratory problems and gastro-intestinal problems). In all four programmes the need variable has a positive and significant effect on mortality, and expenditure has the anticipated negative effect. The diagnostic statistics reveal that, in all four cases, own programme expenditure is endogenous and that the instruments are valid. They also suggest that the instruments are relevant and there is no evidence that the instruments are weak. The Pesaran-Taylor test suggests that there is no evidence of model mis-specification.

The results for the other programmes are similar to but more diverse than those for the big four programmes. This is to be anticipated because mortality is a much rarer outcome in these programmes than it is in, say, the cancer programme. Own programme expenditure is not endogenous in the next two programmes (infectious diseases and neurological problems) and we revert to the use of the OLS estimator. Expenditure has the anticipated negative effect on mortality in the infectious disease programme but this is not statistically significant. The all service measure of need is not relevant for this PBC; instead, we find that a measure of need associated with HIV is positively associated with mortality, as is a measure of deprivation (households with no car). Mortality from epilepsy is negatively associated with expenditure in the neurological programme. The need for health care variables has a positive and significant effect on mortality.

Expenditure and need have the anticipated effects on mortality in the trauma and injuries programme. In addition, the provision of unpaid care appears to be associated with an increase in mortality from fractures. This might be because the availability of care allows the elderly to continue to live in their own home and that they are more likely to fall and die from a fall at home than they are in alternative accommodation (such as in a residential home or sheltered housing).

Expenditure has the anticipated negative effect on mortality in the neonates programme where the generic all service measure of need has been replaced with two more programme specific indicators of need (the proportion of births that are low birth weight births and the proportion of households that are lone parent households).

The final two results both employ the OLS estimator. Expenditure in the genitor-urinary programme has a small negative effect on mortality (from renal problems). The prevalence of one parent households and non-white residents both seem to be positively associated with mortality.

Finally, expenditure has the anticipated negative effect on mortality in the endocrine problems programme where the generic all service measure of need has again been replaced with a more programme specific indicator of need (the diabetes prevalence rate). Mortality in this programme is also positively associated with the Index of Multiple Deprivation 2000.

The first four results in Table B7.5 show the expenditure model for the big four programmes (i.e., for cancer, circulatory disease, respiratory problems and gastro-intestinal problems). In all four programmes both the need and budget variables have a positive and significant effect on own programme expenditure. In addition, the proxy for need in other programmes is negative and significant in all four cases. In the circulatory expenditure programme the provision of unpaid care is associated with more expenditure (patients may buy care in more affluent areas), as is the proportion of residents in the white ethnic group (there might be some unmet need associated with circulatory problems in the non-white ethnic groups).

The PCT budget variable is positive in all of the remaining seven programmes and this variable is significant in six of the seven. The proxy for other programme need (SYLLR all deaths) has the anticipated negative sign in five of the seven programmes and, where it is positive, it is never statistically significant.

The all service proxy for own programme need is positive and significant in three programmes. In the other four programmes, however, it has been replaced various other socio-economic indicators of need: in the trauma programme, for example, with the provision of unpaid care is associated with a reduction in NHS expenditure and, in the neonates programme, the proportion of residents in the white ethnic group is negatively associated with expenditure.

The diagnostic statistics reveal that, for all seven IV models, expenditure is endogenous and the instruments are valid. They also suggest that the instruments are relevant and there is no evidence that the instruments are weak. The Pesaran-Taylor test suggests that there is no evidence of model mis-specification.

B7.5 IV estimates of outcome and expenditure models: the first-stage equations

For the health outcome equation, IV estimation involves finding variables (instruments) that are good predictors of programme expenditure but which are appropriately excluded from the equation of interest (that is, from the outcome equation). The assumption is that the instruments impact upon the health outcome through their impact on expenditure only, and that they do not have a direct effect on the outcome. If, on the other hand, an instrument reflects unobserved factors that affect both expenditure and mortality directly, then the IV estimator becomes both biased and inconsistent. Such an instrument is said to be 'invalid' because it belongs in the equation of interest in its own right.

In our outcome model we typically employ two instruments (call these z1 and z2) for expenditure. IV estimation assumes that these instruments do not belong in the outcome equation. In other words, IV estimation assumes that the coefficients $\gamma 1$ and $\gamma 2$ in the outcome model

$$y = \propto + \beta 1x + \beta 2n + \gamma 1z1 + \gamma 2z2 + \epsilon \qquad (7.1)$$

are identically zero (where y is mortality, x is expenditure, and n is a measure of the own programme need for health care and all variables relate to a particular programme of care). Such exclusion restrictions can be debatable and researchers who employ IV techniques often devote considerable effort towards convincing the reader that their assumed exclusion restrictions are a good approximation [35, 36]. These efforts usually take two forms: first, researchers often offer a strong theoretical economic argument why their instruments do not belong in the equation of interest; and, second, statistical tests for the validity of the exclusion restrictions (Sargan 2SLS, Hansen J-test GMM) are routinely reported as part of the results for any study that employs IV techniques.

It is difficult for us to identify clear theoretical reasons why our instruments (such as the proportion of lone pensioner households, the provision of unpaid care, and an index of multiple deprivation) do not belong in the equation of interest (that is, that they will not *directly* affect mortality). Of necessity, therefore, we must be guided by the available statistical tests for the validity of the exclusion restrictions. However, although our outcome models 'pass' the relevant statistical test, some commentators have argued that the Sargan/Hansen test may have weak power and may fail to reject the null hypothesis of instrument validity even when an exclusion restriction is not valid. As we shall see in section B9, this is likely to be the case when the induced biases in the estimates of $\beta 1$ (the coefficient on the endogenous variable) are the same across all instruments. The Hansen-Sargan J test statistic will be small when the null hypothesis of valid instruments is correct; but it will also be small if the biases induced in $\widehat{\beta 1}$ by invalid instruments all coincide (i.e., the instruments all identify the *same* wrong parameter)[37]. In other words, for the Hansen-Sargan J test to have low power the use of any subset of instruments should generate the *same* asymptotic bias in $\widehat{\beta 1}$.

Our approach, implemented below, is to identify theoretical reasons why our instruments might belong in the first-stage expenditure equation but not in the second-stage outcome equation. Even if our arguments are thought unconvincing, a critic would also have to argue that any subset of our selected instruments will each induce the same bias in the coefficient on the endogenous variable. This is because it is only in these circumstances that the Hansen-Sargan test will be unable to reject the null hypothesis of instrument validity even when an exclusion restriction is not valid.

The first stage regressions associated with the IV outcome results in Table B7.4 can be found in Table BA.3 in the annex. A brief summary of the first-stage regressions is provided below.

Cancer programme of care

The instrument set for the cancer programme of care (see column 1 in Table BA.3) includes the proportion of households that are lone pensioner households and the proportion of the population providing unpaid care. These instruments have intuitive appeal. The first stage regression of cancer expenditure on the instruments and the need for health care (as an exogenous regressor in the 2SLS model) reveals a positive and significant coefficient on lone pensioners and a negative but non-significant coefficient on the proportion of unpaid carers. The proportion of lone pensioners is likely to reflect an additional adjustment for health care need specific to an elderly and needy population. The omission of this variable from the second-stage regression is plausible as the dependent variable relates to mortality under 75 years of age and some of the lone pensioners will be aged over 75 years, and members of this group are, by definition, relatively healthy individuals. Unpaid care might act as a substitute for the provision of health care services and, in these circumstances, a negative relationship with expenditure is to be expected. There is no obvious relationship between the provision of unpaid care and mortality.

Circulatory disease programme of care

The two instruments used for cancer were also employed to predict expenditure in the circulatory disease programme and they were augmented with the addition of the population weighted index of multiple

deprivation (IMD 2000). The relevance of the latter variable is theoretically plausible as circulatory disease is more related to disadvantage than is cancer. In addition, we also employed the proportion of residents in the white ethnic group as an additional instrument for expenditure but its coefficient is very small and it is not statistically significant.

Increased expenditure on circulatory disease in the first stage regression is associated with a greater proportion of pensioners living alone and a greater proportion of unpaid carers. The latter may reflect an increased awareness and compliance with medical intervention, particularly preventative measures, brought about by carers but this will not affect our outcome model if the impact of this additional support is largely on the mortality of those aged over 75 years. Expenditure on circulatory problems is also negatively associated with the IMD 2000. As the IMD incorporates an access to medical services domain, this negative association might reflect some unmet need which largely affects mortality in those aged over 75 years.

Respiratory problems programme of care
The IMD 2000 is negatively associated with expenditure on respiratory problems. As the IMD incorporates an access to medical services domain, this negative association might reflect some unmet need which largely affects mortality in those aged over 75 years. The proportion of the population aged 16-74 that is permanently sick has a positive association with expenditure but might not affect mortality in the under 75s if expenditure is largely directed towards managing chronic disease.

Gastro-intestinal problems programme of care
Increased expenditure on gastro-intestinal problems in the first stage regression is positively associated with the proportion of residents providing unpaid care. This may reflect an increased awareness and compliance with medical intervention, particularly preventative measures, brought about by carers but this will not affect our outcome model if the impact of this additional support is largely on the mortality of those aged over 75 years.

Trauma, burns and injuries programme of care
Increased expenditure on trauma, burns and injuries in the first stage regression is positively associated with the proportion of pensioners living alone. This may reflect longer stays in hospital and an increased need for community care. However, the proportion of pensioners living alone will have little effect on our mortality measure if most of this expenditure is associated with patients aged over 75 years of age.

Neonate programme of care
The percentage of those aged 16-74 that are long-term unemployed and the proportion of households that are in social rented housing are both positively associated with expenditure on neonate care. These are both indicators of socio- economic deprivation and might be associated with the presence of larger families (i.e., more children per family). This would affect expenditure per head of population but not necessarily mortality per 1,000 live births. The negative coefficient on the proportion of those aged 16-74 with no qualifications might reflect the 'emigration' of young adults from those areas that are particularly deprived. This would reduce expenditure per head of population but would have no impact on the mortality measure.

The first stage regressions associated with the IV expenditure results in Table B7.5 can be found in Table BA.4 in the annex.

Cancer programme of care
The first-stage equation for the cancer expenditure model includes two instruments – lone pensioners, and unpaid carers -- that are excluded as regressors from the second stage of estimation. In this model the first stage regression of other programme need (as proxied here by the circulatory disease mortality rate) on the instrument set generates a negative coefficient on both instruments excluded from the second-stage regression. A greater proportion of unpaid carers might reflect an increased level of care (and perhaps increased compliance with care programmes and drug regimes) resulting in a decrease in other programme deaths. The availability of unpaid care in the community might not have a direct effect on cancer expenditure if such care supplements rather than substitutes for NHS funded care. Conditional

on need and the total PCT budget, the negative coefficient on the proportion of lone pensioners may be indicative of the presence of increased networks of social support. If this additional support reduces other programme mortality but does not substitute for NHS care, then the lone pensioner variable will not belong in the expenditure equation.

Circulatory disease programme of care
In the circulatory disease expenditure model, the first stage regression of other programme need (as proxied here by cancer mortality rate) on the instrument set results in a negative coefficient on one instrument (lone pensioners) and a positive coefficient on the other (the IMD 2000). As noted above, the negative coefficient on the proportion of lone pensioners may be indicative of areas with increased networks of social support. If this additional support does not substitute for NHS care then the lone pensioner variable will not belong in the expenditure equation. It is plausible that the IMD 2000 should have a positive effect on other programme need but not belong in the expenditure equation if, for example, there is some unmet need in another (but not the circulatory disease) care programme.

Respiratory problems programme of care
In the respiratory disease expenditure model, the first stage regression of other programme need (as proxied here by the all cause SYLL rate) on the instrument set results in a negative coefficient on one instrument (unpaid care) and a positive coefficient on another (i.e., on the IMD2000). A greater proportion of unpaid carers might reflect an increased level of care (and perhaps increased compliance with care programmes and drug regimes) resulting in a decrease in other programme deaths. The availability of unpaid care might not have a direct effect on own programme expenditure if such care does not substitute for NHS funded care. It is plausible that the IMD 2000 should have a positive effect on other programme need but not belong in the expenditure equation if, for example, there is some unmet need in another (but not the respiratory disease) care programme.

Gastro-intestinal problems programme of care
In the gastro-intestinal problems expenditure model, the first stage regression of other programme need (as proxied here by the all cause SYLL rate) on the instrument set (including need and total budget) results in a negative coefficient on one instrument (lone pensioners) and a positive coefficient on the other (IMD2000). As noted above, the negative coefficient on the proportion of lone pensioners may be indicative of areas with increased networks of social support. If this additional support does not substitute for NHS care then the lone pensioner variable will not belong in the expenditure equation. It is plausible that the IMD 2000 should have a positive effect on other programme need but not belong in the expenditure equation if, for example, there is some unmet need in another (but not the gastro-intestinal) care programme.

Neurological problems programme of care
The first-stage equation for the neurological expenditure model includes three instruments – lone pensioners, unpaid carers and IMD2000 -- that are excluded as regressors from the second stage of estimation. Explanations for the signs on these variables have been outlined above when discussing the other first stage regressions.

Trauma and injuries programme of care
The first-stage equation for the trauma expenditure model includes two instruments – lone pensioners and the IMD2000 -- that are excluded as regressors from the second stage of estimation. Explanations for the signs on these variables have been outlined above when discussing the other first stage regressions.

GMS/PMS programme of care
The first-stage equation for the GMS/PMS expenditure model includes three instruments – households with no car, lone parents, and permanently sick -- that are excluded as regressors from the second stage of estimation. All three are plausibly positively associated with other programme need (as proxied here by the all cause SYLL rate) but do not occur as regressors in the second stage GMS/PMS expenditure model. The latter includes at least one measure of deprivation – the proportion of people aged 16-74

without any qualification – and the Hansen-Sargan test suggests that three excluded instruments offer no additional explanatory power for observed variations in GMS/PMS expenditure.

We appreciate that not everyone will be convinced by our arguments about the validity of our instruments and so in section B9 we undertake a sensitivity analysis that examines the impact of weakening the instrument exclusion restriction.

Table B7.4: table showing preferred outcome models using 2005/6 expenditure data and mortality for 2002/2003/2004

VARIABLES	(1) PBC 2 cancer 2005/6 outcome model instrument spend unweighted second stage	(2) PBC 10 circulation 2005/6 outcome model instrument spend unweighted second stage	(3) PBC 11 respiratory 2005/6 outcome model instrument spend unweighted second stage	(4) PBC 13 gastro-intestinal 2005/6 outcome model instrument spend unweighted second stage	(5) PBC 1 infectious disease 2005/6 outcome model spend exogenous unweighted OLS	(6) PBC 7 neurological 2005/6 outcome model spend exogenous unweighted OLS	(7) PBC 16 trauma 2005/6 outcome model instrument spend unweighted second stage	(8) PBC 19 neonates 2005/6 outcome model instrument spend unweighted second stage	(9) PBC 17 genito-urinary 2005/6 outcome model spend exogenous unweighted OLS	(10) PBC 4 endocrine 2005/6 outcome model spend exogenous unweighted OLS
own programme spend p/head	-0.394***	-1.370***	-1.574***	-2.018***	-0.152	-0.182	-1.332***	-0.237*	-0.034	-0.244*
	[0.100]	[0.156]	[0.483]	[0.364]	[0.117]	[0.143]	[0.469]	[0.127]	[0.220]	[0.129]
need per head	0.905***	2.628***	4.076***	4.254***		1.157***	1.588***			
	[0.083]	[0.163]	[0.562]	[0.412]		[0.252]	[0.445]			
lone pensioner households			-0.930***							
			[0.158]							
born outside EU					0.111*					
					[0.063]					
no car households					0.701***					
					[0.114]					
HIV need per head					0.212**					
					[0.082]					
unpaid carers							1.164***			
							[0.392]			
low birth weight births								0.919***		
								[0.223]		
lone parents households								0.549***	1.035***	
								[0.121]	[0.211]	
white ethic group									-1.246***	
									[0.329]	
population weighted IMD 2000										0.421***
										[0.076]
diabetes prevalence rate 2004/5										14.236***
										[5.195]
diabetes prevalence rate squared										2.026***
										[0.759]
constant	4.101***	1.849***	-2.892**	-2.052**	2.654***	0.917**	0.689	1.621***	2.188***	24.258***
	[0.248]	[0.324]	[1.250]	[0.916]	[0.443]	[0.459]	[1.462]	[0.455]	[0.681]	[8.859]
Observations	295	295	295	295	295	294	295	294	267	294
R-squared					0.328	0.068			0.169	0.203
Endogeneity test statistic	29.216	65.024	12.630	39.106			3.542	4.071		
Endogeneity p-value	6.47e-08	0	0.000380	4.01e-10			0.0598	0.0436		
Hansen-Sargan test statistic	0.786	7.209	1.877	2.468			1.200	5.976		
Hansen-Sargan p-value	0.375	0.0655	0.171	0.291			0.273	0.0504		

Shea's partial R-squared	0.133	0.311	0.0376	0.173			0.112	0.0735		
Kleibergen-Paap LM test statisti	26.59	42.31	20.56	34.83			26.97	19.04		
Kleibergen-Paap p-value	1.68e-06	1.44e-08	3.44e-05	1.32e-07			1.39e-06	0.000268		
Kleibergen-Paap F statistic	16.94	29.51	10.29	23.32			17.76	11.49		
Pesaran-Taylor reset statistic	0.0347	0.162	0.0929	2.196			0.756	1.388		
Pesaran-Taylor p-value	0.852	0.688	0.761	0.138			0.385	0.239		
Ramsey reset F statistic					2.089	0.665			1.075	1.118
Probability > F					0.102	0.574			0.360	0.342

Notes: (i) Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1;
(ii) for the endogeneity test the null is that the specified endogenous regressors can actually be treated as exogenous;
(iii) the instrument validity test is based on the Hansen-Sargan test. The joint null hypothesis is that the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.
(iv) Shea's partial R-squared is an indicator of the degree of instrument relevance (i.e., of the correlation between the instruments and the endogenous regressor). It is the value of R-squared from a regression of the endogenous variable on the excluded instruments.
(v) A statistical test of instrument relevance is provided by the Kleibergen-Paap LM test. The null hypothesis is that the instruments are not relevant.
(vi) Weak identification arises when the excluded instruments are correlated with the endogenous regressor, but only weakly. Estimators can perform poorly when instruments are weak. The Kleibergen-Paap Wald F statistic provides a formal test of weak identification. The null hypothesis is that the instruments are weak.

Table B7.5: table showing preferred expenditure models using 2005/6 expenditure data and mortality for 2002/2003/2004

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	PBC 2	PBC 10	PBC 11	PBC 13	PBC 1	PBC 7	PBC 16	PBC 19	PBC 17	PBC 4	PBC 23
	cancer	circulation	respiratory	gastro-intestinal	infectious disease	neurological	trauma	neonates	genito-urinary	endocrine	GMS/PMS
	2005/6	2005/6	2005/6	2005/6	2005/6	2005/6	2005/6	2005/6	2005/6	2005/6	2005/6
	spend model instrument spend	spend model instrument spend	spend model instrument spend	spend model instrument spend	spend model o/need exogenous	spend model instrument spend	spend model instrument o/need	spend model o/need exogenous	spend model o/need exogenous	spend model o/need exogenous	spend model instrument spend
	unweighted	unweighted	unweighted	unweighted	unweighted	unweighted	unweighted	unweighted	unweighted	unweighted	unweighted
VARIABLES	second stage	second stage	second stage	second stage	OLS	second stage	second stage	OLS	OLS	OLS	second stage
SYLLR cancer		-0.954***									
		[0.249]									
PCT budget per head	0.968***	0.682***	0.849***	0.772***	0.742***	1.111***	0.627***	0.388	1.041***	0.425**	0.926***
	[0.191]	[0.161]	[0.223]	[0.166]	[0.180]	[0.244]	[0.173]	[0.391]	[0.141]	[0.175]	[0.199]
need per head	0.703***	0.885***	2.226***	1.115***		0.773***	1.720***			0.570***	
	[0.248]	[0.261]	[0.436]	[0.230]		[0.298]	[0.401]			[0.207]	
white ethnic group		0.198***						-0.739***			
		[0.066]						[0.181]			
provision of unpaid care		0.364***					-0.339*				
		[0.136]					[0.190]				
SYLLR circulatory disease	-0.577***										
	[0.107]										
lone pensioners			-0.612***								-0.257**
			[0.165]								[0.101]
SYLLR all deaths			-1.367***	-0.639***	-0.437***	-0.899***	-1.157***	0.121	0.035	-0.158	-1.003***
			[0.328]	[0.149]	[0.157]	[0.182]	[0.274]	[0.307]	[0.099]	[0.116]	[0.276]
born outside EU					0.069**						
					[0.029]						
full-time students					-0.165***				0.127***		
					[0.053]				[0.031]		
no car households					0.444***						
					[0.099]						
HIV need per head					0.142***						
					[0.034]						
London boroughs dummy					0.942***						
					[0.106]						
LA/HA rented housing								0.377***			
								[0.126]			
no qualifications											0.521***
											[0.140]
private rented housing											0.102**
											[0.041]
work in agriculture											-0.058***
											[0.022]

Constant	-0.020	3.440***	4.368**	1.241	-0.969	2.066*	3.631**	-4.760**	-2.854***	-2.435***	4.320**
	[0.517]	[1.111]	[1.757]	[0.930]	[1.037]	[1.131]	[1.414]	[1.920]	[0.604]	[0.726]	[1.703]
Observations	295	295	295	295	295	295	295	295	295	295	295
Endogeneity test statistic	6.465	12.921	19.325	7.218		13.865	12.690				5.273
Endogeneity p-value	0.0110	0.000325	1.10e-05	0.00722		0.000196	0.000368				0.0217
Hansen-Sargan test statistic	0.416	1.925	0.00232	2.441		0.826	3.577				3.213
Hansen-Sargan p-value	0.519	0.165	0.962	0.118		0.662	0.0586				0.201
Shea's partial R-squared	0.450	0.141	0.168	0.416		0.450	0.239				0.290
Kleibergen-Paap LM test statistic	63.99	31.27	33.11	57.16		64.15	39.80				47.98
Kleibergen-Paap p-value	0	1.62e-07	6.47e-08	0		0	2.28e-09				2.15e-10
Kleibergen-Paap F statistic	109.7	21.74	19.08	98.29		70.14	40.01				43.10
Pesaran-Taylor reset statistic	2.679	0.231	0.848	0.987		0.0184	0.912				0.668
Pesaran-Taylor p-value	0.102	0.631	0.357	0.320		0.892	0.340				0.414
R-squared					0.709			0.177	0.399	0.267	
Ramsey reset F statistic					1.572			0.250	1.358	0.765	
Probability > F					0.196			0.861	0.256	0.514	

See notes to Table B7.4.

B7.6 Calculation of the cost of a life and life year

The preferred models identified in Tables B7.4 and B7.5 indicate the responsiveness of mortality to changes in expenditure, and of own programme expenditure to changes in budget, using expenditure data for 2005/6. Together with information about programme expenditure and mortality, the coefficients on the own programme expenditure and PCT budget variables listed in Tables B7.4 and B7.5 can be used to calculate the cost of an additional life year for the ten programmes for which outcome and expenditure models are available.[21] For a relatively small budget change:

the cost of an additional life in a particular programme
= the change in expenditure in that programme / the change in mortality in that programme
= (annual spend * expenditure elasticity) / (annual mortality * outcome elasticity * expenditure elasticity)

and

the cost of an additional life year in a particular programme
= the change in expenditure in that programme / the change in life years lost in that programme
= (annual spend * expenditure elasticity) / (annual life years lost * outcome elasticity * expenditure elasticity).

Table B7.6 presents the necessary information to calculate the cost of an additional life (or life year) for each of these ten programmes. There is an assumed small (1%) increase in the national budget and it is also assumed that this increase is applied to each PCT's budget. The total additional spend in each programme associated with this injection (column E) is determined by the initial level of expenditure in the programme (column B) and the programme's expenditure elasticity (column D). And this additional spend, in conjunction with the outcome elasticity (column H) and the number of deaths in the programme (column G), determine the number of lives saved that is associated with the additional expenditure. If we divide the change in programme expenditure (column E) by the change in the number of lives lost (column I) we obtain the cost per life gained (column K).

Alternatively, we can apply the outcome elasticity (column H) to the annual number of life years lost in the programme (column G) to determine the number of life years saved that is associated with the additional expenditure. If we divide the change in programme expenditure (column E) by the change in the number of life years lost (column N) we obtain the cost per life year gained (column O). Note that none of these figures are QALY adjusted and that all costs are at current (2005/6) prices.

The cost per life year associated with the cancer programme is £13,741 and this is almost identical to that calculated using expenditure data for 2004/5 but with the same mortality data as that employed here [2]. Similarly, the cost per life year associated with the circulatory disease programme is £8,328 and this is also almost identical to that calculated using expenditure data for 2004/5 but with the same mortality data as that employed here [2]. The cost per life year for the respiratory programme (£20,601) and for the gastro-intestinal programme (£18,303) are a little larger than these figures but are still of the same order of magnitude. Taken together, the cost per life year for these 'big four' programmes is £12,855.[22] [23]

Table B7.6 also contains cost per life year estimates for the six other programmes for which a mortality-based outcome indicator is available. These cost estimates are much larger than those for the big four

[21] The programme specific cost per life and life year estimates presented here will underestimate the true programme specific costs because not all PCT expenditure can be allocated to a specific programme (for example, all GMS expenditure is allocated to PBC23 rather than being split between cancer, circulatory disease, respiratory problems, etc). However, this more generic expenditure *is* incorporated into the calculation of the cost of a life year when this calculation is undertaken across all programmes.

[22] These are the 'big four' programmes in terms of the number of lives (or life years) lost.

[23] The cost of a life year for a group of PBCs is calculated by dividing (a) the sum of the change in spend on the component PBCs by (b) the sum of the change in the number of lives/life years lost for the component PBCs.

programmes. This is to be expected as mortality is a less relevant outcome indicator for these PBCs than for the big four programmes. The cost per life year across all ten programmes for which a mortality-based outcome indicator is available is £21,256.

Although we have an estimate of the cost per life year for ten programmes, it is unclear how we should adjust this estimate for the expenditure associated with the other 13 programmes. We attempted to estimate an outcome and expenditure model for expenditure and mortality in all 13 of these programmes combined.[24] However, this was not successful with, for example, counter-intuitive signs on some variables.[25]

Instead, we decided to make some assumption about the cost per life year associated with the other 13 programmes. We examined two possibilities. First, we assumed that the other 13 programmes generate no mortality gain at all. This is clearly unrealistic but it does provide an upper bound for the cost per life year across all programmes of care. Table B7.7 is similar to Table B7.6 but it incorporates this zero gain assumption for the 13 other programmes.[26] It shows that the cost per life year across all 23 programmes – assuming a zero mortality gain in the 13 programmes without a mortality based indicator – is £56,799.

Second, the zero mortality gain assumption is an extreme one but possibly relevant for the residual programme (PBC 23) -- where about two-thirds of the expenditure is attributable to primary care -- if we assume that any mortality gain associated with primary care expenditure is reflected in mortality rates associated with other, more disease specific, programmes (e.g., cancer, circulatory disease, etc). But if we assume a zero mortality gain in PBC 23, what assumption should we make about the mortality gain associated with the remaining 12 programmes?

One possibility is to assume that the cost per life (year) in the remaining 12 programmes is on average the same as that associated with the ten programmes for which a mortality-based outcome indicator is available. At first this may sound strange as we have already noted that mortality is not regularly associated with these programmes whereas it is a normal outcome for the ten programmes for which a mortality-based outcome indicator is available (and this is of course why mortality data at PCT level is available for these ten PBCs). However, if we broaden our interpretation of health gain to include non-mortality effects (such as those on the quality of life), then this assumption – that the cost per life (year) in the remaining 12 programmes is on average the same as that associated with the ten programmes for which a mortality-based outcome indicator is available – becomes far more plausible.

Thus Table B7.8 is similar to Table B7.7 but incorporates: (a) a zero gain assumption for the residual (including primary care) programme (PBC23); and (b) an average gain assumption for the remaining 12 programmes for which no mortality based outcome indicator is available. Table B7.8 shows that the cost per life year across all 23 programmes (see row 15) is £24,200. This is, of course, slightly greater than the cost of a life year for the ten programmes for which a mortality-based outcome indicator is available (£21,256) because a small proportion of expenditure (that on primary care) is assumed to have no health benefit beyond that captured by the more disease specific programmes (e.g., in cancer, circulatory disease, etc).

The costs quoted in Tables B7.6, B7.7, and B7.8 make no QALY adjustment but such an adjustment would add between 50% and 66% to the costs quoted [5].

---

[24] We are grateful to Steve Morris for this suggestion.

[25] Instead of estimating programme specific models we also tried estimating an outcome model using the all cause mortality rate and expenditure across all programmes combined but this was not successful (again, counter-intuitive signs were obtained on some variables). We also investigated the possibility of using an overall measure of health derived from the Health Survey for England. Apart from sample size issues at PCT level (4,645 adults in England were interviewed for the 2009 survey), such surveys by definition only provide information about the health status of the living population and reveal nothing about the level of mortality.

[26] The cost of a life year for those 13 programmes where there is no health gain is, of course, undefined.

Table B7.6: table showing cost of life and life year estimates for 2005/6 for the ten programmes for which we have outcome and expenditure elasticities

	A	B	C	D	E (=0.01* B*D)	F	G	H	I (=0.01* D*G*H)	J	K (=E/I)	L	M	N (=0.01*D* H*M/3)	O (=E/N)
	PBC description	Spend (£m) 2005/6		Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2002/04	Outcome elasticity	Change in annual mortality		Cost per life gained (£)		Total life years lost, <75years, 2002/04	Change in annual life years lost	Cost per life year gained (£)
1	Cancer	£4,094		0.968	£39.63		62,259	0.394	237.45		£166,897		2,268,541	2,884	£13,741
2	Circulatory problems	£6,112		0.682	£41.68		45,504	1.370	425.16		£98,042		1,607,171	5,005	£8,328
3	Respiratory problems	£3,421		0.849	£29.04		11,601	1.574	155.03		£187,350		316,506	1,410	£20,601
4	Gastro-intestinal problems	£3,998		0.772	£30.86		5,926	2.018	92.32		£334,318		324,735	1,686	£18,303
5	Big four programmes	£17,625			£141.22		125,290		909.96		£155,196		4,516,953	10,986	£12,855
6	Infectious diseases	£1,161		0.742	£8.61		2,050	0.152	2.31		£3,725,931		106,552	40	£215,054
7	Endocrine problems	£1,832		0.425	£7.79		1,690	0.244	1.75		£4,442,720		60,615	21	£371,601
8	Neurological problems	£2,019		1.111	£22.43		729	0.182	1.47		£15,217,293		66,137	45	£503,201
9	Genito-urinary problems	£3,313		1.041	£34.49		294	0.034	0.10		£331,432,573		10,030	1	£29,144,918
10	Trauma & injuries	£3,758		0.627	£23.56		1,037	1.332	8.66		£2,720,657		30,000	84	£282,132
11	Neonate conditions	£660		0.388	£2.56		2,123	0.237	1.95		£1,311,733		477,675	146	£17,490
12	All ten programmes	£30,368		0.792	£240.67		133,213		926.22		£259,838		5,267,962	11,322	£21,256
	Note:														
	All 23 programmes	£64,310													
	% change in budget	1.00													
	proportionate change	0.01													
	Change in budget	£643.10													

Table B7.7: table showing cost of life and life year estimates for 2005/6 for all programmes (assumes that 13 PBCs offer no health gain)

	A	B	C	D	E (=0.01*B*D)	F	G	H	I (=0.01*D*G*H)	J	K (=E/I)	L	M	N (=0.01*D*H*M/3)	O (=E/N)
	PBC description	Spend (£m) 2005/6		Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2002/04	Outcome elasticity	Change in annual mortality		Cost per life gained (£)		Total life years lost, <75years, 2002/04	Change in annual life years lost	Cost per life year gained (£)
1	Cancer	£4,094		0.968	£39.63		62,259	0.394	237.45		£166,897		2,268,541	2,884	£13,741
2	Circulatory problems	£6,112		0.682	£41.68		45,504	1.370	425.16		£98,042		1,607,171	5,005	£8,328
3	Respiratory problems	£3,421		0.849	£29.04		11,601	1.574	155.03		£187,350		316,506	1,410	£20,601
4	Gastro-intestinal problems	£3,998		0.772	£30.86		5,926	2.018	92.32		£334,318		324,735	1,686	£18,303
5	Big four programmes	£17,625			£141.22		125,290		909.96		£155,196		4,516,953	10,986	£12,855
6	Infectious diseases	£1,161		0.742	£8.61		2,050	0.152	2.31		£3,725,931		106,552	40	£215,054
7	Endocrine problems	£1,832		0.425	£7.79		1,690	0.244	1.75		£4,442,720		60,615	21	£371,601
8	Neurological problems	£2,019		1.111	£22.43		729	0.182	1.47		£15,217,293		66,137	45	£503,201
9	Genito-urinary problems	£3,313		1.041	£34.49		294	0.034	0.10		£331,432,573		10,030	1	£29,144,918
10	Trauma & injuries	£3,758		0.627	£23.56		1,037	1.332	8.66		£2,720,657		30,000	84	£282,132
11	Neonate conditions	£660		0.388	£2.56		2,123	0.237	1.95		£1,311,733		477,675	146	£17,490
12	All ten programmes	£30,368		0.792	£240.67		133,213		926.22		£259,838		5,267,962	11,322	£21,256
	Other 13 programmes?														
13	Assume no health gain	£33,942			£402.43				0.00					0	
14	All 23 programmes	£64,310			£643.10				926.22		£694,330			11,322	£56,799
	Note:														
	All 23 programmes	£64,310													
	% change in budget	1.00													
	proportionate change	0.01													
	Change in budget	£643.10													

Table B7.8: table showing cost of life and life year estimates for 2005/6 for all programmes (assumes GMS/PMS provides no gain, other PBCs provide average gain)

	A	B	C	D	E (=0.01* B*D)	F	G	H	I (=0.01* D*G*H)	J	K (=E/I)	L	M	N (=0.01*D* H*M/3)	O (=E/N)
	PBC description	Spend (£m) 2005/6		Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2002/04	Outcome elasticity	Change in annual mortality		Cost per life gained (£)		Total life years lost, <75years, 2002/04	Change in annual life years lost	Cost per life year gained (£)
1	Cancer	£4,094		0.968	£39.63		62,259	0.394	237.45		£166,897		2,268,541	2,884	£13,741
2	Circulatory problems	£6,112		0.682	£41.68		45,504	1.370	425.16		£98,042		1,607,171	5,005	£8,328
3	Respiratory problems	£3,421		0.849	£29.04		11,601	1.574	155.03		£187,350		316,506	1,410	£20,601
4	Gastro-intestinal problems	£3,998		0.772	£30.86		5,926	2.018	92.32		£334,318		324,735	1,686	£18,303
5	Big four programmes	£17,625			£141.22		125,290		909.96		£155,196		4,516,953	10,986	£12,855
6	Infectious diseases	£1,161		0.742	£8.61		2,050	0.152	2.31		£3,725,931		106,552	40	£215,054
7	Endocrine problems	£1,832		0.425	£7.79		1,690	0.244	1.75		£4,442,720		60,615	21	£371,601
8	Neurological problems	£2,019		1.111	£22.43		729	0.182	1.47		£15,217,293		66,137	45	£503,201
9	Genito-urinary problems	£3,313		1.041	£34.49		294	0.034	0.10		£331,432,573		10,030	1	£29,144,918
10	Trauma & injuries	£3,758		0.627	£23.56		1,037	1.332	8.66		£2,720,657		30,000	84	£282,132
11	Neonate conditions	£660		0.388	£2.56		2,123	0.237	1.95		£1,311,733		477,675	146	£17,490
12	All ten programmes	£30,368		0.792	£240.67		133,213		926.22		£259,838		5,267,962	11,322	£21,256
13	Other 13 programmes? (a) assume no health gain for GMS/PMS	£8,449		0.926	£78.24				0.00					0	
14	(b) assume average gain in the other 12 PBCs	£25,493		1.272	£324.20				1,247.69		£259,838			15,252	£21,256
15	All 23 programmes	£64,310			£643.10				2,173.90		£295,827			26,575	£24,200
	Note:														
	All 23 programmes	£64,310													
	% change in budget	1.00													
	proportionate change	0.01													
	Change in budget	£643.10													

B7.7 Summary and conclusion

In this section we have extended the results reported by Martin, Rice and Smith[5] by obtaining plausible outcome and expenditure models for all ten programmes of care with a mortality based outcome indicator. In addition, we have, for the first time, calculated the cost of a life year across the big four programmes combined (£12,855) and across all ten programmes (£21,256). Moreover, with the aid of an assumption about the productivity (health gain) of programmes without a meaningful mortality based outcome indicator, we have extended our individual programme estimates to incorporate expenditure across all programmes of care.

If we assume that the other 13 programmes without a mortality-based outcome indicator generate no health gain then the cost of an additional life year across all expenditure for 2005/6 is £56,799.

Alternatively, if we assume that any health care gain associated with primary care expenditure is reflected in mortality rates associated with other, more disease specific, and that the health gain associated with the remaining 12 programmes is, on average, the same as that recorded by the PBCs with a mortality based indicator, then the cost per life year across all expenditure for 2005/6 is £24,200.

This concludes our analysis of the 2005/6 programme budgeting data. In the next section we apply our model to the 2006/7 programme budgeting data.

B8. Analysis of programme budgeting expenditure for 2006/7

B8.1 Construction of an alternative measure of need

The analysis of the 2005/6 programme budgeting data employed a measure of the need for health care that incorporated the AREA resource allocation formula for acute services. As was described in section B6.2, we attempted to construct a better measure of need using a recently developed person-based approach [34]. However, we were unable to construct a viable alternative PBRA based measure of need for use with the PB data for 2005/6 because the PBRA formula only relates to acute services yet the PB data incorporates elements for acute, maternity, mental health, prescribing and primary care, and we were unable to separate these components parts.

The construction of an alternative measure of need is, however, possible for use with the 2006/7 PB data. Spend and mortality data are available for the new (152) PCTs, and the Department of Health's resource allocation exposition book for 2009/10 (which employs the CARAN model) provides separate measures of need for acute, maternity, mental health, prescribing and GMS/PMS services. We can therefore replace the (CARAN-based) measure of acute need for the 2009/10 allocation with our own PBRA based measure of acute need (albeit for 2006/7) to calculate an alternative to the AREA based measure of need across all health care services.

The PBRA model was applied to all patients on Practice lists as at 1 April 2006 to generate a PCT level measure of acute need (see section B7.1 for a description of this approach as applied to patients on Practice lists as at 1 April 2005). The resulting PBRA measure of acute need can be compared with the CARAN based measure of acute need as reported in the Department of Health's resource allocation exposition book for 2009/10. The correlation coefficient for these two measures is 0.8514 and descriptive statistics for the two measures are shown below in Table B8.1.

Table B8.1: table showing summary statistics for the CARAN and PBRA based measures of acute need

Variable	Number of PCTs	Mean	Std. Dev.	Min	Max
CARAN_acute need	152	1.0033	0.1113	0.7659	1.2153
PBRA_acute need	152	1.0037	0.1218	0.7606	1.3420

The all service measure of need (which is a weighted average of the acute, maternity, mental health, prescribing and GMS/PMS measures) as reported in the Department of Health's resource allocation exposition book for 2009/10 can be re-calculated by replacing the CARAN-based acute measure with the PBRA-based acute measure of need. The correlation coefficient for these two all service measures of need is 0.9714 and Figure B8.1 shows a scatter plot of these two measures. Descriptive statistics for these two all service measures of need along with the (AREA-based) PB measure of need are shown below in Table B8.2.

Figure B8.1: graph showing scatter plot of all service measures of need: incorporating CARAN or PBRA based measures of acute need

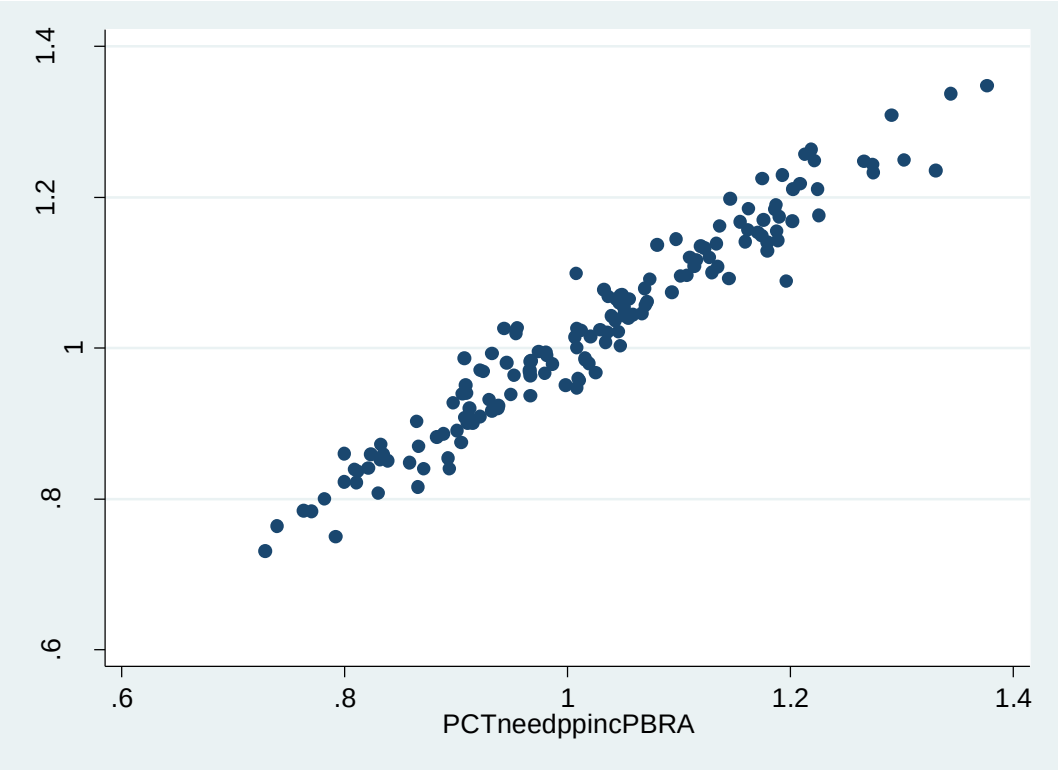


Table B8.2: table showing all service measures of need: incorporating CARAN, PBRA or AREA based measures of acute need

Variable	Obs	Mean	Std. Dev.	Min	Max
PCTneed_CARAN	152	1.0240	0.1339	0.7311	1.3479
PCTneed_PBRA	152	1.0242	0.1395	0.7287	1.3769
PCTneed_AREA	152	1.0293	0.1380	0.7165	1.4006

The correlation coefficients for the three measures are shown in Table B8.3.

Table B8.3: table showing correlation coefficients for alternative measures of all service need

Variable	PCTneed_AREA	PCTneed_PBRA	PCTneed_CARAN
PCTneed_AREA	1		
PCTneed_PBRA	0.9583	1	
PCTneed_CARAN	0.9839	0.9714	1

B8.2 Estimation issues associated with the use of 2006/7 expenditure data

As well as having to select a preferred measure of need from the three available, the estimation of our model using PB data for 2006/7 requires the resolution of several other issues.

Estimation issue 1: 'net spend' or 'own population' spend?
The Department of Health reports two sets of PB spend data: the first is on a 'net spend' basis and the second is on an 'own population' basis. The 'own population' data starts with the 'net spend' figure, adds any expenditure funded from non-NHS sources, and adjusts for expenditure made under PCT lead/host commissioning arrangements. These adjustments are usually very small. For 2005/6 we used the net spend data (because only net spend data was produced in the first year and we were hoping to build a

panel) but given the now regular production of own population data this would seem to be the more appropriate data set to use as, for example, it includes all expenditure irrespective of its funding source.

Estimation issue 2: to weight or not to weight?
OLS and IV estimation implicitly gives the same weight to each PCT when estimating our expenditure and outcome models. With the re-organisation of PCTs in October 2006, the number of such organisations was reduced from 303 to 152. However, far from making them more similar in terms of size (as measured by their population), this re-organisation actually increased the disparity in size between the largest and the smallest PCTs, with the largest PCT now being 14 times the size of the smallest. Unless we explicitly weight each observation (PCT) by its size, we will be giving the same weight (influence) to PCTs that are much smaller than other PCTs.

Estimation issue 3: which MFF?
This study builds on previous work using PB data. Martin, Rice and Smith[3] report the results of the estimation of our model using PB data for 2006/7. One essential step in this estimation is the removal of the impact of unavoidable variations in local costs from the reported measure of the 'unified weighted' population. At the time of the earlier study the authors only had access to an MFF based on HCHS for the new 152 PCTs. Now, however, a more broadly-based MFF is available, that is, one based on a weighted average of MFFs for HCHS, prescribing, and GMS/PMS. Should we use an MFF for HCHS only, or one that incorporates HCHS and prescribing, or one that incorporates HCHS, prescribing and GMS/PMS?

Estimation issue 4: SMRs or SYLLRs, and which proxy for the other programme need variable?
Previous studies have reported results using both SMRs and SYLLRs but the sheer number of models being estimated requires that we focus on one measure only. Various proxies for other programme need have been employed in previous studies (see section B6.3 for further discussion). In this sub-section we persevere with this variety but consistency demands that we focus in on a preferred proxy for other programme need.

This study builds on previous work using PB data. Martin, Rice and Smith[3] report the results of the estimation of our model using PB data for 2006/7. With several alternative measures of need and MFF available, we undertook a preliminary empirical analysis of the 2006/7 PB data using the outcome and expenditure models for the big four programmes as reported in Martin, Rice and Smith[3] as our starting point. These models incorporated the AREA-based measure of need and an MFF based on HCHS only.

We first re-estimated the outcome and expenditure models by replacing the AREA based measure of need with one incorporating the PBRA formula. Then we re-estimated these models again with a measure of need incorporating the CARAN model. The results suggest that: (a) for the outcome models, the use of the PBRA measure of need generates a smaller coefficient on expenditure than does the AREA measure of need; and (b) that for the spend models, the use of the PBRA measure of need generates a larger coefficient on PCT budget than does the AREA measure of need. For both the outcome and expenditure models, the use of the CARAN measure of need generates outcome and expenditure elasticities that lie between those generated by the AREA and PBRA measures.

Next, the results reported by Martin, Rice and Smith[3] employ an MFF based on HCHS only to remove unavoidable variations in local costs from the reported measure of the (unified weighted) need for health care services. This was the only MFF available for the new PCTs at the time of that study. Now, however, a more broadly-based MFF is available (that is, one based on a weighted average of MFFs for HCHS, prescribing, and GMS/PMS).

To examine the consequences of using the CARAN MFF (i.e., a weighted average of the HCHS, prescribing, and GMS/PMS MFFs), this MFF was used to calculate the implied level of need given the unified weighted populations for 2006/7 which are reported alongside the PB spend data by the

Department of Health.[27] We found that the use of an extended set of MFFs can sometimes affect the coefficient on the variable of interest.

Models were also estimated using a weighted average of the CARAN MFFs for HCHS and prescribing only. The latter results were very similar to those using all three of the CARAN MFFs (i.e., a weighted average of the HCHS, prescribing, and GMS/PMS MFFs).

We also tried re-estimating the outcome and expenditure models from Martin, Rice and Smith [3] using the 'own population' expenditure data rather than the 'net spend' data but this adjustment had very little effect on the results. In addition, the impact of 'weighting' each observation by PCT size was usually rather modest.

Because of the sheer number of variations possible, we decided to estimate 13 particular variants of our model and details of these variants are summarised in Table B8.4. These variants were estimated for each of the big four programmes using both the outcome and expenditure equations. The results are presented in Tables B8.5 to B8.12.

Table B8.4: table showing variants of the outcome and expenditure models estimated using 2006/7 spend data

Variant	PCTs weighted?	MFF indicator	Indicator of need	Mortality indicator
1	No weights	HCHS	AREA-based UWP/HCHS MFF	SMR
2	No weights	HCHS	PBRA model applied to patients on list at 1 April 2006	SMR
3	No weights	HCHS	CARAN model used for allocations in 2009/10	SMR
4	No weights	HCHS	AREA-based UWP/HCHS MFF	SYLLR
5	No weights	HCHS	PBRA model applied to patients on list at April 2006	SYLLR
6	No weights	HCHS	CARAN model used for allocations in 2009/10	SYLLR
7	Yes	HCHS	AREA-based UWP/HCHS MFF	SMR
8	No weights	HCHS, prescribing & GMS	AREA-based UWP/(HCHS, prescribing & GMS) MFF	SMR
9	Yes	HCHS, prescribing & GMS	AREA-based UWP/(HCHS, prescribing & GMS) MFF	SMR
10	No weights	HCHS & prescribing	AREA-based UWP/(HCHS & prescribing) MFF	SMR
11	Yes	HCHS & prescribing	AREA-based UWP/(HCHS & prescribing) MFF	SMR
12	No weights	HCHS, prescribing & GMS	CARAN model used for allocations in 2009/10	SMR
13	Yes	HCHS, prescribing & GMS	CARAN model used for allocations in 2009/10	SMR

Note: UWP=unified weighted population.

[27] Note that implied need=unified weighted population/(CARAN MFF*raw population).

Table B8.5: table showing cancer spend models with various indicators of MFF and need

	(1) PBC 2 cancer 2006/7 spend model uses SMR no weighting second stage iAREA need 1 HCHS MFF	(2) PBC 2 cancer 2006/7 spend model uses SMR no weighting second stage PBRA need HCHS MFF	(3) PBC 2 cancer 2006/7 spend model uses SMR no weighting second stage CARAN need HCHS MFF	(4) PBC 2 cancer 2006/7 spend model uses SYLLR no weighting second stage iAREA need 1 HCHS MFF	(5) PBC 2 cancer 2006/7 spend model uses SYLLR no weighting second stage PBRA need HCHS MFF	(6) PBC 2 cancer 2006/7 spend model uses SYLLR no weighting second stage CARAN need HCHS MFF	(7) PBC 2 cancer 2006/7 spend model uses SMR weighted second stage iAREA need 1 HCHS MFF	(8) PBC 2 cancer 2006/7 spend model uses SMR no weighting second stage iAREA need 2 CARAN 3 MFFs	(9) PBC 2 cancer 2006/7 spend model uses SMR weighted second stage iAREA need 2 CARAN 3 MFFs	(10) PBC 2 cancer 2006/7 spend model uses SMR no weighting second stage iAREA need 3 CARAN 2 MFFs	(11) PBC 2 cancer 2006/7 spend model uses SMR weighted second stage iAREA need 3 CARAN 2 MFFs	(12) PBC 2 cancer 2006/7 spend model uses SMR no weighting second stage CARAN need CARAN 3 MFFs	(13) PBC 2 cancer 2006/7 spend model uses SMR weighted second stage CARAN need CARAN 3 MFFs
PCT budget per head	0.353	0.681***	0.572**	0.388	0.752***	0.618**	0.326	0.250	0.245	0.246	0.241	0.552**	0.544*
	[0.273]	[0.235]	[0.247]	[0.272]	[0.238]	[0.247]	[0.362]	[0.284]	[0.357]	[0.284]	[0.357]	[0.239]	[0.308]
needAREA1	1.513***			1.557***			1.351***						
	[0.288]			[0.284]			[0.367]						
other programme need 1	-0.654***	-0.771***	-0.733***				-0.604***	-0.749***	-0.661***	-0.749***	-0.661***	-0.680***	-0.616***
	[0.124]	[0.125]	[0.124]				[0.131]	[0.152]	[0.154]	[0.151]	[0.154]	[0.126]	[0.139]
needPBRA		1.320***			1.352***								
		[0.261]			[0.265]								
needCARAN			1.431***			1.477***						1.347***	1.160***
			[0.287]			[0.289]						[0.246]	[0.294]
other programme need 2				-0.649***	-0.773***	-0.728***							
				[0.119]	[0.124]	[0.119]							
needAREA2								1.778***	1.554***				
								[0.329]	[0.389]				
needAREA3										1.765***	1.545***		
										[0.328]	[0.388]		
Constant	0.271	0.707	0.569	0.356	0.835	0.665	0.067	5.901***	5.559**	5.932***	5.582**	3.430**	3.220
	[0.544]	[0.551]	[0.554]	[0.542]	[0.567]	[0.554]	[0.572]	[2.062]	[2.550]	[2.063]	[2.550]	[1.645]	[2.087]
Observations	152	152	152	152	152	152	152	152	152	152	152	152	152
Endogeneity test statistic	13.112	18.683	18.420	13.313	18.736	18.716	11.940	13.098	11.708	13.017	11.504	16.985	13.460
Endogeneity p-value	0.000293	1.54e-05	1.77e-05	0.000264	1.50e-05	1.52e-05	0.000549	0.000296	0.000622	0.000309	0.000695	3.77e-05	0.000244
Hansen-Sargan test statistic	0.870	1.139	0.560	0.504	0.748	0.281	1.089	1.730	1.875	1.711	1.857	0.381	0.321
Hansen-Sargan p-value	0.351	0.286	0.454	0.478	0.387	0.596	0.297	0.188	0.171	0.191	0.173	0.537	0.571
Shea's partial R-squared	0.607	0.526	0.586	0.570	0.482	0.548	0.612	0.511	0.537	0.510	0.536	0.572	0.583
Kleibergen-Paap LM test statistic	40.55	38.73	40.45	40.49	38.17	41.14	38.91	38.41	37.13	38.38	37.11	43.82	41.36
Kleibergen-Paap p-value	1.57e-09	3.88e-09	1.64e-09	1.61e-09	5.14e-09	1.16e-09	3.55e-09	4.56e-09	8.66e-09	4.63e-09	8.75e-09	3.06e-10	1.05e-09
Kleibergen-Paap F statistic	73.17	63.72	68.50	67.14	51.14	60.78	72.14	57.74	61.44	58.09	61.58	68.29	66.96
Pesaran-Taylor reset statistic	0.233	0.299	0.0271	0.198	0.211	0.00518	0.000324	0.00529	0.0391	0.0345	0.00971	9.41e-07	0.0158
Pesaran-Taylor p-value	0.629	0.585	0.869	0.656	0.646	0.943	0.986	0.942	0.843	0.853	0.922	0.999	0.900

Notes: (a) iAREA need 1=AREA unified weighted population/HCHS MFF

(b) iAREA need 2=AREA unified weighted population/HCHS, prescribing & GMS MFFs

(c) iAREA need 3=AREA unified weighted population/HCHS & prescribing MFFs

(d) other programme need 1=circulatory disease SMR

(e) other programme need 2=circulatory disease SYLLR

(f) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table B8.6: table showing circulatory disease spend models with various indicators of MFF and need

	(1) PBC 10 circulation 2006/7 spend model uses SMR no weighting second stage iAREA need1 HCHS MFF	(2) PBC 10 circulation 2006/7 spend model uses SMR no weighting second stage PBRA need HCHS MFF	(3) PBC 10 circulation 2006/7 spend model uses SMR no weighting second stage CARAN need HCHS MFF	(4) PBC 10 circulation 2006/7 spend model uses SYLLR no weighting second stage iAREA need1 HCHS MFF	(5) PBC 10 circulation 2006/7 spend model uses SYLLR no weighting second stage PBRA need HCHS MFF	(6) PBC 10 circulation 2006/7 spend model uses SYLLR no weighting second stage CARAN need HCHS MFF	(7) PBC 10 circulation 2006/7 spend model uses SMR weighted second stage iAREA need1 HCHS MFF	(8) PBC 10 circulation 2006/7 spend model uses SMR no weighting second stage iAREA need2 CARAN 3 MFFs	(9) PBC 10 circulation 2006/7 spend model uses SMR weighted second stage iAREA need2 CARAN 3 MFFs	(10) PBC 10 circulation 2006/7 spend model uses SMR no weighting second stage iAREA need3 CARAN 2 MFFs	(11) PBC 10 circulation 2006/7 spend model uses SMR weighted second stage iAREA need3 CARAN 2 MFFs	(12) PBC 10 circulation 2006/7 spend model uses SMR no weighting second stage CARAN need CARAN 3 MFFs	(13) PBC 10 circulation 2006/7 spend model uses SMR weighted second stage CARAN need CARAN 3 MFFs
other programme need 1	-0.766**	-0.736**	-0.811**				-0.939***	-0.776**	-0.927***	-0.781**	-0.935***	-0.831**	-1.112***
	[0.298]	[0.299]	[0.370]				[0.305]	[0.315]	[0.324]	[0.316]	[0.326]	[0.367]	[0.392]
PCT budget per head	0.861***	1.162***	1.035***	0.836***	1.191***	0.998***	0.719***	0.832***	0.661**	0.829***	0.657**	0.983***	0.914***
	[0.240]	[0.218]	[0.219]	[0.229]	[0.220]	[0.210]	[0.259]	[0.242]	[0.264]	[0.242]	[0.264]	[0.213]	[0.231]
needAREA1	0.624*			0.732*			0.967**						
	[0.355]			[0.389]			[0.378]						
white ethnic group	0.215***	0.187**	0.207**	0.232***	0.199**	0.225**	0.278***	0.219**	0.284***	0.219**	0.284***	0.209**	0.286***
	[0.079]	[0.080]	[0.086]	[0.083]	[0.083]	[0.095]	[0.084]	[0.085]	[0.091]	[0.085]	[0.091]	[0.086]	[0.098]
provision of unpaid care	0.457**	0.554***	0.488**	0.437**	0.549***	0.466*	0.239	0.527***	0.336	0.528***	0.335	0.477**	0.200
	[0.205]	[0.186]	[0.227]	[0.212]	[0.183]	[0.247]	[0.227]	[0.190]	[0.210]	[0.190]	[0.211]	[0.227]	[0.268]
needPBRA		0.250			0.295								
		[0.275]			[0.280]								
needCARAN			0.480			0.610						0.546	0.925**
			[0.401]			[0.479]						[0.369]	[0.399]
other programme need 2				-0.904**	-0.871**	-0.956*							
				[0.365]	[0.347]	[0.491]							
needAREA2								0.655*	1.020**				
								[0.383]	[0.424]				
needAREA3										0.652*	1.018**		
										[0.381]	[0.422]		
Constant	2.380**	2.377*	2.621*	3.246**	3.244**	3.533*	2.744**	3.763*	5.275**	3.803*	5.337**	2.827	4.028**
	[1.212]	[1.251]	[1.469]	[1.572]	[1.553]	[2.088]	[1.222]	[2.133]	[2.306]	[2.143]	[2.319]	[1.863]	[1.995]
Observations	152	152	152	152	152	152	152	152	152	152	152	152	152
Endogeneity test statistic	8.506	10.727	8.136	6.475	8.793	5.743	9.019	8.939	8.654	9.036	8.745	8.315	9.729
Endogeneity p-value	0.00354	0.00106	0.00434	0.0109	0.00302	0.0166	0.00267	0.00279	0.00326	0.00265	0.00310	0.00393	0.00181
Hansen-Sargan test statistic	2.454	0.640	1.841	2.364	0.423	2.166	2.993	1.770	2.225	1.792	2.237	1.777	2.030
Hansen-Sargan p-value	0.117	0.424	0.175	0.124	0.515	0.141	0.0836	0.183	0.136	0.181	0.135	0.183	0.154
Shea's partial R-squared	0.235	0.205	0.184	0.207	0.184	0.148	0.238	0.225	0.230	0.224	0.228	0.183	0.183
Kleibergen-Paap LM test statistic	22.99	24.15	20.59	21.32	25.01	17.46	26.79	23.47	27.59	23.63	27.59	20.91	24.11
Kleibergen-Paap p-value	1.02e-05	5.70e-06	3.39e-05	2.35e-05	3.70e-06	0.000161	1.52e-06	8.02e-06	1.02e-06	7.41e-06	1.02e-06	2.88e-05	5.80e-06
Kleibergen-Paap F statistic	23.14	22.53	18.64	17.28	20.44	12.62	22.37	21.47	22.07	21.27	21.63	19.10	18.93
Pesaran-Taylor reset statistic	0.00329	0.156	0.102	0.0384	0.333	0.288	0.0270	0.123	0.152	0.189	0.235	0.0165	0.190
Pesaran-Taylor p-value	0.954	0.693	0.750	0.845	0.564	0.592	0.869	0.726	0.696	0.664	0.628	0.898	0.663

Notes: (a) iAREA need 1=AREA unified weighted population/HCHS MFF

(b) iAREA need 2=AREA unified weighted population/HCHS, prescribing & GMS MFFs

(c) iAREA need 3=AREA unified weighted population/HCHS & prescribing MFFs

(d) other programme need 1=cancer SMR

(e) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table B8.7: table showing respiratory problems spend models with various indicators of MFF and need

	(1) PBC 11 respiratory 2006/7 spend model uses SMR no weighting second stage iAREA need1 HCHS MFF	(2) PBC 11 respiratory 2006/7 spend model uses SMR no weighting second stage PBRA need HCHS MFF	(3) PBC 11 respiratory 2006/7 spend model uses SMR no weighting second stage CARAN need HCHS MFF	(4) PBC 11 respiratory 2006/7 spend model uses SYLLR no weighting second stage iAREA need1 HCHS MFF	(5) PBC 11 respiratory 2006/7 spend model uses SYLLR no weighting second stage PBRA need HCHS MFF	(6) PBC 11 respiratory 2006/7 spend model uses SYLLR no weighting second stage CARAN need HCHS MFF	(7) PBC 11 respiratory 2006/7 spend model uses SMR weighted second stage iAREA need1 HCHS MFF	(8) PBC 11 respiratory 2006/7 spend model uses SMR no weighting second stage iAREA need2 CARAN 3 MFFs	(9) PBC 11 respiratory 2006/7 spend model uses SMR weighted second stage iAREA need2 CARAN 3 MFFs	(10) PBC 11 respiratory 2006/7 spend model uses SMR no weighting second stage iAREA need3 CARAN 2 MFFs	(11) PBC 11 respiratory 2006/7 spend model uses SMR weighted second stage iAREA need3 CARAN 2 MFFs	(12) PBC 11 respiratory 2006/7 spend model uses SMR no weighting second stage CARAN need CARAN 3 MFFs	(13) PBC 11 respiratory 2006/7 spend model uses SMR weighted second stage CARAN need CARAN 3 MFFs
PCT budget per head	0.781**	0.992***	0.957***	1.045***	1.315***	1.204***	0.808**	0.592**	0.591*	0.588**	0.585*	0.865***	0.958***
	[0.318]	[0.330]	[0.363]	[0.370]	[0.409]	[0.432]	[0.334]	[0.282]	[0.310]	[0.283]	[0.310]	[0.287]	[0.329]
needAREA1	1.714***			1.741***			1.813***						
	[0.597]			[0.497]			[0.563]						
lone pensioner households	-0.497	-0.243	-0.483	-0.419*	-0.240	-0.380	-0.595*	-0.078	-0.304	-0.089	-0.320	-0.447	-0.556
	[0.346]	[0.243]	[0.356]	[0.252]	[0.210]	[0.271]	[0.346]	[0.285]	[0.378]	[0.286]	[0.380]	[0.337]	[0.344]
other programme need 1	-0.803**	-0.602**	-0.890**				-0.866**	-0.391	-0.664	-0.407	-0.687	-0.834*	-0.931**
	[0.397]	[0.294]	[0.439]				[0.392]	[0.364]	[0.478]	[0.364]	[0.481]	[0.428]	[0.437]
needPBRA		1.176***			1.226***								
		[0.391]			[0.366]								
needCARAN			1.686***			1.720***						1.680***	1.782***
			[0.627]			[0.536]						[0.609]	[0.561]
other programme need 2				-1.109**	-0.955**	-1.197**							
				[0.455]	[0.406]	[0.552]							
needAREA2								1.312**	1.770**				
								[0.667]	[0.803]				
needAREA3										1.325**	1.788**		
										[0.661]	[0.800]		
Constant	-0.143	-0.649	0.258	2.954	2.293	3.531	-0.046	1.605	2.457	1.686	2.578	1.022	0.595
	[1.250]	[0.961]	[1.415]	[2.335]	[2.094]	[2.856]	[1.220]	[2.372]	[2.547]	[2.389]	[2.574]	[1.980]	[2.082]
Observations	152	152	152	152	152	152	152	152	152	152	152	152	152
Endogeneity test statistic	7.821	8.431	9.089	9.157	10.215	9.242	6.984	4.679	4.326	4.853	4.475	9.016	6.863
Endogeneity p-value	0.00516	0.00369	0.00257	0.00248	0.00139	0.00236	0.00822	0.0305	0.0375	0.0276	0.0344	0.00268	0.00880
Hansen-Sargan test statistic	1.655	3.108	0.983	0.214	1.502	0.0135	0.615	4.704	2.922	4.621	2.855	0.866	0.156
Hansen-Sargan p-value	0.198	0.0779	0.321	0.644	0.220	0.908	0.433	0.0301	0.0874	0.0316	0.0911	0.352	0.693
Shea's partial R-squared	0.164	0.211	0.149	0.183	0.203	0.172	0.167	0.161	0.131	0.161	0.131	0.142	0.146
Kleibergen-Paap LM test statistic	18.58	21.08	18.17	22.04	17.51	18.27	18.78	20.00	13.94	20.14	13.90	17.12	16.62
Kleibergen-Paap p-value	9.22e-05	2.64e-05	0.000113	1.64e-05	0.000158	0.000108	8.34e-05	4.55e-05	0.000940	4.23e-05	0.000960	0.000192	0.000246
Kleibergen-Paap F statistic	8.163	14.31	7.772	8.729	8.513	6.885	8.383	13.56	7.241	13.77	7.297	7.776	7.220
Pesaran-Taylor reset statistic	0.238	0.0135	0.141	0.0704	0.0139	0.0164	1.083	2.231	2.206	2.311	2.283	3.699	4.984
Pesaran-Taylor p-value	0.625	0.907	0.707	0.791	0.906	0.898	0.298	0.135	0.138	0.128	0.131	0.0545	0.0256

Notes: (a) iAREA need 1=AREA unified weighted population/HCHS MFF

(b) iAREA need 2=AREA unified weighted population/HCHS, prescribing & GMS MFFs

(c) iAREA need 3=AREA unified weighted population/HCHS & prescribing MFFs

(d) other programme need 1=SMR for all causes of death amenable to health care (see Martin, Rice and Smith 2012)

(e) other programme need 2=SYLLR for all causes of death

(f) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table B8.8: table showing gastro-intestinal problems spend models with various indicators of MFF and need

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	PBC 13 gastro 2006/7 spend model uses SMR no weighting second stage iAREA need1 HCHS MFF	PBC 13 gastro 2006/7 spend model uses SMR no weighting second stage PBRA need HCHS MFF	PBC 13 gastro 2006/7 spend model uses SMR no weighting second stage CARAN need HCHS MFF	PBC 13 gastro 2006/7 spend model uses SYLLR no weighting second stage iAREA need1 HCHS MFF	PBC 13 gastro 2006/7 spend model uses SYLLR no weighting second stage PBRA need HCHS MFF	PBC 13 gastro 2006/7 spend model uses SYLLR no weighting second stage CARAN need HCHS MFF	PBC 13 gastro 2006/7 spend model uses SMR weighted second stage iAREA need1 HCHS MFF	PBC 13 gastro 2006/7 spend model uses SMR no weighting second stage iAREA need2 CARAN 3 MFFs	PBC 13 gastro 2006/7 spend model uses SMR weighted second stage iAREA need2 CARAN 3 MFFs	PBC 13 gastro 2006/7 spend model uses SMR no weighting second stage iAREA need3 CARAN 2 MFFs	PBC 13 gastro 2006/7 spend model uses SMR weighted second stage iAREA need3 CARAN 2 MFFs	PBC 13 gastro 2006/7 spend model uses SMR no weighting second stage CARAN need CARAN 3 MFFs	PBC 13 gastro 2006/7 spend model uses SMR weighted second stage CARAN need CARAN 3 MFFs
PCT budget per head	0.538	0.876**	0.862**	1.058**	1.461***	1.396***	0.627*	0.240	0.271	0.237	0.265	0.509*	0.692**
	[0.355]	[0.371]	[0.414]	[0.446]	[0.513]	[0.533]	[0.371]	[0.301]	[0.334]	[0.301]	[0.334]	[0.305]	[0.340]
lneed	2.627***			2.840***			2.775***						
	[0.851]			[0.758]			[0.755]						
lone pensioner households	-0.838*	-0.269	-0.740	-0.793**	-0.362	-0.736*	-1.080**	0.122	-0.262	0.121	-0.270	-0.612	-0.901**
	[0.492]	[0.347]	[0.520]	[0.385]	[0.314]	[0.422]	[0.460]	[0.290]	[0.375]	[0.287]	[0.374]	[0.453]	[0.445]
other programme need 1	-1.386**	-0.820**	-1.416**				-1.572***	-0.279	-0.751	-0.281	-0.763	-1.216**	-1.510***
	[0.566]	[0.402]	[0.634]				[0.520]	[0.366]	[0.476]	[0.363]	[0.475]	[0.562]	[0.552]
needPBRA		1.422***			1.627***								
		[0.488]			[0.478]								
needCARAN			2.369***			2.740***						2.375***	2.619***
			[0.889]			[0.839]						[0.836]	[0.748]
other programme need 2				-2.093***	-1.524***	-2.250***							
				[0.663]	[0.588]	[0.817]							
needAREA2								1.267*	1.916**				
								[0.662]	[0.794]				
needAREA3										1.264*	1.920**		
										[0.650]	[0.784]		
Constant	2.133	0.486	2.386	8.379**	5.626*	9.374**	2.512	4.107	5.367*	4.129	5.450*	5.180**	4.691**
	[1.763]	[1.237]	[1.998]	[3.338]	[2.958]	[4.164]	[1.605]	[2.520]	[2.774]	[2.524]	[2.787]	[2.437]	[2.310]
Observations	152	152	152	152	152	152	152	152	152	152	152	152	152
Endogeneity test statistic	3.530	1.329	4.118	11.390	5.366	9.900	7.202	0.041	1.085	0.052	1.144	4.849	8.194
Endogeneity p-value	0.0603	0.249	0.0424	0.000738	0.0205	0.00165	0.00728	0.839	0.298	0.820	0.285	0.0277	0.00420
Hansen-Sargan test statistic	7.192	9.154	4.655	2.282	6.867	1.169	5.058	12.98	11.45	12.91	11.31	4.276	2.414
Hansen-Sargan p-value	0.00732	0.00248	0.0310	0.131	0.00878	0.280	0.0245	0.000316	0.000715	0.000327	0.000771	0.0387	0.120
Shea's partial R-squared	0.164	0.211	0.149	0.183	0.203	0.172	0.167	0.161	0.131	0.161	0.131	0.142	0.146
Kleibergen-Paap LM test statistic	18.58	21.08	18.17	22.04	17.51	18.27	18.78	20.00	13.94	20.14	13.90	17.12	16.62
Kleibergen-Paap p-value	9.22e-05	2.64e-05	0.000113	1.64e-05	0.000158	0.000108	8.34e-05	4.55e-05	0.000940	4.23e-05	0.000960	0.000192	0.000246
Kleibergen-Paap F statistic	8.163	14.31	7.772	8.729	8.513	6.885	8.383	13.56	7.241	13.77	7.297	7.776	7.220
Pesaran-Taylor reset statistic	0.00544	0.107	0.00251	0.0613	0.0576	0.000667	0.167	1.735	2.598	1.752	2.633	2.450	3.579
Pesaran-Taylor p-value	0.941	0.743	0.960	0.804	0.810	0.979	0.683	0.188	0.107	0.186	0.105	0.118	0.0585

Notes: (a) iAREA need 1=AREA unified weighted population/HCHS MFF

(b) iAREA need 2=AREA unified weighted population/HCHS, prescribing & GMS MFFs

(c) iAREA need 3=AREA unified weighted population/HCHS & prescribing MFFs

(d) other programme need 1= SMR for all causes of death amenable to health care (see Martin, Rice and Smith 2012)

(e) other programme need 2= SYLLR for all causes of death

(f) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table B8.9: table showing cancer outcome models with various indicators of MFF and need

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	PBC 2 cancer 2006/7 outcome model uses SMR no weighting second stage iAREA need1 HCHS MFF	PBC 2 cancer 2006/7 outcome model uses SMR no weighting second stage PBRA need HCHS MFF	PBC 2 cancer 2006/7 outcome model uses SMR no weighting second stage CARAN need HCHS MFF	PBC 2 cancer 2006/7 outcome model uses SYLLR no weighting second stage iAREA need1 HCHS MFF	PBC 2 cancer 2006/7 outcome model uses SYLLR no weighting second stage PBRA need HCHS MFF	PBC 2 cancer 2006/7 outcome model uses SYLLR no weighting second stage CARAN need HCHS MFF	PBC 2 cancer 2006/7 outcome model uses SMR weighted second stage iAREA need1 HCHS MFF	PBC 2 cancer 2006/7 outcome model uses SMR no weighting second stage iAREA need2 CARAN 3 MFFs	PBC 2 cancer 2006/7 outcome model uses SMR weighted second stage iAREA need2 CARAN 3 MFFs	PBC 2 cancer 2006/7 outcome model uses SMR no weighting second stage implied need CARAN 2 MFFs	PBC 2 cancer 2006/7 outcome model uses SMR weighted second stage iAREA need3 CARAN 2 MFFs	PBC 2 cancer 2006/7 outcome model uses SMR no weighting second stage CARAN need CARAN 3 MFFs	PBC 2 cancer 2006/7 outcome model uses SMR weighted second stage CARAN need CARAN 3 MFFs
lneedAREA1	1.142***			1.048***			1.121***						
	[0.161]			[0.143]			[0.169]						
cancer spend per head	-0.426***	-0.287***	-0.351***	-0.356***	-0.223***	-0.291***	-0.487***	-0.284***	-0.367***	-0.284***	-0.366***	-0.421***	-0.494***
	[0.125]	[0.080]	[0.098]	[0.110]	[0.068]	[0.082]	[0.148]	[0.109]	[0.133]	[0.109]	[0.133]	[0.126]	[0.156]
needPBRA		0.972***			0.881***								
		[0.104]			[0.090]								
needCARAN			1.087***			1.005***						1.126***	1.112***
			[0.130]			[0.110]						[0.153]	[0.167]
needAREA2								1.048***	1.070***				
								[0.128]	[0.143]				
needAREA3										1.035***	1.058***		
										[0.128]	[0.142]		
Constant	3.689***	4.049***	3.884***	4.139***	4.482***	4.309***	3.536***	6.012***	6.375***	6.009***	6.373***	6.614***	6.938***
	[0.318]	[0.202]	[0.249]	[0.278]	[0.172]	[0.207]	[0.372]	[0.476]	[0.583]	[0.476]	[0.583]	[0.552]	[0.684]
Observations	152	152	152	152	152	152	152	152	152	152	152	152	152
Endogeneity test statistic	18.518	16.063	19.086	15.982	12.454	17.096	18.965	11.026	13.552	10.985	13.490	19.697	19.500
Endogeneity p-value	1.68e-05	6.13e-05	1.25e-05	6.40e-05	0.000417	3.55e-05	1.33e-05	0.000898	0.000232	0.000918	0.000240	9.07e-06	1.01e-05
Hansen-Sargan test statistic	0.248	0.239	0.0431	0.163	0.161	0.00820	0.192	0.860	0.690	0.857	0.686	0.000632	0.0933
Hansen-Sargan p-value	0.619	0.625	0.835	0.686	0.688	0.928	0.661	0.354	0.406	0.355	0.407	0.980	0.760
Shea's partial R-squared	0.200	0.246	0.226	0.200	0.246	0.226	0.176	0.202	0.167	0.202	0.166	0.169	0.142
Kleibergen-Paap LM test statistic	18.49	23.43	21.27	18.49	23.43	21.27	16.89	19.40	17.25	19.39	17.20	18.35	15.71
Kleibergen-Paap p-value	9.66e-05	8.17e-06	2.41e-05	9.66e-05	8.17e-06	2.41e-05	0.000215	6.12e-05	0.000179	6.17e-05	0.000184	0.000104	0.000389
Kleibergen-Paap F statistic	19.15	28.08	24.39	19.15	28.08	24.39	15.16	19.90	14.69	19.81	14.58	16.81	12.05
Pesaran-Taylor reset statistic	2.789	5.422	3.218	3.506	5.986	3.796	3.838	4.129	5.271	4.234	5.259	4.399	5.890
Pesaran-Taylor p-value	0.0949	0.0199	0.0728	0.0611	0.0144	0.0514	0.0501	0.0422	0.0217	0.0396	0.0218	0.0360	0.0152

Notes: (a) iAREA need 1=AREA unified weighted population/HCHS MFF

(b) iAREA need 2=AREA unified weighted population/HCHS, prescribing & GMS MFFs

(c) iAREA need 3=AREA unified weighted population/HCHS & prescribing MFFs

(d) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table B8.10: table showing circulatory disease outcome models with various indicators of MFF and need

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	PBC 10 circulation 2006/7 outcome model uses SMR no weighting second stage iAREA need 1 HCHS MFF	PBC 10 circulation 2006/7 outcome model uses SMR no weighting second stage PBRA need HCHS MFF	PBC 10 circulation 2006/7 outcome model uses SMR no weighting second stage CARAN need HCHS MFF	PBC 10 circulation 2006/7 outcome model uses SYLLR no weighting second stage iAREA need 1 HCHS MFF	PBC 10 circulation 2006/7 outcome model uses SYLLR no weighting second stage PBRA need HCHS MFF	PBC 10 circulation 2006/7 outcome model uses SYLLR no weighting second stage CARAN need HCHS MFF	PBC 10 circulation 2006/7 outcome model uses SMR weighted second stage iAREA need 1 HCHS MFF	PBC 10 circulation 2006/7 outcome model uses SMR no weighting second stage iAREA need 2 CARAN 3 MFFs	PBC 10 circulation 2006/7 outcome model uses SMR weighted second stage iAREA need 2 CARAN 3 MFFs	PBC 10 circulation 2006/7 outcome model uses SMR no weighting second stage iAREA need 3 CARAN 2 MFFs	PBC 10 circulation 2006/7 outcome model uses SMR weighted second stage iAREA need 3 CARAN 2 MFFs	PBC 10 circulation 2006/7 outcome model uses SMR no weighting second stage CARAN need CARAN 3 MFFs	PBC 10 circulation 2006/7 outcome model uses SMR weighted second stage CARAN need CARAN 3 MFFs
needAREA1	2.442***			2.657***			2.554***						
	[0.239]			[0.256]			[0.251]						
circulation spend per head	-1.166***	-0.945***	-1.080***	-1.245***	-0.983***	-1.138***	-1.258***	-0.968***	-1.077***	-0.966***	-1.075***	-1.285***	-1.379***
	[0.203]	[0.180]	[0.195]	[0.215]	[0.190]	[0.205]	[0.207]	[0.191]	[0.194]	[0.191]	[0.194]	[0.243]	[0.238]
needPBRA		2.104***			2.262***								
		[0.208]			[0.220]								
needCARAN			2.394***			2.587***						2.508***	2.624***
			[0.242]			[0.257]						[0.278]	[0.282]
needAREA2								2.303***	2.452***				
								[0.218]	[0.233]				
needAREA3										2.281***	2.426***		
										[0.217]	[0.231]		
Constant	1.971***	2.456***	2.165***	1.983***	2.555***	2.222***	1.771***	9.078***	9.596***	9.073***	9.584***	10.605***	11.050***
	[0.429]	[0.379]	[0.411]	[0.454]	[0.400]	[0.432]	[0.438]	[0.916]	[0.931]	[0.914]	[0.928]	[1.168]	[1.145]
Observations	152	152	152	152	152	152	152	152	152	152	152	152	152
Endogeneity test statistic	32.774	30.750	39.253	38.776	28.130	42.881	28.691	28.410	28.030	28.471	27.939	40.272	38.934
Endogeneity p-value	1.04e-08	2.94e-08	3.72e-10	4.75e-10	1.13e-07	5.82e-11	8.49e-08	9.82e-08	1.19e-07	9.51e-08	1.25e-07	2.21e-10	4.38e-10
Hansen-Sargan test statistic	7.315	11.76	3.706	5.337	12.53	3.288	5.937	10.43	7.965	10.24	7.888	2.449	1.678
Hansen-Sargan p-value	0.0625	0.00827	0.295	0.149	0.00576	0.349	0.115	0.0153	0.0467	0.0166	0.0484	0.484	0.642
Shea's partial R-squared	0.368	0.383	0.376	0.368	0.383	0.376	0.349	0.370	0.346	0.371	0.347	0.305	0.291
Kleibergen-Paap LM test statistic	29.32	26.93	31.11	29.32	26.93	31.11	32.68	31.06	34.79	31.25	34.98	28.69	31.43
Kleibergen-Paap p-value	6.73e-06	2.05e-05	2.90e-06	6.73e-06	2.05e-05	2.90e-06	1.39e-06	2.97e-06	5.14e-07	2.72e-06	4.70e-07	9.03e-06	2.50e-06
Kleibergen-Paap F statistic	20.24	19.21	20.32	20.24	19.21	20.32	19.89	20.09	19.66	20.15	19.70	15.50	16.33
Pesaran-Taylor reset statistic	0.0847	0.0884	2.19e-07	0.257	0.0261	0.0107	0.0185	0.0138	0.00282	0.0143	0.00315	2.369	0.196
Pesaran-Taylor p-value	0.771	0.766	1.000	0.612	0.872	0.918	0.892	0.906	0.958	0.905	0.955	0.124	0.658

Notes: (a) iAREA need 1=AREA unified weighted population/HCHS MFF

(b) iAREA need 2=AREA unified weighted population/HCHS, prescribing & GMS MFFs

(c) iAREA need 3=AREA unified weighted population/HCHS & prescribing MFFs

(d) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table B8.11: table showing respiratory disease outcome models with various indicators of MFF and need

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend no weighting second stage iAREA need1 HCHS MFF	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend no weighting second stage PBRA need HCHS MFF	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend no weighting second stage CARAN need HCHS MFF	PBC 11 respiratory 2006/7 outcome model uses SYLLR instrument spend no weighting second stage iAREA need1 HCHS MFF	PBC 11 respiratory 2006/7 outcome model uses SYLLR instrument spend no weighting second stage PBRA need HCHS MFF	PBC 11 respiratory 2006/7 outcome model uses SYLLR instrument spend no weighting second stage CARAN need HCHS MFF	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend weighted second stage iAREA need1 HCHS MFF	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend no weighting second stage iAREA need2 CARAN 3 MFFs	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend weighted second stage iAREA need2 CARAN 3 MFFs	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend no weighting second stage iAREA need3 CARAN 2 MFFs	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend weighted second stage iAREA need3 CARAN 2 MFFs	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend no weighting second stage CARAN need CARAN 3 MFFs	PBC 11 respiratory 2006/7 outcome model uses SMR instrument spend weighted second stage CARAN need CARAN 3 MFFs
needAREA1	8.008***			9.158***			8.647***						
	[2.969]			[3.298]			[3.317]						
respiratory spend per head	-4.845**	-3.364***	-4.149**	-5.568**	-3.894***	-4.808**	-5.182**	-3.535**	-3.773**	-3.536**	-3.764**	-6.738*	-6.640*
	[2.147]	[1.225]	[1.734]	[2.388]	[1.359]	[1.945]	[2.352]	[1.412]	[1.503]	[1.400]	[1.479]	[3.799]	[3.464]
needPBRA		5.941***			6.789***								
		[1.706]			[1.887]								
needCARAN			7.238***			8.306***						10.184**	10.352**
			[2.492]			[2.788]						[5.006]	[4.635]
needAREA2								6.501***	7.025***				
								[1.985]	[2.176]				
needAREA3										6.460***	6.965***		
										[1.959]	[2.130]		
Constant	-10.277*	-6.163*	-8.328*	-12.218*	-7.567**	-10.087*	-11.234*	17.749***	18.712***	17.755***	18.675***	31.101**	30.667**
	[5.898]	[3.355]	[4.749]	[6.563]	[3.723]	[5.328]	[6.465]	[5.877]	[6.252]	[5.824]	[6.151]	[15.828]	[14.436]
Observations	152	152	152	152	152	152	152	152	152	152	152	152	152
Endogeneity test statistic	51.569	49.552	54.608	55.731	52.069	58.974	48.137	42.431	44.464	42.671	44.683	57.889	53.094
Endogeneity p-value	0	0	0	0	0	0	0	7.32e-11	0	6.48e-11	0	0	0
Hansen-Sargan test statistic	0.302	1.253	0.123	0.305	1.828	0.211	0.179	0.785	0.354	0.700	0.303	0.00383	0.0915
Hansen-Sargan p-value	0.582	0.263	0.726	0.581	0.176	0.646	0.673	0.376	0.552	0.403	0.582	0.951	0.762
Shea's partial R-squared	0.0491	0.0791	0.0593	0.0491	0.0791	0.0593	0.0462	0.0654	0.0624	0.0661	0.0633	0.0235	0.0246
Kleibergen-Paap LM test statistic	5.660	8.303	7.499	5.660	8.303	7.499	5.437	7.311	6.973	7.461	7.117	3.866	3.772
Kleibergen-Paap p-value	0.0590	0.0157	0.0235	0.0590	0.0157	0.0235	0.0660	0.0258	0.0306	0.0240	0.0285	0.145	0.152
Kleibergen-Paap F statistic	3.344	5.857	4.507	3.344	5.857	4.507	2.959	4.328	3.804	4.402	3.875	2.030	1.859
Pesaran-Taylor reset statistic	0.791	4.049	0.000560	1.490	5.225	0.00355	0.327	0.0202	0.00861	0.0218	0.0116	3.788	1.716
Pesaran-Taylor p-value	0.374	0.0442	0.981	0.222	0.0223	0.952	0.568	0.887	0.926	0.883	0.914	0.0516	0.190

Notes: (a) iAREA need 1=AREA unified weighted population/HCHS MFF

(b) iAREA need 2=AREA unified weighted population/HCHS, prescribing & GMS MFFs

(c) iAREA need 3=AREA unified weighted population/HCHS & prescribing MFFs

(d) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table B8.12: table showing gastro-intestinal disease outcome models with various indicators of MFF and need

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	PBC 13 gastro-intestinal 2006/7 outcome model uses SMR instrument spend no weighting second stage iAREA need1 HCHS MFF	PBC 13 gastro-intestinal 2006/7 outcome model uses SMR instrument spend no weighting second stage PBRA need HCHS MFF	PBC 13 gastro-intestinal 2006/7 outcome model uses SMR instrument spend no weighting second stage CARAN need HCHS MFF	PBC 13 gastro-intestinal 2006/7 outcome model uses SYLLR instrument spend no weighting second stage iAREA need1 HCHS MFF	PBC 13 gastro-intestinal 2006/7 outcome model uses SYLLR instrument spend no weighting second stage PBRA need HCHS MFF	PBC 13 gastro-intestinal 2006/7 outcome model uses SYLLR instrument spend no weighting second stage CARAN need HCHS MFF	PBC 13 gastro-intestinal 2006/7 outcome model uses SMR instrument spend weighted second stage iAREA need1 HCHS MFF	PBC 13 gastro-intestinal 2006/7 outcome model uses SMR instrument spend no weighting second stage iAREA need2 CARAN 3 MFFs	PBC 13 gastro-intestinal 2006/7 outcome model uses SMR instrument spend weighted second stage iAREA need2 CARAN 3 MFFs	PBC 13 gastro-intestinal 2006/7 outcome model uses SMR instrument spend no weighting second stage iAREA need3 CARAN 2 MFFs	PBC 13 gastro-intestinal 2006/7 outcome model uses SMR instrument spend weighted second stage iAREA need3 CARAN 2 MFFs	PBC 13 gastro-intestinal 2006/7 outcome model uses SMR instrument spend no weighting second stage CARAN need CARAN 3 MFFs	PBC 13 gastro 2006/7 outcome model uses SMR instrument spend weighted second stage CARAN need CARAN 3 MFFs
needAREA1	3.853***			3.966***			3.779***						
	[0.551]			[0.558]			[0.499]						
gastro spend per head	-1.755***	-1.420***	-1.641***	-1.544***	-1.180***	-1.404***	-1.750***	-1.275***	-1.317***	-1.275***	-1.315***	-2.056***	-2.192***
	[0.397]	[0.353]	[0.427]	[0.399]	[0.358]	[0.429]	[0.385]	[0.335]	[0.326]	[0.335]	[0.325]	[0.589]	[0.574]
needPBRA		3.342***			3.413***								
		[0.486]			[0.498]								
needCARAN			3.794***			3.887***						4.140***	4.250***
			[0.612]			[0.621]						[0.768]	[0.710]
needAREA2								3.426***	3.479***				
								[0.466]	[0.419]				
needAREA3										3.393***	3.443***		
										[0.462]	[0.415]		
Constant	-2.155**	-1.251	-1.838	-0.954	0.028	-0.566	-2.166**	7.919***	8.073***	7.916***	8.064***	11.273***	11.835***
	[1.047]	[0.928]	[1.121]	[1.054]	[0.943]	[1.127]	[1.016]	[1.431]	[1.391]	[1.430]	[1.387]	[2.524]	[2.460]
Observations	152	152	152	152	152	152	152	152	152	152	152	152	152
Endogeneity test statistic	23.347	18.985	25.405	17.048	11.389	17.857	16.834	16.638	11.980	16.689	11.942	25.632	22.341
Endogeneity p-value	1.35e-06	1.32e-05	4.65e-07	3.65e-05	0.000739	2.38e-05	4.08e-05	4.52e-05	0.000538	4.40e-05	0.000549	4.13e-07	2.28e-06
Hansen-Sargan test statistic	3.067	4.604	1.555	4.936	7.575	2.637	7.476	5.029	8.762	4.907	8.714	1.284	3.554
Hansen-Sargan p-value	0.216	0.100	0.459	0.0847	0.0227	0.268	0.0238	0.0809	0.0125	0.0860	0.0128	0.526	0.169
Shea's partial R-squared	0.193	0.231	0.200	0.193	0.231	0.200	0.191	0.208	0.198	0.208	0.198	0.139	0.135
Kleibergen-Paap LM test statistic	16.47	17.51	16.32	16.47	17.51	16.32	17.68	17.09	18.26	17.14	18.34	13.39	13.98
Kleibergen-Paap p-value	0.000910	0.000556	0.000974	0.000910	0.000556	0.000974	0.000511	0.000679	0.000389	0.000661	0.000375	0.00386	0.00293
Kleibergen-Paap F statistic	12.12	13.24	10.79	12.12	13.24	10.79	11.96	13.23	12.98	13.24	13.00	7.550	7.248
Pesaran-Taylor reset statistic	0.233	0.0427	0.0897	1.246	1.121	0.258	0.170	0.0935	0.443	0.0893	0.411	0.00841	0.117
Pesaran-Taylor p-value	0.629	0.836	0.765	0.264	0.290	0.611	0.680	0.760	0.506	0.765	0.521	0.927	0.732

Notes: (a) iAREA need 1=AREA unified weighted population/HCHS MFF

(b) iAREA need 2=AREA unified weighted population/HCHS, prescribing & GMS MFFs

(c) iAREA need 3=AREA unified weighted population/HCHS & prescribing MFFs

(d) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

The assimilation of the impact of alternative measures of need, weights, and MFFs proved overwhelming. Instead, we approached the selection of the appropriate need~weighting~MFF combination from an a priori perspective. The AREA-based need formula has been replaced by the CARAN formula for the purposes of resource allocation and therefore it must be believed to be a better indicator of relative health care need. The PBRA approach is relatively new and has not been implemented yet. We therefore decided to use the CARAN based measure as our indicator of the level of need.

With some PCTs several times larger than others, it is difficult to justify giving them all the same weighting. It was therefore decided to weight all of our models by PCT size (where size is measured by the PCT's population).

We also decided to use the 'own population' expenditure data on the grounds that all NHS expenditure, irrespective of its funding source, should be included in the analysis (although there is the issue about how this income is split between PBCs).

Finally, it was decided to focus on the use of the SYLLR as the outcome indicator, and to proxy 'other programme need' in the expenditure equation using the all cause SYLLR minus the own programme SYLLR.

B8.3 Model estimation using 2006/7 expenditure data and mortality data for 2004/2006: CARAN need and three MFFs

Initially, acceptable models were obtained using the CARAN measure of need and adjusting expenditure for local input prices using a weighted average of the MFFs for all three services (HCHS, prescribing, and GMS/PMS). The outcome and expenditure results for the big four programmes are shown in Table B8.13 with the relevant outcome and expenditure elasticities highlighted.

In all four outcome models expenditure has a significant negative effect on mortality and, in three of these; the all service measure of need has a significant positive effect. In the respiratory outcome model, where the all service need term is not significant, there is another indicator of need – the proportion of the population that are permanently sick – and this is both positive and statistically significant. The diagnostic statistics suggest that, in all four cases, own programme expenditure is endogenous and that the instruments are valid. They also suggest that the instruments are relevant. There is some evidence that the instruments are slightly weak in one of the four outcome results (the respiratory model).[28] The Pesaran-Taylor test suggests that there is no evidence of model mis-specification.

In all four expenditure models both the need and budget variables have a positive and significant effect on own programme expenditure. In addition, the proxy for need in other programmes is negative and significant in all four cases. In the gastro-intestinal expenditure programme the prevalence of lone pensioners households is associated with less NHS expenditure; there might be some unmet need here or perhaps this is a self-selecting group.

The diagnostic statistics suggest that, for all four expenditure models, the proxy for other programme need is endogenous and that the instruments are valid. They also suggest that the instruments are relevant and, with the possible exception of the gastro-intestinal expenditure result, there is no evidence that the instruments are weak. The Pesaran-Taylor test suggests that there is no evidence of model mis-specification.

The elasticities shown in Table B8.13 can be used to calculate the cost of a life year in each programme and these calculations -- for both these four programmes as well as for the other six programmes with a mortality based outcome indicator -- are shown in Tables B8.14 and B8.15 (the full outcome and expenditure models for the other six programmes with a mortality based outcome indicator are not shown here).

[28] Ideally, the test F statistic should equal to or greater than ten.

Table B8.14 reveals that the cost of a life year for the big four programmes combined is £11,298. This is remarkably close to the figure obtained using expenditure data for 2005/6, an AREA-based measure of need, and a HCHS MFF (£12,855). The cost of a life year for all ten programmes with a mortality based measure of need the cost of a life year is £21,743, which is even closer to the figure obtained using 2005/6 expenditure data (£21,256). If we assume a zero gain in the 13 programmes without a mortality based indicator then the cost per life year across all 23 programmes is £66,318 (it is £56,799 for 2005/6 data).

Alternatively, if we assume that PBC23 generates a zero health gain and that the gain attributable to the remaining 12 programmes is, on average, the same as that attributable to those with a mortality outcome measure, then Table B8.15 shows that the cost of a life year across all programmes is £25,038 (it is £24,200 for 2005/6 data).

Table B8.13: table showing outcome and expenditure models for the big four programmes using spend data (incorporating three MFFs) for 2006/7

Regressors	(1) PBC 2 cancer 2006/7 outcome model	(2) PBC 2 cancer 2006/7 spend model	(3) PBC 10 circulation 2006/7 outcome model	(4) PBC 10 circulation 2006/7 spend model	(5) PBC 11 respiratory 2006/7 outcome model	(6) PBC 11 respiratory 2006/7 spend model	(7) PBC 13 gastro 2006/7 outcome model	(8) PBC 13 gastro 2006/7 spend model
all cause SYLLR excluding cancer		-0.952***						
		[0.179]						
budget per head (HPG MFF)		0.542**		0.694**		0.712***		0.650**
		[0.242]		[0.292]		[0.252]		[0.289]
need CARAN	0.958***	1.765***	2.830***	2.185***	1.764	1.371***	4.609***	2.696***
	[0.129]	[0.286]	[0.252]	[0.355]	[1.192]	[0.297]	[0.700]	[0.679]
own programme spend per head	-0.351***		-1.441***		-2.830***		-2.125***	
	[0.117]		[0.219]		[0.767]		[0.563]	
all cause SYLLR excluding circulatory problems				-1.782***				
				[0.336]				
permanently sick					1.371***			
					[0.405]			
all cause SYLLR excluding respiratory problems						-0.670**		
						[0.288]		
all cause SYLLR excluding gastro-intestinal problems								-1.856***
								[0.612]
lone pensioner households								-0.593**
								[0.297]
Constant	6.588***	5.937***	11.538***	10.299***	18.965***	3.117	12.208***	9.752***
	[0.515]	[1.775]	[1.050]	[2.384]	[3.853]	[1.976]	[2.416]	[3.053]
Observations	152	152	152	152	152	152	152	152
Endogeneity test statistic	14.496	20.274	43.352	25.784	27.923	7.922	21.862	13.531
Endogeneity p-value	0.000140	6.71e-06	0	3.82e-07	1.26e-07	0.00488	2.93e-06	0.000235
Hansen-Sargan test statistic	0.208	0.293	1.507	0.542	1.879	0.356	1.006	0.0267
Hansen-Sargan p-value	0.649	0.588	0.681	0.462	0.170	0.550	0.316	0.870
Shea's partial R-squared	0.163	0.445	0.303	0.296	0.0802	0.366	0.142	0.206
Kleibergen-Paap LM test statistic	16.97	42.38	32.53	32.70	10.51	36.33	15.00	19.07
Kleibergen-Paap p-value	0.000207	6.28e-10	1.49e-06	7.93e-08	0.00523	1.29e-08	0.000553	7.22e-05
Kleibergen-Paap F statistic	12.47	48.32	17.31	25.71	7.482	24.32	11.80	8.660
Pesaran-Taylor reset statistic	5.471	0.00111	0.0912	0.0183	3.090	1.915	0.267	0.0880
Pesaran-Taylor p-value	0.0193	0.973	0.763	0.892	0.0788	0.166	0.605	0.767

Robust standard errors in brackets. ***p<0.01, ** p<0.05, * p<0.1

Note: all spend figures are on a net population basis and are adjusted for local prices using three MFFs from the Department of Health's resource allocation exposition book for 2009/10. All estimated models use 152 PCTs and are weighted by PCT population. The SYLLR is the mortality indicator.

There are several differences between the models estimated here and those reported in Martin, Rice and Smith (2008b):

(i) here we use net population spend data (not net spend data);

(ii) here we use three MFFs (not solely the HCHS MFF); and

(iii) here we use a consistent definition of the 'other programme need' proxy across all programmes (i.e., all cause SYLLR minus the own programme SYLLR).

Table B8.14: table showing cost of life and life year estimates using spend data for 2006/7 (three MFFs) and outcome data for 2004/06 (assumes zero gain for 13 programmes)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
				=0.01*C*D				=0.01*D*G*H		=E/I			=0.01*D*H*M/3	=E/N
	PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)		Average annual mortality, <75 years, 2004/06	Outcome elasticity	Change in annual mortality		Cost per life gained (£)		Total life years lost, <75years, 2004/06	Change in annual life years lost	Cost per life year gained (£)
1	Cancer	£4,122	0.542	£22.34		62,259	0.351	118.44		£188,625		2,221,530	1,409	£15,859
2	Circulatory problems	£6,161	0.694	£42.76		45,504	1.441	455.06		£93,959		1,463,912	4,880	£8,762
3	Respiratory problems	£3,285	0.712	£23.39		11,601	2.83	233.76		£100,058		321,264	2,158	£10,839
4	Gastro-intestinal problems	£3,700	0.65	£24.05		5,926	2.125	81.85		£293,820		328,853	1,514	£15,884
5	Big four programmes	£17,268		£112.54		125,290		889.12		£126,573		4,335,559	9,961	£11,298
6	*Big four programmes 2005/6*	*£17,625*		*£141.22*		*125,290*		*909.96*		*£155,196*		*4,516,953*	*10,986*	*£12,855*
7	Infectious diseases	£1,053	0.725	£7.63		2,050	0.03	0.45		£17,121,951		101604	7	£1,036,377
8	Endocrine problems	£1,852	0.954	£17.67		1,690	0.965	15.56		£1,135,604		60,615	186	£94,985
9	Neurological problems	£2,790	0.64	£17.86		729	0.1	0.47		£38,271,605		68,808	15	£1,216,428
10	Genito-urinary problems	£3,482	0.799	£27.82		294	0.074	0.17		£160,047,803		11,554	2	£12,217,601
11	Trauma & injuries	£2,892	0.609	£17.61		1,037	0.527	3.33		£5,291,867		30,000	32	£548,767
12	Maternity* & neonates	£3,574	0.601	£21.48		2,123	0.036	0.46		£46,762,966		484,950	35	£614,153
13	Other six programmes	£15,643		£110.07		7,923		20.43		£5,387,190		757,531	277	£396,796
14	*Other six programmes 2005/6*	*£12,743*		*£99.44*		*7,923*		*16.26*		*£6,115,621*		*751,009*	*337*	*£295,074*
15	All ten programmes	£32,911	0.676	£222.61		133,213		909.55		£244,747		5,093,090	10,238	£21,743
16	*All ten programmes 2005/6*	*£30,368*	*0.792*	*£240.67*		*133,213*		*926.22*		*£259,838*		*5,267,962*	*11,322*	*£21,256*
	Assume zero health gain in the other 13 programmes													
18	Other 13 programmes	£34,985	1.304	£456.35				0.00					0	
19	*Other 13 programmes 2005/6*	*£33,942*	*1.186*	*£402.43*				*0.00*					*0*	
20	All 23 programmes	£67,896		£678.96				909.55		£746,481			10,238	£66,318
21	*All 23 programmes 2005/6*	*£64,310*		*£643.10*				*926.22*		*£694,330*			*11,322*	*£56,799*
	Note:	2006/7	2005/6											
22	All 23 programme spend	£67,896	£64,310											
23	% change in budget	1.00	1.00											
24	proportionate change	0.01	0.01											
25	Change in budget	£678.96	£643.10											

Note that the annual mortality figures reported in cells G5 & G6 and G13 & G14 are identical because we do not have mortality data for 2002/04.

Note that, for 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

Table B8.15: table showing cost of life and life year estimates using spend data for 2006/7 (three MFFs) and outcome data for 2004/06 (assumes some gain in other 13 programmes)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
				=0.01*C*D				=0.01*D*G*H		=E/I			=0.01*D*H*M/3	=E/N
	PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)		Average annual mortality, <75years, 2004/06	Outcome elasticity	Change in annual mortality		Cost per life gained (£)		Total life years lost, <75years, 2004/06	Change in annual life years lost	Cost per life year gained (£)
1	Cancer	£4,122	0.542	£22.34		62,259	0.351	118.44		£188,625		2,221,530	1,409	£15,859
2	Circulatory problems	£6,161	0.694	£42.76		45,504	1.441	455.06		£93,959		1,463,912	4,880	£8,762
3	Respiratory problems	£3,285	0.712	£23.39		11,601	2.83	233.76		£100,058		321,264	2,158	£10,839
4	Gastro-intestinal disease	£3,700	0.65	£24.05		5,926	2.125	81.85		£293,820		328,853	1,514	£15,884
5	Big four programmes	£17,268		£112.54		125,290		889.12		£126,573		4,335,559	9,961	£11,298
6	*Big four programmes 2005/6*	*£17,625*		*£141.22*		*125,290*		*909.96*		*£155,196*		*4,516,953*	*10,986*	*£12,855*
7	Infectious diseases	£1,053	0.725	£7.63		2,050	0.03	0.45		£17,121,951		101604	7	£1,036,377
8	Endocrine problems	£1,852	0.954	£17.67		1,690	0.965	15.56		£1,135,604		60,615	186	£94,985
9	Neurological problems	£2,790	0.64	£17.86		729	0.1	0.47		£38,271,605		68,808	15	£1,216,428
10	Genito-urinary problems	£3,482	0.799	£27.82		294	0.074	0.17		£160,047,803		11,554	2	£12,217,601
11	Trauma & injuries	£2,892	0.609	£17.61		1,037	0.527	3.33		£5,291,867		30,000	32	£548,767
12	Maternity* & neonates	£3,574	0.601	£21.48		2,123	0.036	0.46		£46,762,966		484,950	35	£614,153
13	Other six programmes	£15,643		£110.07		7,923		20.43		£5,387,190		757,531	277	£396,796
14	*Other six PBCs 2005/6*	*£12,743*		*£99.44*		*7,923*		*16.26*		*£6,115,621*		*751,009*	*337*	*£295,074*
15	All ten programmes	£32,911	0.676	£222.61		133,213		909.55		£244,747		5,093,090	10,238	£21,743
16	*All ten programmes 2005/6*	*£30,368*	*0.792*	*£240.67*		*133,213*		*926.22*		*£259,838*		*5,267,962*	*11,322*	*£21,256*
	Assume zero health gain in PBC23, and gain in ten PBCs applies to other 12 PBCs													
17	PBC23	£10,585	0.844	£89.34				0.00					0.00	
18	*PBC23 2005/6*	*£8,449*	*0.926*	*£78.24*				*0.00*					*0.00*	
19	Other 12 programmes	£24,400		£367.01				1,499.56		£244,747			16,880	£21,743
20	*Other 12 PBCs 2005/6*	*£25,493*		*£324.20*				*1,247.69*		*£259,838*			*15,252*	*£21,256*
21	All 23 programmes	£67,896		£678.96				2,409.11		£281,830			27,118	£25,038
22	*All 23 programmes 2005/6*	*£64,310*		*£643.10*				*2,173.90*		*£295,827*			*26,575*	*£24,200*
	Note:	2006/7	2005/6											
23	All 23 programme spend	£67,896	£64,310											
24	% change in budget	1.00	1.00											
25	proportionate change	0.01	0.01											
26	Change in budget	£678.96	£643.10			See also notes to Table B8.14 immediately above.								

B8.4 Model estimation using 2006/7 expenditure data and mortality data for 2004/2006: CARAN need and two MFFs

Further discussion by the project team noted that the PB data incorporates all PCT expenditure and that, as there is a separate category for GMS/PMS expenditure (PBC23a), it seems appropriate that the GMS/PMS MFF should be applied to this category. However, other categories of expenditure exclude GMS/PMS expenditure but incorporate both HCHS and prescribing expenditure. It therefore seems appropriate that a weighted averaged of the HCHS and prescribing MFFs should be applied to these other (non-GMS/PMS) categories of expenditure.

We therefore re-estimated the outcome and expenditure models for those programmes with a mortality based outcome indicator using the CARAN measure of need and adjusting expenditure for local input prices using the MFFs for HCHS and prescribing services. The outcome and expenditure results for the big four programmes are shown in Table B8.16 with the relevant outcome and expenditure elasticities again highlighted (the first-stage regressions associated with these results can be found in Table BA.5 in the annex).

In all four outcome models expenditure has a significant negative effect on mortality and, in three of these, the all service measure of need has a significant positive effect. In the respiratory outcome model, where the all service need term is not significant, there is another indicator of need – the proportion of the population that are permanently sick – and this is both positive and statistically significant. The all service measure of need squared is also positive and significant in the cancer outcome equation. The diagnostic statistics suggest that, in all four cases, own programme expenditure is endogenous and that the instruments are valid. They also suggest that the instruments are relevant. There is a little evidence that the instruments are weak in one of the four outcome results, namely the respiratory model. Re-estimation of the latter model but without the least significant instrument generates a coefficient of -3.507 on expenditure and the Kleibergen-Paap F statistic now exceeds ten (it is 11.799). The Pesaran-Taylor test suggests that there is no evidence of model mis-specification in any of the outcome models.

In all four expenditure models both the need and budget variables have a positive and significant effect on own programme expenditure. In addition, the proxy for need in other programmes is negative and significant in all four cases. In the gastro-intestinal expenditure programme the prevalence of lone pensioners households is associated with less NHS expenditure; there might be some unmet need here or perhaps this is self-selecting group.

The diagnostic statistics suggest that, for all four expenditure models, expenditure is endogenous and the instruments are valid. They also suggest that the instruments are relevant and, with the possible exception of the gastro-intestinal expenditure result, there is no evidence that the instruments are weak. Re-estimation of the gastro-intestinal expenditure model without the least significant instrument generates a coefficient of 0.667 on the budget variable and the Kleibergen-Paap F statistic now exceeds ten (it is 16.871). The Pesaran-Taylor test suggests that there is no evidence of model mis-specification.

The outcome and expenditure elasticities are little changed from those presented in Table B8.13 and, like those, these new elasticities can be used to calculate the cost of a life year in each programme. These calculations -- for both these four programmes as well as for the other six programmes with a mortality based outcome indicator -- are shown in Tables B8.17 and B8.18 (the full outcome and expenditure models for the other six programmes with a mortality based outcome indicator are not shown here).

The figures for 2006/7 in Table B8.17 (which incorporate two MFFs) can be compared with those for 2006/7 in Table B8.14 (which incorporate three MFFs). Table B8.17 reveals that the use of a different MFF has little impact on the cost of a life year for the big four PBCs (it was £11,298, it is now £10,783) as well as on the cost of a life year for all programmes with a mortality outcome measure (was £21,743, now £20,893).

In addition, Table B8.18 shows that if we assume that PBC23 generates a zero health gain and that the gain attributable to the remaining 12 programmes is, on average, the same as that attributable to those

with a mortality outcome measure, then the cost of a life year across all programmes is now £23,697 (it was £25,038 for 2006/7 in Table B8.15).

The figures in Table B8.18 also reveal that the cost of a life year in 2006/7 for all programmes (£23,697) is little changed from the comparable figure for 2005/6 (£24,200).

Table B8.16: table showing outcome and expenditure models for the big four programmes using spend data for 2006/7 (incorporating two MFFs) and mortality data for 2004/5/6

	(1) PBC 2 cancer 2006/7 outcome model	(2) PBC 2 cancer 2006/7 spend model	(3) PBC 10 circulation 2006/7 outcome model	(4) PBC 10 circulation 2006/7 spend model	(5) PBC 11 respiratory 2006/7 outcome model	(6) PBC 11 respiratory 2006/7 spend model	(7) PBC 13 gastro 2006/7 outcome model	(8) PBC 13 gastro 2006/7 spend model
own programme spend per head	-0.337***		-1.447***		-2.839***		-2.137***	
	[0.104]		[0.220]		[0.772]		[0.569]	
need CARAN per head	0.974***	1.772***	2.860***	2.191***	1.782	1.375***	4.657***	2.697***
	[0.110]	[0.287]	[0.257]	[0.355]	[1.198]	[0.297]	[0.716]	[0.676]
needCARAN per head squared	1.314***							
	[0.352]							
all cause SYLLR excluding cancer		-0.951***						
		[0.180]						
PCT budget per head		0.548**		0.701**		0.718***		0.655**
		[0.242]		[0.292]		[0.253]		[0.289]
all cause SYLLR excluding circulatory disease				-1.778***				
				[0.336]				
permanently sick aged 16-74					1.385***			
					[0.405]			
all cause SYLLR excluding respiratory problems						-0.663**		
						[0.288]		
all cause SYLLR excluding gastro-intestinal problems								-1.847***
								[0.609]
lone pensioner households								-0.590**
								[0.295]
Constant	6.506***	5.881***	11.567***	10.227***	19.047***	3.032	12.260***	9.664***
	[0.455]	[1.778]	[1.058]	[2.387]	[3.877]	[1.977]	[2.441]	[3.046]
Endogeneity test statistic	15.173	20.248	43.405	25.854	27.876	7.863	21.853	13.607
Endogeneity p-value	9.81e-05	6.80e-06	0	3.68e-07	1.29e-07	0.00505	2.94e-06	0.000225
Hansen-Sargan test statistic	0.00201	0.306	1.440	0.530	1.912	0.344	1.011	0.0294
Hansen-Sargan p-value	0.964	0.580	0.696	0.467	0.167	0.557	0.315	0.864
Shea's partial R-squared	0.164	0.445	0.300	0.296	0.0793	0.366	0.140	0.206
Kleibergen-Paap LM test statistic	17.85	42.38	32.37	32.70	10.42	36.33	14.86	19.07
Kleibergen-Paap p-value	0.000133	6.28e-10	1.61e-06	7.93e-08	0.00545	1.29e-08	0.000592	7.22e-05
Kleibergen-Paap F statistic	13.28	48.32	17.14	25.71	7.390	24.32	11.63	8.660
Pesaran-Taylor reset statistic	0.00226	0.00178	0.0945	0.0215	3.139	1.908	0.266	0.0605
Pesaran-Taylor p-value	0.962	0.966	0.759	0.883	0.0764	0.167	0.606	0.806

Note: robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table B8.17: table showing cost of life and life year estimates using spend data for 2006/7 (two MFFs) and outcome data for 2004/06 (assumes zero gain for 13 programmes)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
				=0.01*C*D				=0.01*D*G*H		=E/I			=0.01*D*H*M/3	=E/N
	PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)		Average annual mortality, <75years, 2004/06	Outcome elasticity	Change in annual mortality		Cost per life gained (£)		Total life years lost, <75years, 2004/06	Change in annual life years lost	Cost per life year gained (£)
1	Cancer	£4,122	0.548	£22.59		62,259	0.337	114.98		£196,461		2,221,530	1,368	£16,518
2	Circulatory problems	£6,161	0.701	£43.19		45,504	1.447	461.57		£93,569		1,463,912	4,950	£8,725
3	Respiratory problems	£3,285	0.718	£23.59		11,601	3.507	292.12		£80,743		321,264	2,697	£8,747
4	Gastro-intestinal problems	£3,700	0.667	£24.68		5,926	2.137	84.47		£292,170		328,853	1,562	£15,795
5	Big four programmes	£17,268		£114.04		125,290		953.13		£119,650		4,335,559	10,576	£10,783
6	*Big four programmes 2005/6*	*£17,625*		*£141.22*		*125,290*		*909.96*		*£155,196*		*4,516,953*	*10,986*	*£12,855*
7	Infectious diseases	£1,053	0.731	£7.70		2,050	0.03	0.45		£17,121,951		101,604	7	£1,036,377
8	Endocrine problems	£1,852	0.966	£17.89		1,690	0.812	13.26		£1,349,579		60,615	158	£112,882
9	Neurological problems	£2,790	0.648	£18.08		729	0.098	0.46		£39,052,658		68,808	15	£1,241,253
10	Genito-urinary problems	£3,482	0.837	£29.14		294	0.073	0.18		£162,240,239		11,554	2	£12,384,965
11	Trauma & injuries	£2,892	0.617	£17.84		1,037	0.527	3.37		£5,291,867		30,000	33	£548,767
12	Maternity* & neonates	£3,574	0.601	£21.48		2,123	0.035	0.45		£48,099,051		484,950	34	£631,700
13	Other six programmes	£15,643		£112..13		7,923		18.17		£6,172,491		757,531	249	£449,706
14	*Other six PBCs 2005/6*	*£12,743*		*£99.44*		*7,923*		*16.26*		*£6,115,621*		*751,009*	*337*	*£295,074*
15	All ten programmes	£32,911	0.687	£226.18		133,213		971.30		£232,861		5,093,090	10,826	£20,893
16	*All ten programmes 2005/6*	*£30,368*	*0.792*	*£240.67*		*133,213*		*926.22*		*£259,838*		*5,267,962*	*11,322*	*£21,256*
	Assume zero health gain in the other 13 programmes													
18	Other 13 programmes	£34,985	1.294	£452.78				0.00					0	
19	*Other 13 PBCs 2005/6*	*£33,942*	*1.186*	*£402.43*				*0.00*					*0*	
20	All 23 programmes	£67,896		£678.96				971.30		£699,024			10,826	£62,718
21	*All 23 programmes 2005/6*	*£64,310*		*£643.10*				*926.22*		*£694,330*			*11,322*	*£56,799*
	Note:	2006/7	2005/6											
22	All 23 programme spend	£67,896	£64,310											
23	% change in budget	1.00	1.00											
24	proportionate change	0.01	0.01											
25	Change in budget	£678.96	£643.10											

Note that the annual mortality figures reported in cells G5 & G6 and G13 & G14 are identical because we do not have mortality data for 2002/04.

Note that, for 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

Table B8.18: table showing cost of life and life year estimates using spend data for 2006/7 (two MFFs) and outcome data for 2004/06 (assumes some gain in other 13 programmes)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
				=0.01*C*D				=0.01*D*G*H		=E/I			=0.01*D*H*M/3	=E/N
	PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)		Average annual mortality, <75years, 2004/06	Outcome elasticity	Change in annual mortality		Cost per life gained (£)		Total life years lost, <75years, 2004/06	Change in annual life years lost	Cost per life year gained (£)
1	Cancer	£4,122	0.548	£22.59		62,259	0.337	114.98		£196,461		2,221,530	1,368	£16,518
2	Circulatory problems	£6,161	0.701	£43.19		45,504	1.447	461.57		£93,569		1,463,912	4,950	£8,725
3	Respiratory problems	£3,285	0.718	£23.59		11,601	3.507	292.12		£80,743		321,264	2,697	£8,747
4	Gastro-intest problems	£3,700	0.667	£24.68		5,926	2.137	84.47		£292,170		328,853	1,562	£15,795
5	Big four programmes	£17,268		£114.04		125,290		953.13		£119,650		4,335,559	10,576	£10,783
6	*Big four PBCs 2005/6*	*£17,625*		*£141.22*		*125,290*		*909.96*		*£155,196*		*4,516,953*	*10,986*	*£12,855*
7	Infectious diseases	£1,053	0.731	£7.70		2,050	0.03	0.45		£17,121,951		101,604	7	£1,036,377
8	Endocrine problems	£1,852	0.966	£17.89		1,690	0.812	13.26		£1,349,579		60,615	158	£112,882
9	Neurological problems	£2,790	0.648	£18.08		729	0.098	0.46		£39,052,658		68,808	15	£1,241,253
10	Genito-urinary problems	£3,482	0.837	£29.14		294	0.073	0.18		£162,240,239		11,554	2	£12,384,965
11	Trauma & injuries	£2,892	0.617	£17.84		1,037	0.527	3.37		£5,291,867		30,000	33	£548,767
12	Maternity* & neonates	£3,574	0.601	£21.48		2,123	0.035	0.45		£48,099,051		484,950	34	£631,700
13	Other six programmes	£15,643		£112..13		7,923		18.17		£6,172,491		757,531	249	£449,706
14	*Other six PBCs 2005/6*	*£12,743*		*£99.44*		*7,923*		*16.26*		*£6,115,621*		*751,009*	*337*	*£295,074*
15	All ten programmes	£32,911	0.687	£226.18		133,213		971.30		£232,861		5,093,090	10,826	£20,893
16	*All ten programmes 2005/6*	*£30,368*	*0.792*	*£240.67*		*133,213*		*926.22*		*£259,838*		*5,267,962*	*11,322*	*£21,256*
	Assume zero health gain in PBC23, and gain in ten PBCs applies to other 12 PBCs													
17	PBC23	£10,585	0.759	£80.34				0.00					0.00	
18	*PBC23 2005/6*	*£8,449*	*0.926*	*£78.24*				*0.00*					*0.00*	
19	Other 12 programmes	£24,400		£372.44				1,599.42		£232,861			17,826	£20,893
20	*Other 12 PBCs 2005/6*	*£25,493*		*£324.20*				*1,247.69*		*£259,838*			*15,252*	*£21,256*
21	All 23 programmes	£67,896		£678.96				2,570.72		£264,113			28,652	£23,697
22	*All 23 programmes 2005/6*	*£64,310*		*£643.10*				*2,173.90*		*£295,827*			*26,575*	*£24,200*
	Note:	2006/7	2005/6											
23	All23 programme spend	£67,896	£64,310											
24	% change in budget	1.00	1.00											
25	proportionate change	0.01	0.01											
26	Change in budget	£678.96	£643.10											

Note that the annual mortality figures reported in cells G5 & G6 and G13 & G14 are identical because we do not have mortality data for 2002/04.

Note that, for 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

B8.5 Model estimation using 2006/7 expenditure data and mortality data for 2006/2008: CARAN need and two MFFs

One shortcoming with the models presented above is that they relate expenditure in 2006/7 to mortality in the same period and in the two previous periods (i.e., in 2004, 2005 and 2006). The difficulty with this is that one would expect expenditure in year *t* to affect mortality in year *t* and possibly *subsequent* years (*t+1*, *t+2*, etc) but not mortality in *previous* years (*t-1*, *t-2*, etc). However, if we assume that PCTs have reached some sort of equilibrium in the expenditure choices they make and the outcomes they secure, so that expenditure levels change relatively little from one year to the next, then mortality over the three year period *t*, *t-1* and *t-2* might be a good proxy for mortality in *t*, *t+1* and *t+2*. Indeed, this is probably not an unreasonable assumption given the relatively slow pace at which both types of variable change.

Although this assumption of equilibrium is not an unreasonable one, it is one that ideally we would like to be able to drop. Fortunately, with the recent availability of more up-to-date mortality data, we have the opportunity to relate expenditure in 2006 to mortality in the same year and in the two following years (i.e., in 2006, 2007 and 2008).[29] Thus the models reported in Table B8.16 were re-estimated replacing the mortality rate for 2004/5/6 with that for 2006/7/8. The outcome and expenditure results for the big four programmes are shown in Table B8.19 with the relevant outcome and expenditure elasticities again highlighted (the first-stage regressions associated with these results can be found in Table BA.6 in the annex). These elasticities are similar to those presented previously in Table B8.16 but there are some changes (e.g, the outcome elasticity in the respiratory outcome equation falls from -2.839 to -2.029).

In all four outcome models expenditure has a significant negative effect on mortality and the all service measure of need has a significant positive effect. The all service measure of need squared is also positive and significant in the cancer outcome equation. In the respiratory outcome model, there is an additional indicator of need – the proportion of the population that are permanently sick – and this is both positive and statistically significant. The diagnostic statistics suggest that, in all four cases, own programme expenditure is endogenous and that the instruments are valid. They also suggest that the instruments are relevant. There is no evidence that the instruments are weak in three of the four outcome results. The Pesaran-Taylor test suggests that there is no evidence of model mis-specification

However, the Kleibergen-Paap F statistic for the respiratory disease outcome model is 7.022 and this is less than the 'critical' target of 10.0. This indicates that the instruments may be weak. However, if we re-estimate this model having dropped the least significant instrument, the coefficient on own programme expenditure is now -2.622 and this is significant at the 1% level. Moreover, there is now no evidence of weak instruments (the Kleibergen-Paap F statistic is 11.025) and it is this coefficient that we use for the respiratory outcome model in the cost of a life year calculations below.

In three of the four expenditure models both the need and budget variables have a positive and significant effect on own programme expenditure. In addition, the proxy for need in other programmes is negative and significant in all four cases. The diagnostic statistics suggest that, for all four expenditure models, expenditure is endogenous and the instruments are valid. They also suggest that the instruments are relevant and there is no evidence that the instruments are weak. The Pesaran-Taylor test suggests that there is no evidence of model mis-specification.

The outcome and expenditure elasticities presented in Table B8.19 can be used to calculate the cost of a life year in each programme. These calculations -- for both the big four programmes as well as for the other six programmes with mortality based outcome indicator -- are shown in Table B8.20. They show that the use of a more appropriate measure of mortality (i.e., for 2006/2007/2008 rather than for 2004/2005/2006) slightly increases the cost of a life year for the big four PBCs (from £10,783 to £12,333) as well as for all ten programmes with a mortality outcome measure (from £20,893 to £23,780).

[29] Clearly, some expenditure in year *t* will have an effect on mortality beyond *t+2* but we have no mortality data that would allow us to include this in our modelling work. We must assume that, for expenditure that affects mortality beyond *t+2,* PCTs have reached some sort of equilibrium position in terms of their expenditure choices and the outcomes secured.

In addition, Table B8.21 shows that if we assume that PBC23 generates a zero health gain and that the gain attributable to the remaining 12 programmes is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes is now £26,876 (it was £23,697 using mortality for 2004/5/6).

Table B8.19: table showing outcome and expenditure models for the big four programmes using spend data for 2006/7 (two MFFs) and mortality data for 2006/7/8

	(1) PBC 2 cancer outcome model	(2) PBC 2 cancer spend model	(3) PBC 10 circulation outcome model	(4) PBC 10 circulation spend model	(5) PBC 11 respiratory outcome model	(6) PBC 11 respiratory spend model	(7) PBC 13 gastro outcome model	(8) PBC 13 gastro spend model
own programme spend per head	-0.342***		-1.434***		-2.029***		-1.536***	
	[0.099]		[0.218]		[0.636]		[0.468]	
need CARAN per head	0.995***	1.626***	2.860***	2.306***	2.696***	1.449***	4.160***	2.040***
	[0.106]	[0.343]	[0.252]	[0.372]	[1.044]	[0.331]	[0.577]	[0.378]
need CARAN per head squared	1.163***				2.451			
	[0.348]				[1.561]			
SYLLR all deaths exclude cancer		-0.855***						
		[0.191]						
PCT budget per head		0.465		0.540*		0.679***		0.446*
		[0.300]		[0.299]		[0.251]		[0.263]
SYLLR all deaths exc circulatory				-1.666***				
				[0.295]				
permanently sick					0.759**			
					[0.367]			
SYLLR all deaths exc respiratory						-0.672**		
						[0.305]		
SYLLR all deaths exclude gastro								-1.206***
								[0.314]
lone pensioner households								
Constant	6.501***	5.913***	11.413***	10.696***	13.756***	3.346	9.719***	8.370***
	[0.436]	[2.815]	[1.046]	[2.379]	[3.279]	[2.075]	[2.009]	[2.299]
Endogeneity test statistic	13.695	19.421	42.548	24.461	17.687	8.439	16.373	15.211
Endogeneity p-value	0.000215	1.05e-05	6.90e-11	7.58e-07	2.60e-05	0.00367	5.20e-05	9.61e-05
Hansen-Sargan test statistic	0.685	0.021	0.949	1.262	1.462	0.302	2.761	0.0164
Hansen-Sargan p-value	0.408	0.084	0.814	0.261	0.227	0.583	0.0966	0.0898
Shea's partial R-squared	0.164	0.445	0.300	0.296	0.0785	0.327	0.140	0.356
Kleibergen-Paap LM test statistic	17.85	41.88	32.37	32.02	10.02	34.98	14.86	35.72
Kleibergen-Paap p-value	0.000133	8.04e-10	1.61e-06	1.11e-07	0.00666	2.54e-08	0.000592	1.75e-08
Kleibergen-Paap F statistic	13.28	56.69	17.14	31.84	7.022	20.94	11.63	22.40
Pesaran-Taylor reset statistic	0.00537	0.18	0.136	0.00349	0.0120	1.497	1.669	0.007
Pesaran-Taylor p-value	0.942	0.668	0.712	0.953	0.913	0.221	0.196	0.935

Note: robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Table B8.20: table showing cost of life and life year estimates using spend data for 2006/7 and outcome data for 2006/7/8 (assumes zero health gain for 13 programmes)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
				=0.01*C*D				=0.01*D*G*H		=E/I			=0.01*D*H*M/3	=E/N
	PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2006/08	Outcome elasticity (without negative sign)	Change in annual mortality		Cost per life gained (£)		Total life years lost, <75years, 2006/08	Change in annual life years lost	Cost per life year gained (£)
1	Cancer	£ 4,122	0.465	£19.17		61,961	0.342	98.54		£194,520		2,207,021	1,170	£16,383
2	Circulatory problems	£6,161	0.540	£33.27		41,106	1.434	318.31		£104,519		1,361,634	3,515	£9,466
3	Respiratory problems	£3,285	0.679	£22.31		11,574	2.622	206.06		£108,248		324,223	1,924	£11,593
4	Gastro-intestinal problems	£3,700	0.446	£16.50		6,160	1.536	42.20		£391,048		345,908	790	£20,892
	Big four programmes summary:													
5	Spend 2006 & mortality 2006/8	£17,268		£91.24		120,801		665.10		£137.188		4,238,786	7,399	£12,333
6	*Spend 2006 & mortality 2004/6*	*£17,268*		*£114.04*		*125,290*		*953.13*		*£119,650*		*4,335,559*	*10,576*	*£10,783*
7	*Spend 2005 & mortality 2002/4*	*£17,625*		*£141.22*		*125,290*		*909.96*		*£155,196*		*4,516,953*	*10,986*	*£12,855*
8	Infectious diseases	£1,053	0.792	£8.34		2,050	0.047	0.76		£10,928,905		106,552	13	£630,798
9	Endocrine problems	£1,852	0.953	£17.65		1,542	0.842	12.37		£1,426,410		57,672	154	£114,416
10	Neurological problems	£2,790	0.616	£17.19		727	0.112	0.50		£34,265,082		66,137	15	£1,129,960
11	Genito-urinary problems	£3,482	0.912	£31.76		294	0.051	0.14		£232,226,224		10,030	2	£20,421,090
12	Trauma & injuries*	£2,892	0.358	£10.35		1,037	0	0.00		#DIV/0!		30,000	0	#DIV/0!
13	Maternity & neonates*	£3,574	0.224	£8.01		2,189	0.482	2.36		£3,387,363		492,600	177	£45,158
	Other six programmes summary:													
14	Spend 2006 & mortality 2006/8	£15,643		£93.29		7,839		16.14		£5,780,723		762,991	362	£258,046
15	*Spend 2006 & mortality 2004/6*	*£15,643*		*£112.13*		*7,923*		*18.17*		*£6,172,491*		*757,531*	*249*	*£449,706*
16	*Spend 2005 & mortality 2002/4*	*£12,743*		*£99.44*		*7,923*		*16.26*		*£6,115,621*		*751,009*	*337*	*£295,074*
	All ten programmes summary:													
17	Spend 2006 & mortality 2006/8	£32,911	0.561	£184.53		128,640		681.24		£270,881		5,001,777	7,760	£23,780
18	*Spend 2006 & mortality 2004/6*	*£32,911*	*0.687*	*£226.18*		*133,213*		*971.30*		*£232,861*		*5,093,090*	*10,826*	*£20,893*
19	*Spend 2005 & mortality 2002/4*	*£30,368*	*0.792*	*£240.67*		*133,213*		*926.22*		*£259,838*		*5,267,962*	*11,322*	*£21,256*
	Assume zero health gain in the other 13 programmes													
	Other 13 programmes summary:													
20	Spend 2006 & mortality 2006/8	*£34,985*	1.413	£494.43				0.00					0	
21	*Spend 2006 & mortality 2004/6*	*£34,985*	*1.294*	*£452.78*				*0.00*					*0*	
22	*Spend 2005 & mortality 2002/4*	*£33,942*	*1.186*	*£402.43*				*0.00*					*0*	
	All 23 programmes													
23	Spend 2006 & mortality 2006/8	£67,896		£678.96				681.24		£996,655			7,760	£87,494
24	*Spend 2006 & mortality 2004/6*	*£67,896*		*£678.96*				*971.30*		*£699,024*			*10,826*	*£62,718*
25	*Spend 2005 & mortality 2002/4*	*£64,310*		*£643.10*				*926.22*		*£694,330*			*11,322*	*£56,799*
	Note:	2006/7	2005/6											
26	All 23 programme spend	£67,896	£64,310		Note that the annual mortality figures reported in cells G6 & G7 and G15 & G16 are identical because we do not have mortality data for 2002/04.									
27	% change in budget	1.00	1.00		Note that we have been unable to obtain a satisfactory outcome model for trauma & injuries and have assumed a zero outcome elasticity.									
28	proportionate change	0.01	0.01		Note that, for expenditure in 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.									
29	Change in budget	£678.96	£643.10											

Table B8.21: table showing cost of life and life year estimates using spend data for 2006/7 and outcome data for 2006/7/8 (assumes average health gain for 12 other programmes)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
				=0.01*C*D				=0.01*D*G*H		=E/I			=0.01*D*H*M/3	=E/N
	PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2006/08	Outcome elasticity (without negative sign)	Change in annual mortality		Cost per life gained (£)		Total life years lost, <75years, 2006/08	Change in annual life years lost	Cost per life year gained (£)
1	Cancer	£4,122	0.465	£19.17		61,961	0.342	98.54		£194,520		2,207,021	1,170	£16,383
2	Circulatory problems	£6,161	0.540	£33.27		41,106	1.434	318.31		£104,519		1,361,634	3,515	£9,466
3	Respiratory problems	£3,285	0.679	£22.31		11,574	2.622	206.06		£108,248		324,223	1,924	£11,593
4	Gastro-intestinal problems	£3,700	0.446	£16.50		6,160	1.536	42.20		£391,048		345,908	790	£20,892
	Big four programmes summary:													
5	Spend 2006 & mortality 2006/8	£17,268		£91.24		120,801		665.10		£137.188		4,238,786	7,399	£12,333
6	*Spend 2006 & mortality 2004/6*	*£17,268*		*£114.04*		*125,290*		*953.13*		*£119,650*		*4,335,559*	*10,576*	*£10,783*
7	*Spend 2005 & mortality 2002/4*	*£17,625*		*£141.22*		*125,290*		*909.96*		*£155,196*		*4,516,953*	*10,986*	*£12,855*
8	Infectious diseases	£1,053	0.792	£8.34		2,050	0.047	0.76		£10,928,905		106,552	13	£630,798
9	Endocrine problems	£1,852	0.953	£17.65		1,542	0.842	12.37		£1,426,410		57,672	154	£114,416
10	Neurological problems	£2,790	0.616	£17.19		727	0.112	0.50		£34,265,082		66,137	15	£1,129,960
11	Genito-urinary problems	£3,482	0.912	£31.76		294	0.051	0.14		£232,226,224		10,030	2	£20,421,090
12	Trauma & injuries*	£2,892	0.358	£10.35		1,037	0	0.00		#DIV/0!		30,000	0	#DIV/0!
13	Maternity & neonates*	£3,574	0.224	£8.01		2,189	0.482	2.36		£3,387,363		492,600	177	£45,158
	Other six programmes summary:													
14	Spend 2006 & mortality 2006/8	£15,643		£93.29		7,839		16.14		£5,780,723		762,991	362	£258,046
15	*Spend 2006 & mortality 2004/6*	*£15,643*		*£112.13*		*7,923*		*18.17*		*£6,172,491*		*757,531*	*249*	*£449,706*
16	*Spend 2005 & mortality 2002/4*	*£12,743*		*£99.44*		*7,923*		*16.26*		*£6,115,621*		*751,009*	*337*	*£295,074*
	All ten programmes summary:													
17	Spend 2006 & mortality 2006/8	£32,911	0.561	£184.53		128,640		681.24		£270,881		5,001,777	7,760	£23,780
18	*Spend 2006 & mortality 2004/6*	*£32,911*	*0.687*	*£226.18*		*133,213*		*971.30*		*£232,861*		*5,093,090*	*10,826*	*£20,893*
19	*Spend 2005 & mortality 2002/4*	*£30,368*	*0.792*	*£240.67*		*133,213*		*926.22*		*£259,838*		*5,267,962*	*11,322*	*£21,256*
	Other 13 PBCs? Assume zero health gain in PBC23…													
20	PBC23: spend 2006 & mortality 2006/8	£10,585	0.739	£78.22				0.00					0.00	
21	PBC23: spend 2006 & mortality 2004/6	£10,585	0.759	£80.34				0.00					0.00	
22	PBC23: spend 2005 & mortality 2002/4	*£8,449*	*0.926*	*£78.24*				*0.00*					*0.00*	
	...and that the gain in ten PBCs (see above) applies to the remaining 12 PBCs													
23	12 PBCs: spend 2006 & mortality 2006/8	£24,400		£416.20				1,536.48		£270,881			17,502	£23,780
24	*12 PBCs: spend 2006 & mortality 2004/6*	*£24,400*		*£372.44*				*1,599.42*		*£232,861*			*17,826*	*£20,893*
25	*12 PBCs: spend 2005 & mortality 2002/4*	*£25,493*		*£324.20*				*1,247.69*		*£259,838*			*15,252*	*£21,256*
	All 23 programmes													
26	23 PBCs: spend 2006 & mortality 2006/8	£67,896		£678.96				2,217.72		£306,153			25,262	£26,876
27	*All 23 PBCs: spend 2006 & mortality 2004/6*	*£67,896*		*£678.96*				*2,570.72*		*£264,113*			*28,652*	*£23,697*
28	*All 23 PBCs: spend 2005 & mortality 2002/4*	*£64,310*		*£643.10*				*2,173.90*		*£295,827*			*26,575*	*£24,200*
	Note:	2006/7	2005/6											
29	All 23 programme spend	£67,896	£64,310											
30	% change in budget	1.00	1.00											
31	proportionate change	0.01	0.01											
32	Change in budget	£678.96	£643.10											

Note the annual mortality figures reported in cells G5 & G6 and G13 & G14 are identical because we do not have mortality data for 2002/4.
Note that, for 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

B8.6 Adjusting the cost of life (year) estimates for the mismatch in the ICD10 coverage of the expenditure and the mortality data

The cost of a life (year) estimates presented in Tables B8.20 and B8.21 assume a 1% increase in each PCT's budget and are calculated as:

the cost of an additional life in a particular programme
= the change in expenditure in that programme / the change in mortality in that programme
= (annual spend * expenditure elasticity) / (annual mortality * outcome elasticity * expenditure elasticity)

and

the cost of an additional life year in a particular programme
= the change in expenditure in that programme / the change in life years lost in that programme
= (annual spend * expenditure elasticity) / (annual mortality * outcome elasticity * expenditure elasticity)

Thus an integral part of the calculation of the cost of a life (year) is the annual mortality (life years lost) figure associated with a particular programme. Ideally, the ICD10 coverage of the expenditure data should coincide with that of the mortality data but, as know from Table B5.1, the ICD10 coverage of the mortality data typically falls short of that for the expenditure data. Unless we adjust the annual mortality figure so that its ICD10 coverage approximates that of the expenditure data, our cost of life (year) estimates will usually be too large because they will usually underestimate the mortality gain.

Table B8.22 reproduces Table B8.20 but incorporates this ICD 10 coverage adjustment (see columns L and R in Table B8.22). This adjustment reduces the cost of a life year:

- for the big four programmes from £12,333 to £10,604
- for the ten programmes with a mortality based outcome indicator from £23,780 to £19,965
- for all programmes assuming a zero gain for the 13 PBCs without an outcome indicator from £87,494 to £73,457.

Similarly, Table B8.23 reproduces Table B8.21 but incorporates this ICD 10 coverage adjustment (see columns L and R again). If we assume a zero health gain in PBC23 and an average gain in the other 12 PBCs without a mortality based outcome indicator, then this adjustment reduces the cost of a life year for all programmes from £26,876 to £22,565.

TableB8.22: table showing Cost of life and life year estimates using expenditure data for 2006 and outcome data for 2006/7/8 (assumes zero health gain for 13 programmes) adjusted for the ICD10 coverage of the expenditure and outcome data

A	B	C	D	E =0.01*C*D	F	G	H	I =0.01*D*G*H	J	K =E/I	L	M =I/L	N =E/M	O	P	Q =0.01*D*H*P/3	R	S =Q/R	T =E/Q	U =E/S
	PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2006/08	Outcome elasticity (without negative sign)	Change in annual mortality		Cost per life gained (£)	Coverage of mortality data relative to spend data	Change in annual mortality adj for coverage	Cost per life gained (£) adj for coverage		Total life years lost, <75years, 2006/08	Change in annual life years lost	Coverage of mortality data relative to spend data	Change in annual life years lost adj for YLL	Cost per life year gained (£)	Cost per life year gained adj for YLL coverage (£)
1	Cancer	£4,122	0.465	£19.17		61,961	0.342	98.54		£194,520	0.984	100.14	£191,407		2,207,021	1,170	0.984	1,,189	£16,383	£16,121
2	Circulatory problems	£6,161	0.540	£33.27		41,106	1.434	318.31		£104,519	0.992	320.88	£103,683		1,361,634	3,515	0.992	3,543	£9,466	£9,390
3	Respiratory problems	£3,285	0.679	£22.31		11,574	2.622	206.06		£108,248	0.773	262.57	£83,676		324,223	1,924	0.773	2,489	£11,593	£8,961
4	Gastro-intestinal problems	£3,700	0.446	£16.50		6,160	1.536	42.20		£391,048	0.571	73.90	£223,288		345,908	790	0.571	1,383	£20,892	£11,929
	Big four programmes summary:																			
5	Spend 2006 & mortality 2006/8	£17,268		£91.24		120,801		665.10		£137,188		761.49	£119,823		4,238,786	7,399		8,604	£12,333	£10,604
6	Spend 2006 & mortality 2004/6	£17,268		£114.04		125,290		953.13		£119,650					4,335,559	10,576			£10,783	
7	Spend 2005 & mortality 2002/4	£17,625		£141.22		125,290		909.96		£155,196					4,516,953	10,986			£12,855	
8	Infectious diseases	£1,053	0.792	£8.34		2,050	0.047	0.76		£10,928,905	1.000	0.76	£10,928,905		106,552	13	1.000	13	£630,798	£630,798
9	Endocrine problems	£1,852	0.953	£17.65		1,542	0.842	12.37		£1,426,410	0.634	19.52	£904,344		57,672	154	0.634	243	£114,416	£72,539
10	Neurological problems	£2,790	0.616	£17.19		727	0.112	0.50		£34,265,082	0.136	3.69	£4,660,051		66,137	15	0.136	112	£1,129,960	£153,675
11	Genito-urinary problems	£3,482	0.912	£31.76		294	0.051	0.14		£232,226,224	0.172	0.80	£39,942,910		10,030	2	0.172	9	£20,421,090	£3,512,427
12	Trauma & injuries*	£2,892	0.358	£10.35		1,037	0	0.00		#DIV/0!	0.175	0.00	#DIV/0!		30,000	0	0.175	0	#DIV/0!	#DIV/0!
13	Maternity & neonates*	£3,574	0.224	£8.01		2,189	0.482	2.36		£3,387,363	8.213	0.29	£27,820,413		492,600	177	0.679	261	£45,158	£30,662
	Other six programmes summary:																			
14	Spend 2006 & mortality 2006/8	£15,643		£93.29		7,839		16.14		£5,780,723		25.05	£3,724,129		762,991	362		639	£258,046	£146,108
15	Spend 2006 & mortality 2004/6	£15,643		£112.13		7,923		18.17		£6,172,491					757,531	249			£449,706	
16	Spend 2005 & mortality 2002/4	£12,743		£99.44		7,923		16.26		£6,115,621					751,009	337			£295,074	
	All ten programmes summary:																			
17	Spend 2006 & mortality 2006/8	£32,911	0.561	£184.53		128,640		681.24		£270,881		786.54	£234,617		5,001,777	7,760		9,243	£23,780	£19,965
18	Spend 2006 & mortality 2004/6	£32,911	0.687	£226.18		133,213		971.30		£232,861					5,093,090	10,826			£20,893	
19	Spend 2005 & mortality 2002/4	£30,368	0.792	£240.67		133,213		926.22		£259,838					5,267,962	11,322			£21,256	
	Assume zero health gain in the other 13 programmes																			
	Other 13 programmes summary:																			
20	Spend 2006 & mortality 2006/8	£34,985	1.413	£494.43				0.00								0		0		
21	Spend 2006 & mortality 2004/6	£34,985	1.294	£452.78				0.00								0				
22	Spend 2005 & mortality 2002/4	£33,942	1.186	£402.43				0.00								0				
	All 23 programmes																			
23	Spend 2006 & mortality 2006/8	£67,896		£678.96				681.24		£996,655		786.54	£863,228			7,760		9,243	£87,494	£73,457
24	Spend 2006 & mortality 2004/6	£67,896		£678.96				971.30		£699,024						10,826			£62,718	
25	Spend 2005 & mortality 2002/4	£64,310		£643.10				926.22		£694,330						11,322			£56,799	
	Note:	2006/7	2005/6																	
26	All 23 programme spend	£67,896	£64,310			Note that the annual mortality figures reported in cells G6 & G7 and G15 & G16 are identical because we do not have mortality data for 2002/04.														
27	% change in budget	1.00	1.00			Note that we have been unable to obtain a satisfactory outcome model for trauma & injuries and have assumed a zero outcome elasticity.														
28	proportionate change	0.01	0.01			Note that, for expenditure in 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.														
29	Change in budget	£678.96	£643.10			Note that the adjustment for the coverage of the YLL data relative to the spend data uses deaths under age 75 in England in 2008.														

Table B8.23: table showing cost of life and life year estimates using expenditure data for 2006 and outcome data for 2006/7/8 (assumes average health gain for 12 other programmes) adjusted for the ICD10 coverage of the expenditure and outcome data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
				=0.01*C*D				=0.01*D *G*H		=E/I		=I/L	=E/M			=0.01*D*H* P/3		=Q/R	=E/Q	=E/S
	PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2006/08	Outcome elasticity (without negative sign)	Change in annual mortality		Cost per life gained (£)	Coverage of mortality data relative to spend data	Change in annual mortality adj for coverage	Cost per life gained (£) adj for coverage		Total life years lost, <75years, 2006/08	Change in annual life years lost	Coverage of mortality data relative to spend data	Change in annual life years lost adj for coverage	Cost per life year gained (£)	Cost per life year gained adj for YLL coverage (£)
1	Cancer	£4,122	0.465	£19.17		61,961	0.342	98.54		£194,520	0.984	100.14	£191,407		2,207,021	1,170	0.984	1,,189	£16,383	£16,121
2	Circulatory problems	£6,161	0.540	£33.27		41,106	1.434	318.31		£104,519	0.992	320.88	£103,683		1,361,634	3,515	0.992	3,543	£9,466	£9,390
3	Respiratory problems	£3,285	0.679	£22.31		11,574	2.622	206.06		£108,248	0.773	262.57	£83,676		324,223	1,924	0.773	2,489	£11,593	£8,961
4	Gastro-intestinal problems	£3,700	0.446	£16.50		6,160	1.536	42.20		£391,048	0.571	73.90	£223,288		345,908	790	0.571	1,383	£20,892	£11,929
	Big four programmes summary:																			
5	Spend 2006 & mortality 2006/8	£17,268		£91.24		120,801		665.10		£137,188		761.49	£119,823		4,238,786	7,399		8,604	£12,333	£10,604
6	Spend 2006 & mortality 2004/6	£17,268		£114.04		125,290		953.13		£119,650					4,335,559	10,576			£10,783	
7	Spend 2005 & mortality 2002/4	£17,625		£141.22		125,290		909.96		£155,196					4,516,953	10,986			£12,855	
8	Infectious diseases	£1,053	0.792	£8.34		2,050	0.047	0.76		£10,928,905	1.000	0.76	£10,928,905		106,552	13	1.000	13	£630,798	£630,798
9	Endocrine problems	£1,852	0.953	£17.65		1,542	0.842	12.37		£1,426,410	0.634	19.52	£904,344		57,672	154	0.634	243	£114,416	£72,539
10	Neurological problems	£2,790	0.616	£17.19		727	0.112	0.50		£34,265,082	0.136	3.69	£4,660,051		66,137	15	0.136	112	£1,129,960	£153,675
11	Genito-urinary problems	£3,482	0.912	£31.76		294	0.051	0.14		£232,226,224	0.172	0.80	£39,942,910		10,030	2	0.172	9	£20,421,090	£3,512,427
12	Trauma & injuries*	£2,892	0.358	£10.35		1,037	0	0.00		#DIV/0!	0.175	0.00	#DIV/0!		30,000	0	0.175	0	#DIV/0!	#DIV/0!
13	Maternity & neonates*	£3,574	0.224	£8.01		2,189	0.482	2.36		£3,387,363	8.213	0.29	£27,820,413		492,600	177	0.679	261	£45,158	£30,662
	Other six programmes summary:																			
14	Spend 2006 & mortality 2006/8	£15,643		£93.29		7,839		16.14		£5,780,723		25.05	£3,724,129		762,991	362		639	£258,046	£146,108
15	Spend 2006 & mortality 2004/6	£15,643		£112.13		7,923		18.17		£6,172,491					757,531	249			£449,706	
16	Spend 2005 & mortality 2002/4	£12,743		£99.44		7,923		16.26		£6,115,621					751,009	337			£295,074	
	All ten programmes summary:																			
17	Spend 2006 & mortality 2006/8	£32,911	0.561	£184.53		128,640		681.24		£270,881		786.54	£234,617		5,001,777	7,760		9,243	£23,780	£19,965
18	Spend 2006 & mortality 2004/6	£32,911	0.687	£226.18		133,213		971.30		£232,861					5,093,090	10,826			£20,893	
19	Spend 2005 & mortality 2002/4	£30,368	0.792	£240.67		133,213		926.22		£259,838					5,267,962	11,322			£21,256	
	Other 13 PBCs? Assume zero health gain in PBC23…																			
20	PBC23: spend 2006 & mortality 2006/8	£10,585	0.739	£78.22				0.00								0.00		0.00		
21	PBC23: spend 2006 & mortality 2004/6	£10,585	0.759	£80.34				0.00								0.00				
22	PBC23: spend 2005 & mortality 2002/4	£8,449	0.926	£78.24				0.00								0.00				
	...and that the gain in ten PBCs (see row 17) applies to the remaining 12 PBCs																			
23	12 PBCs: spend 2006 & mortality 2006/8	£24,400		£416.20				1,536.48		£270,881		1,773.97	£234,617			17,502		20,847	£23,780	£19,965
24	12 PBCs: spend 2006 & mortality 2004/6	£24,400		£372.44				1,599.42		£232,861						17,826			£20,893	
25	12 PBCs: spend 2005 & mortality 2002/4	£25,493		£324.20				1,247.69		£259,838						15,252			£21,256	
	All 23 programmes																			
26	23 PBCs: spend 2006 & mortality 2006/8	£67,896		£678.96				2,217.72		£306,153		2,560.50	£265,167			25,262		30,090	£26,876	£22,565
27	23 PBCs: spend 2006 & mortality 2004/6	£67,896		£678.96				2,570.72		£264,113						28,652			£23,697	
28	23 PBCs: spend 2005 & mortality 2002/4	£64,310		£643.10				2,173.90		£295,827						26,575			£24,200	

	Note:	2006/7	2005/6
29	All 23 programme spend	£67,896	£64,310
30	% change in budget	1.00	1.00
31	proportionate change	0.01	0.01
32	Change in budget	£678.96	£643.10

Note that the annual mortality figures reported in cells G6 & G7 and G15 & G16 are identical because we do not have mortality data for 2002/04 yet.

Note that we have been unable to obtain a satisfactory outcome model for trauma & injuries and have assumed a zero outcome elasticity.

Note that, for expenditure in 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

Note that the adjustment for the coverage of the YLL data relative to the spend data uses deaths under age 75 in England in 2008.

B8.7 Adjusting the cost of life (year) estimates for Department of Health funded expenditure that is not undertaken by PCTs

PCT expenditure accounts for a large proportion of Department of Health expenditure but PCTs do not account for all of the Department's budget. In 2006/7 the Department of Health's gross expenditure totalled £83.5bn. Charges raised £3.4bn so net expenditure totalled £80.1bn. Of this net expenditure, PCTs accounted for £67.3bn (that is, 84%) and various other bodies accounted for the remaining £12.8bn. A breakdown of this gross and net expenditure by major body is shown in Table B8.24.

Table B8.24: table showing department of Health funded expenditure by major bodies, 2006/7

Body	Gross spend £ billion	Income £ billion	Net spend £ billion
PCTs	69.8	2.5	67.3
Strategic Health Authorities	3.8	0.0	3.8
Special Health Authorities*	2.8	1.3	1.5
Department of Health Own Costs (eg PSS grants, grants to LAs)	7.1	-0.4	7.5
Total Department of Health	83.5	3.4	80.1

*This includes, for example, NICE, the NHS Business Services Authority, the Information Centre, the NHS Litigation Authority, and the National Patient Safety Agency.

The Department of Health has allocated net non-PCT expenditure across the 23 PBCs and the impact of this allocation on total spend by PBC is shown in Table B8.25 below. No geographic breakdown (e.g., by PCT) of this expenditure is available.

Of the additional £12bn of net expenditure, £11.2bn (93%) has been allocated to PBC23. This largely reflects: (a) the allocation of almost all Strategic Health Authority expenditure to either PBC23B ('other: SHAs including workforce development committees') or PBC23X ('other: miscellaneous'), and (b) the allocation of almost two-thirds of Department of Health expenditure to PBC23X ('other: miscellaneous').

The remaining £0.8bn of additional net expenditure is spread across all PBCs according to various allocation rules. For example, the majority of expenditure on Special Health Authorities is apportioned across programme categories on the basis of the PCT and SHA expenditure breakdown. The exception is the NHS Business Services Authority expenditure which is apportioned on the basis of Primary Care prescribing expenditure splits. Although this approach avoids allocating expenditure to the 'Other: Miscellaneous' category, this allocation of expenditure does not necessarily reflect actual expenditure. For example, NHS Litigation Authority expenditure may not be incurred in the same areas as overall PCT expenditure [38].

Table B8.25: table showing net PCT and other Department of Health funded expenditure by PBC, 2006/7

PBC #	Programme budget category	Net spend, £bn, 2006/7 all PCTs	others	all DH	Others' spend as % of PCT spend
		(a)	(b)	c)	(d)
01	Infectious Diseases	1.1	0.1	1.2	13.3%
02	Cancers & Tumours	4.1	0.0	4.2	0.8%
03	Disorders of Blood	0.8	0.1	0.9	12.0%
04	Endocrine, Nutritional and Metabolic Issues	1.9	0.1	2.0	5.4%
05	Mental Health Disorders	8.4	0.3	8.7	3.2%
06	Problems of Learning Disability	2.4	-0.1	2.4	-2.5%
07	Neurological	2.8	0.0	2.8	1.3%
08	Problems of Vision	1.4	0.0	1.3	-1.8%
09	Problems of Hearing	0.3	0.0	0.3	0.5%
10	Problems of Circulation	6.2	0.2	6.4	4.0%
11	Problems of the Respiratory System	3.3	0.1	3.3	1.6%
12	Dental Problems	2.6	-0.2	2.4	-7.0%
13	Problems of the Gastro Intestinal System	3.7	0.0	3.7	-0.5%
14	Problems of the Skin	1.4	0.0	1.5	2.2%
15	Problems of the Musculoskeletal System	3.4	0.0	3.4	-0.3%
16	Problems due to Trauma and Injuries	2.9	0.0	2.9	-0.5%
17	Problems of the Genito Urinary System	3.5	0.1	3.6	2.1%
18	Maternity and Reproductive Health	2.9	-0.1	2.8	-2.2%
19	Conditions of Neonates	0.7	0.1	0.7	10.3%
20	Adverse Effects and Poisoning	0.7	0.0	0.7	-0.8%
21	Healthy Individuals	1.4	0.0	1.4	0.8%
22	Social Care Needs	1.5	0.0	1.5	0.9%
23	Other Areas of Spend/Conditions	10.6	11.2	21.8	106.1%
	All Categories	67.9	12.0	79.9	17.7%

Note: the figures in Tables B8.24 and B8.25 draw on various sources (e.g., Department of Health resource accounts and programme budgeting returns) and may (a) disagree slightly and (b) create some unusual results (e.g., the aggregate PCT figure for dental problems exceeds the all England level[38]).

It is clear that most of the non-PCT expenditure is not specific to any disease area and that, to avoid putting all of it into a residual category, the Department has identified what are reasonable but largely arbitrary rules to spread what is a relatively small proportion of this non-PCT expenditure across all PBCs.

The cost of a life (year) estimates presented above are based on the impact of a 1% exogenous change in total net PCT spend. All of our outcome and expenditure models have been estimated using net PCT expenditure, and all of our elasticities relate to this expenditure. Implicitly we assume that any budgetary shock only affects PCT funding and that it leaves non-PCT funding unchanged.

Suppose instead we assume a 1% exogenous change in the Departmental budget. How might this budgetary shock be split between PCT and non-PCT expenditure? There are two obvious options to consider. We could assume either: (a) that all of this change is applied to PCT budgets and that there is no change in the non-PCT budget (as we do implicitly at the moment) or (b) that the budgetary shock affects both PCT and non-PCT budgets.

If the non-PCT budget is wholly unresponsive to the exogenous shock then our cost of a life year estimates will be unchanged because this expenditure category attracts none of the budgetary change (although this expenditure will clearly contribute to a measure of average productivity).

If the non-PCT budget is to some degree responsive to the exogenous shock then it will affect our cost of a life (year) estimates. To calculate the size of this impact we would need to know:

(a) how responsive the non-PCT budget is to a total Departmental budgetary shock;
(b) how the responsive part of the non-PCT budget is allocated across PBCs;
and
(c) the size of the health effects associated with changes in the non-PCT budget at PBC level.

We have no evidence on how responsive the non-PCT budget is likely to be to a total budgetary shock. However, from Table B8.25 and the discussion about the rather arbitrary (but understandable) rules employed by the Department to allocate non-PCT expenditure to PBCs, it would seem reasonable to assume that any change in the non-PCT budget should all be allocated to PBC23. This 'solves' the problem of identifying the health gains associated with this change in the non-PCT budget because, in our cost of a life year calculations, we assume that expenditure in this category attracts no health gains.

Thus although we have no evidence on how responsive the non-PCT budget is likely to be to a total budgetary shock, we can present two scenarios. In the first scenario, the non-PCT budget is wholly unresponsive to a budgetary shock and any budgetary change is fully implemented via PCT expenditure. In this case, there is no impact on the cost of a life year.

In the second scenario, one might assume that the non-PCT budget is as responsive to Departmental budgetary changes as is the PCT budget. In this case a 1% change in the Departmental budget is translated into a 1% change in both the total PCT and total non-PCT budgets, and this will increase the cost of a life year by 17.7% for 2006/7, that is, from £22,565 to £26,553. This percentage increase is, of course, the same figure as total non-PCT expenditure expressed as a percentage of PCT expenditure. This is because all of the additional non-PCT expenditure is allocated to PBC23 and the assumption is that all expenditure in this category offers no health gain.

We have no information on how any Departmental budgetary shock is likely to be split between PCT and non-PCTs budgets. Our cost of a life year estimates implicitly assume that the non-PCT budget is wholly unresponsive to any budgetary shock. This is clearly a possibility. Alternatively, one might assume that the non-PCT budget is as responsive to a Departmental budgetary shock as is the PCT budget. If this was the case then it would add 17.7% to our cost of a life year estimate for 2006/7. However, in the absence of any information about the responsiveness of the non-PCT budget, it is difficult to come to any firm conclusion about the impact of non-PCT expenditure on our cost of a life year estimates. We therefore persevere with the assumption that the non-PCT budget is wholly unresponsive to Departmental budgetary shocks.

B8.8 Application of method to other non-mortality based outcome indicators

Not all health care expenditure will be directed towards the reduction of mortality but it is relatively easy to envisage how our methods might be applied to other, non-mortality based, outcome indicators. To illustrate how our approach might be applied to other such indicators we note that PROMs (health gain) data for various operations is available from the HES online website. For each PCT this data set reports the average health gain for those survey respondents who have had a specific operation (e.g., for hip replacement, for knee replacement, for varicose veins, and for groin hernia) over the survey period.

As a starting point, and to illustrate the principles involved, we focus on hip and knee replacements. As our outcome indicator for these procedures, we calculate

[(average health gain per hip operation*number of hip operations) +
(average health gain per knee operation*number of knee operations)]
/total PCT population

for each PCT (this ignores age standardisation). This health gain measure is broadly comparable with our usual mortality measure, which is a 'years of life lost' rate per 10,000 of population (again, ignoring age standardisation).

Ideally the expenditure, number of operations, and PROMs data should all relate to the same time period but here the PROMs data covers operations undertaken between April 2009 and October 2010 yet the expenditure and number of operations data relate to 2006/7. Implicitly we are assuming that the average gain per operation in 2006/7 is the same as over the PROMs survey period (although this is not particularly important as we are only illustrating principles here).

Unfortunately, the Department of Health does not report the number of patients undergoing an eligible operation by commissioner (PCT) so we use the HES dataset for 2006/7 to obtain this information. Eligible hip and knee operations are defined in Annex 1 of the 'Guide to PROMs methodology' (on the HES website) and we use these definitions (of eligible operation codes) to obtain a count of eligible hip and knee finished consultant episodes (FCEs) by PCT for 2006/7.

With data for both the average health gain per operation and the number of operations, we are now in a position to calculate 'the health gain per head of population' for hip and knee replacements as defined above. We can then use this as an outcome indicator for expenditure in the 'problems of the musculo-skeletal system' programme (i.e., PBC15) because the vast majority of hip and knee replacements are for osteoarthritis and this diagnosis is included in PBC15.

Table B8.26 reports the estimated outcome equation for PBC15 using the PROMs based outcome indicator. The result is intuitively plausible. More expenditure boosts the health gain but, for a given spend, more need reduces the gain. Of course we should remember that the health gain data relates to operations undertaken between April 2009 and October 2010 yet the expenditure and number of operations data (FCEs) relate to 2006/7. However, one might assume that the gain associated with each operation in 2009/10 is the same as the gain associated with each operation in 2006/7.

The diagnostic statistics suggest that expenditure is endogenous and that the instruments are valid. They also suggest that the instruments are relevant and there is no evidence that the instruments are weak. The Pesaran-Taylor test suggests that there is no evidence of model mis-specification.

This brief example illustrates the principles involved in extending our modelling approach beyond those programmes with a mortality indicator.

Table B8.26: table showing outcome model for the trauma and injuries programme, 2006/7

Regressors	Coefficient (standard error)
Expenditure per person in PBC 15	1.9068*** (0.4289)
Need CARAN per person	-1.6807*** (0.4533)
Constant	-6.3486*** (1.794)
Number of PCTs	143
Diagnostic test statistics	
Endogeneity test statistic	24.677
Endogeneity p-value	0.0000
Hansen-Sargan test statistic	1.136
Hansen-Sargan p-value	0.2865
Kleibergen-Paap LM test statistic	14.702
Kleibergen-Paap p-value	0.0006
Kleibergen-Paap Wald F statistic	10.367
Pesaran-Taylor reset statistic	0.03
Pesaran-Taylor p-value	0.8588

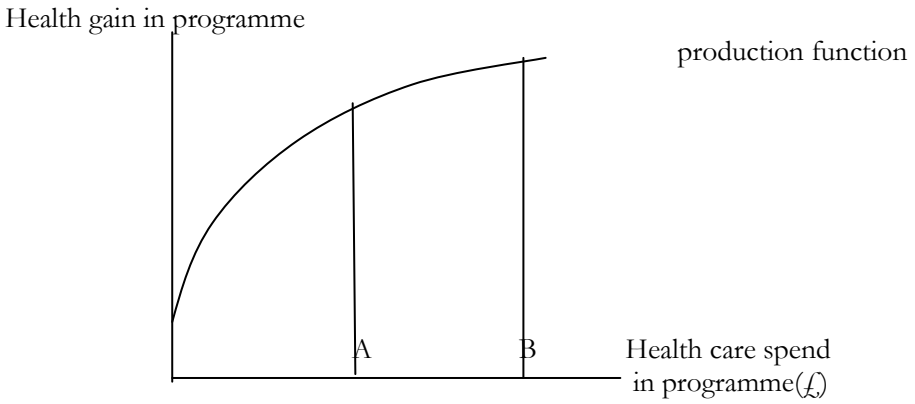
Notes: (i) the dependent variable is the health gain per head of population associated with eligible hip and knee operations undertaken during 2006/7; (ii) that there are only 143 observations and not the usual 152 because, for the other nine PCTs, there are fewer than 30 completed PROMs questionnaires on which to compute the average health gain and, as a result of such a low number of respondents, these PCTs have been dropped from the sample; (ii) the first-stage regression includes the IMD2007 (coefficient = -0.439, standard error=0.144) and the proportion of residents providing unpaid care (coefficient = 0.219, standard error=0.367); and (iii) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

B8.9 Comparing outcome models for 'high' spending and 'low' spending PCTs

As we have already noted, not all PCTs spend the same amount in each programme of care. Even after allowing for differences in local circumstances (such as input prices and need), some PCTs spend more than others, and it is this variation in expenditure that facilitates the estimation of our outcome and expenditure models.

Figure B8.2 illustrates the familiar health gain production function; as expenditure increases so too does health output but it increases at a diminishing rate. If all PCTs face the same production function (having controlled for input prices and need), and all PCTs are wholly efficient, then we would expect those PCTs that spend more (e.g., at point B) to experience a lower outcome elasticity than those that spend less (e.g., at point A) simply because they are further along the production function and are experiencing greater diminishing marginal returns.

Figure B8.2: graph showing Health gain production function



To test this hypothesis we used the expenditure model for each of the big four programmes to divide the 152 PCTs into two groups: those whose predicted spend is greater than the average predicted spend in that programme (ceteris paribus), and those whose predicted spend is smaller than the average predicted spend (ceteris paribus). We then re-estimated our outcome model for each of these two groups of PCTs and the results of this re-estimation are shown in Table B8.27. [30]

The first column in Table B8.27 presents the IV regression results for the outcome model for all PCTs; the second column reports the results for the 'high spend' PCTs; and the third column reports the results for the 'low spend' PCTs. For all four programmes, the coefficient on the expenditure variable is larger (in an absolute sense) for the 'high spend' PCTs than for the 'low spend' PCTs. This result contradicts our hypothesis that 'high spenders' will have a lower elasticity than 'low spenders'.

However, if we drop the assumption that all PCTs are equally efficient – so that that some lie within the frontier defined by the production function – then it is clearly possible for 'high' spending PCTs to experience a larger outcome elasticity than a 'low' spending one. And, of course, it is rather difficult to defend the assumption that all PCTs are equally efficient.

We can use the outcome elasticities reported in Table B8.27 to calculate the cost of a life year for 'high' and 'low' spenders in each of the big four programmes. These calculations are shown in Table B8.28.[31] As is to be anticipated, they reveal that the cost of a life year is much smaller in 'high' spend' PCTs than it is in 'low spend' PCTs. For example, the cost of a life year in the cancer programme is £16,383 across all PCTs but for 'high spenders' it is much less than this (£11,350) and for 'low spenders' it is much greater than this (£76,620). Presumably 'high spending' PCTs are high spenders because the cost of a life year is relatively low and additional health gains in a particular programme can be had relatively cheaply. Similarly, 'low spending' PCTs are low spenders because the cost of a life year is relatively high and additional health gains are relatively expensive.

[30] When re-estimating the all PCT model for 'high spenders' and then for 'low spenders' no attempt was made to adjust the estimating equation for any implied model mis-specifiaction.

[31] The cost of a life year estimates presented in Table 8.28 are not adjusted for the mismatch in the ICD10 coverage of the expenditure and mortality data because such an adjustment would not affect our conclusions.

Table B8.27: table showing re-estimating the 2006/7 outcome model for 'high' spending and 'low' spending PCTs

Cancer outcome equation	All PCTs	'High spend' PCTs	'Low spend' PCTs
Regressors			
1 Constant	6.500***	7.132***	5.352***
2 Need for health care	0.995***	1.265***	0.848***
3 Need for health care squared	1.162***	0.588	0.842
4 Cancer expenditure per person	-0.342***	-0.488***	-0.074
Number observations	152	76	76
Endogeneity test statistic	13.695***	7.165***	0.501
Instrument validity: Hansen J statistic	0.685	0.734	1.587
Instrument relevance: K-P LM statistic	17.847***	7.102**	13.617***
Weak instrument: K-P Wald F statistic	13.279	6.722	7.436
Re-set test	0.01	0.68	1.95

Circulatory disease outcome equation	All PCTs	'High spend' PCTs	'Low spend' PCTs
Regressors			
1 Constant	11.413***	11.254***	9.356***
2 Need for health care	2.859***	2.741***	2.636***
3 Circulatory expenditure per person	-1.434***	-1.403***	-0.995***
Number observations	152	76	76
Endogeneity test statistic	42.548***	9.424***	20.489***
Instrument validity: Hansen J statistic	0.949	4.782	0.366
Instrument relevance: K-P LM statistic	32.372***	12.658**	15.123***
Weak instrument: K-P Wald F statistic	17.143	6.275	12.421
Re-set test	0.14	0	1.29

Respiratory problems outcome equation	All PCTs	'High spend' PCTs	'Low spend' PCTs
Regressors			
1 Constant	17.023***	22.617**	11.695***
2 Need for health care	2.683**	2.512	3.095**
3 Need for health care squared	3.08	5.537	8.097***
4 Permanently sick	1.031**	1.401	0.844
5 Respiratory expenditure per person	-2.622***	-3.697*	-1.461*
Number observations	152	76	76
Endogeneity test statistic	20.860***	10.254***	5.380**
Instrument validity: Hansen J statistic	n/a	n/a	n/a
Instrument relevance: K-P LM statistic	9.091***	3.591	5.108**
Weak instrument: K-P Wald F statistic	11.025	4.568	6.227
Re-set test	0	0.08	0.21

Gastro-intestinal outcome equation	All PCTs	'High spend' PCTs	'Low spend' PCTs
Regressors			
1 Constant	9.718***	9.306***	6.675***
2 Need for health care	4.159***	5.156***	3.236***
3 Gastro-intestinal spend per person	-1.536***	-1.471***	-0.819
Number observations	152	76	76
Endogeneity test statistic	16.373***	7.781***	3.700*
Instrument validity: Hansen J statistic	2.761	1.529	3.824*
Instrument relevance: K-P LM statistic	14.865***	10.094***	7.956**
Weak instrument: K-P Wald F statistic	11.629	10.607	7.985
Re-set test	1.67	0.15	0.56

Table B8.28: table showing calculation of the cost of a life year for the big four programmes in 2006/7 by type of PCT: 'high spenders' and 'low spenders'

	A	B	C	D	E	F	G	H	I
	PBC description	Type of PCT	Spend (£m) FY2006/7	1% of spend (£m) FY2006/7	Outcome elasticity (without negative sign)	Total life years lost, <75years, 2006/08	Annual average life years lost (=F/3)	Change in annual life years lost associated with 1% increase in spend (=E*G)/100	Cost (£) per life year gained (=D/H) unadj for ICD10 coverage
	Split PCTs according to whether they are 'high spender' (n=76) or 'low spenders' (n=76)								
1	Cancers	All	4,122	41.22	0.342	2,207,021	735,674	2,516	16,383
2	Cancers	High spend	2,080	20.8	0.488	1,126,580	375,527	1,833	11,350
3	Cancers	Low spend	2,042	20.42	0.074	1,080,442	360,147	267	76,620
4	Circulatory problems	All	6,161	61.61	1.434	1,361,634	453,878	6,509	9,466
5	Circulatory problems	High spend	3,148	31.48	1.403	695,890	231,963	3,254	9,673
6	Circulatory problems	Low spend	3,012	30.12	0.995	665,744	221,915	2,208	13,641
7	Respiratory problems	All	3,285	32.85	2.622	324,223	108,074	2,834	11,593
8	Respiratory problems	High spend	1,645	16.45	3.697	174,639	58,213	2,152	7,644
9	Respiratory problems	Low spend	1,640	16.4	1.461	149,584	49,861	728	22,513
10	Gastro- problems	All	3,700	37	1.536	345,908	115,303	1,771	20,892
11	Gastro- problems	High spend	1,868	18.68	1.471	190,231	63,410	933	20,026
12	Gastro- problems	Low spend	1,832	18.32	0.819	155,676	51,892	425	43,106

Note: 'high spending' PCTs are those whose predicted spend per person is greater than the average predicted spend per person (ceteris paribus), and 'low spending' PCTs are those whose predicted spend per person is less than the average predicted spend per person (ceteris paribus).

B8.10 Comparing outcome models for over target and under target PCTs

The Department of Health has a well-developed resource allocation formula that determines the size of each PCTs 'target' budget given local conditions (such as population size and the need for health care). Every few years an improved resource allocation formula is developed and this generates a new 'target' budget for each PCT. The new target might be quite different from the old target and the immediate implementation of the new formula might lead to a large change in the budget for some PCTs. To avoid the difficulties that sudden large budgetary changes might bring, actual annual financial allocations are gradually moved towards the latest target budget. This means that in any year some PCTs receive an actual allocation which is greater than their target allocation, and that others receive an actual allocation which is less than their target allocation.

To examine whether being over or under the target allocation has any impact on the results, we split the 152 PCTs into two groups: those that received a budget over their target allocation in 2006/7, and those that received a budget under their target allocation in 2006/7. The outcome elasticities from the estimation of these models are shown in column E of Table B8.29, and these elasticities are used to calculate the cost of a life year for each of these two groups of PCTs for each of the big four programmes (see column I).

The results are consistent for each programme: PCTs whose budget is beyond their target allocation record a smaller outcome elasticity and a larger cost of a life year than PCTs whose budget is less than their target allocation. For example, in the cancer programme and across all PCTs the outcome elasticity is -0.342 and the cost of a life year is £16,383 (unadjusted for the ICD10 coverage of the mortality data). For PCTs with a budget that exceeds their target allocation, the outcome elasticity is smaller (-0.179) and the cost of a life year is larger (£32,365) than for all PCTs combined. However, for PCTs with a budget that falls short of their target allocation, the outcome elasticity is larger (-0.476) and the cost of a life year is smaller (£11,502) than for all PCTs combined.

One explanation for this result is that PCTs whose budget is beyond their target allocation are under less financial pressure than other PCTs, and that one consequence of this is that there is less pressure on them to behave in the most efficient manner possible. There is some evidence in the literature to support the hypothesis that the degree of PCT inefficiency is positively related to the amount by which a PCTs is over its target allocation. [39]

If we also re-estimate the expenditure models for both groups of PCTs we can calculate the cost of a life year for the big four programmes combined. The relevant expenditure elasticities are shown in column D of Table B8.30. These expenditure elasticities are far larger for under target PCTs than they are for over target PCTs. One reason for this might be that the big four programmes are priority ('hard') programmes. Over target PCTs are able to devote sufficient resources to the big four so that any additional budget is directed towards other ('softer') programmes which are less well funded than the priority programmes. In contrast, under target PCTs are struggling to devote sufficient resources to the priority programmes so that, when further funding does become available, this is directed towards the priority programmes.

These expenditure and outcome elasticities in Table B8.30 can be used to calculate the cost of a life year for the big four programmes combined (adjusted for the ICD10 coverage of the mortality data). This cost is:

- £10,604 for all PCTs combined
- £14,083 for PCTs whose budget is beyond its target allocation
- £8,441 for PCTs whose budget falls short of its target allocation.

Again, the cost of a life year is much smaller for PCTs whose budget falls short of its target allocation.

Table B8.29: table showing calculation of the cost of a life year for the big four programmes by type of PCT: over target and under target allocations

	A	B	C	D	E	F	G	H	I
	PBC description	Type of PCT	Spend (£m) FY2006/7	1% of spend (£m) FY2006/7	Outcome elasticity (without negative sign)	Total life years lost, <75years, 2006/08	Annual average life years lost (=F/3)	Change in annual life years lost associated with 1% increase in spend (=E*G)/100	Cost (£) per life year gained (=D/H)
	Split PCTs according to whether they are over target allocation (n=67) or under target allocation (n=85)								
1	Cancers	All	4,122	41.22	0.342	2,207,021	735,674	2,516	16,383
2	Cancers	Over target	1,733	17.33	0.179	897,403	299,134	535	32,365
3	Cancers	Under target	2,390	23.9	0.476	1,309,618	436,539	2,078	11,502
4	Circulatory problems	All	6,161	61.61	1.434	1,361,634	453,878	6,509	9,466
5	Circulatory problems	Over target	2,587	25.87	1.115	544,326	181,442	2,023	12,787
6	Circulatory problems	Under target	3,574	35.74	1.947	817,308	272,436	5,304	6,738
7	Respiratory problems	All	3,285	32.85	2.622	324,223	108,074	2,834	11,593
8	Respiratory problems	Over target	1,357	13.57	2.637	127,810	42,603	1,123	12,079
9	Respiratory problems	Under target	1,928	19.28	2.674	196,413	65,471	1,751	11,013
10	Gastro- problems	All	3,700	37	1.536	345,908	115,303	1,771	20,892
11	Gastro- problems	Over target	1,566	15.66	0.569	142,281	47,427	270	58,030
12	Gastro- problems	Under target	2,134	21.34	1.869	203,626	67,875	1,269	16,822

Note that for those over target, the average amount (percentage) is £13.415m (3.6%); for those under target, the average amount (percentage) is £10.575m (2.6%)

Table B8.30: table showing cost of life and life year estimates using spend data for 2006 and outcome data for 2006/08 for the big four PBCs for: (i) all PCTs; (ii) PCTs that are over target; and (iii) PCTs that are under target.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
				=0.01*C*D					=0.01*D*G*I/H		=E/J			=0.01*D*I*N/3		=O/P	=E/O	=E/Q
	PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)		Annual mortality <75 years, 2006/08	Coverage of mortality data relative to spend data	Outcome elasticity (without negative sign)	Change in annual mortality, adj for coverage		Cost per life gained adj for coverage (£)		Total life years lost, <75years, 2006/08	Change in annual life years lost	Coverage of mortality data relative to spend data	Change in annual life years lost, adj for coverage	Cost per life year gained (£)	Cost per life year gained adj for YLL coverage (£)
	All PCTs together																	
1	Cancer	£4,122	0.465	£19.17		61,961	0.984	0.342	100.14		£191,407		2,207,021	1,170	0.984	1,189	£16,383	£16,121
2	Circulatory problems	£6,161	0.540	£33.27		41,106	0.992	1.434	320.88		£103,683		1,361,634	3,515	1.000	3,543	£9,466	£9,390
3	Respiratory problems	£3,285	0.679	£22.31		11,574	0.773	2.622	266.57		£83,676		324,223	1,924	0.773	2,489	£11,593	£8,961
4	Gastro-intestinal problems	£3,700	0.446	£16.50		6,160	0.571	1.536	73.90		£223,228		345,908	790	0.650	1,383	£20,892	£11,929
	Big four programmes summary:																	
5	Spend 2006 & mortality 2006/8	£17,268		£91.24					761.49		£119,823			7,399		8,604	£12,333	£10,604
	For over target PCTs only (n=67)																	
6	Cancer	£1,733	0.193	£3.34		24,918	0.984	0.179	8.75		£382,320		897,403	103	0.984	105	£32,365	£31,847
7	Circulatory problems	£2,587	0.150	£3.88		16,346	0.992	1.115	27.56		£140,806		544,326	303	1.000	303	£12,787	£12,685
8	Respiratory problems	£1,357	0.326	£4.42		4,588	0.773	2.637	51.02		£86,701		127,810	366	0.773	474	£12,079	£9,337
9	Gastro-intestinal problems	£1,566	0.090	£1.41		2,525	0.571	0.569	2.26		£622,378		142,281	24	0.650	43	£58,030	£33,135
	Big four programmes summary:																	
10	Spend 2006 & mortality 2006/8	£7,243		£13.06					89.60		£145,748			797		927	£16,378	£14,083
	For under target PCTs only (n=85)																	
11	Cancer	£2,390	0.785	£18.76		37,043	0.984	0.476	140.67		£133,377		1,309,618	1,631	0.984	1,658	£11,502	£11,318
12	Circulatory problems	£3,574	0.748	£26.73		24,760	0.992	1.947	363.50		£73,544		817,308	3,968	1.000	4,000	£6,738	£6,684
13	Respiratory problems	£1,982	1.035	£20.51		6,986	0.773	2.674	250.12		£82,015		196,413	1,812	0.773	2,344	£11,321	£8,751
14	Gastro-intestinal problems	£2,134	0.592	£12.63		3,602	0.571	1.869	69.80		£181,000		203,626	751	0.650	1,315	£16,822	£9,605
	Big four programmes summary:																	
15	Spend 2006 & mortality 2006/8	£10,080		£78.64					824.09		£95,429			8,162		9,317	£9,635	£8,441
	Note:	2006/7	2005/6															
	All 23 programme spend	£67,896	£64,310															
	% change in budget	1.00	1.00															
	proportionate change	0.01	0.01															
	Change in budget	£678.96	£643.10															

Note that the adjustment for the coverage of the YLL data relative to the spend data uses deaths under age 75 in England in 2008.

B8.11 The correlation between the outcome and expenditure elasticities

To investigate the correlation between the outcome and expenditure elasticities for any given programme, a random sample (with replacement) of 152 PCTs was drawn from the population of 152 PCTs. In this random drawing, some of the original observations will appear once, some more than once, and some not at all. Using this re-sampled dataset, outcome and expenditure models for the selected programme were estimated (as per Table B8.19) and the outcome and expenditure elasticities saved. This step was repeated 500 times and the correlation coefficient for the outcome and expenditure elasticities was calculated. Table B8.31 shows these correlation coefficients for each of the big four programmes.

Table B8.31: table showing correlation coefficient for the outcome and expenditure elasticities

Programme of care	Correlation coefficient between the outcome and expenditure elasticities
Cancers and tumours	0.1542
Circulatory disease	0.1968
Respiratory problems	0.0368
Gastro-intestinal problems	0.0611

Note: the estimated elasticities are from unweighted IV regressions because there is no weight option with the bootstrap command in Stata. However, weighting makes little difference to our IV results. For example, in the cancer outcome model the coefficient on spend is -0.342 with weighting applied but it is -0.299 without any weighting applied. For the cancer spend model the coefficient on budget is 0.465 with weighting but 0.520 without weighting.

B8.12 Summary and conclusion

In this section we have undertaken several tasks. First, we have identified and resolved several estimation issues relating to the appropriate measure of need, the appropriate price index to be used to adjust PCT expenditure for local variations in input prices, and the fact that PCTs vary in size.

Second, we have derived plausible outcome and expenditure models for ten care programmes using expenditure data for 2006/7 and mortality data for 2004/5/6. The cost of a life year across these ten programmes is £20,893 (it was £21,256 using expenditure data for 2005/6).

Third, we have re-estimated the outcome and expenditure models using the same expenditure data but replacing the mortality data for 2004/5/6 with data for 2006/7/8. The advantage of this is that it assumes that the health benefits associated with expenditure occur either in the same period as the expenditure or in the next two periods. This is an improvement on past practice where data constraints forced researchers to relate expenditure to the current and two previous periods. This re-estimation increased the cost of a life year across all ten programmes by 14%, from £20,893 to £23,780.

Fourth, we have adjusted the cost of a life year calculations for the mismatch in the ICD10 coverage of the expenditure and mortality data. This reduces the cost of a life year for 2006/7 for those 10 PBCs with a mortality indicator by 14%, from £23,780 to £19,965 (a decrease of 16%).

Fifth, we have noted that our cost of a life year estimates are based on the assumption that any Departmental budgetary change falls entirely on PCTs. Although PCTs account for most of the Department of Health's budget, non-PCTs still accounted for 15% of the budget in 2006/7. Because we have no information on how any budgetary change would be split between PCTs and non-PCTs, our estimates implicitly assume that any Departmental budgetary change falls entirely on PCTs. If, on the other hand, the non-PCT budget is responsive to changes in the Department's budget then our cost of a life year estimates will be slightly too low (for example, if the non-PCT budget is as responsive as the PCT budget, then our cost of a life year estimate for 2006/7 will be increased by 17.7% (that is, from £22,565 to £26,553).

We have also illustrated how our modelling framework can be applied to other non-mortality based outcome indicators, and the cost of a life year estimates that are obtained if PCTs are split into different groups (e.g., those that under and those that are over their target budget allocations). In the next section we examine the impact of relaxing the instrument validity restriction on our results.

B9. The sensitivity of the outcome elasticity to the validity of the instrument exclusion restrictions

B9.1 Introduction

One of the crucial elements in the calculation of the cost of a life year for any programme of care is the coefficient on the expenditure variable in the outcome equation. This coefficient indicates the amount by which mortality changes following a (small) change in expenditure in that care programme. It is to be expected that this coefficient will have a negative sign so that as expenditure increases, for example, mortality will decline. If this coefficient is small (in an absolute sense) then it implies that any change in expenditure will have little effect on mortality and so the cost of a life year will be relatively large (*ceteris paribus*). Alternatively, if this coefficient is large (in an absolute sense) then any change in expenditure will have a large effect on mortality and so the cost of a life year will be relatively small (*ceteris paribus)*. For this reason it is important that we correctly identify the magnitude of this 'treatment parameter'.

Our basic outcome model for each programme of care is

$$y = \propto + \beta 1x + \beta 2n + \epsilon \qquad (9.1)$$

where y is mortality, x is expenditure, and n is a measure of the need for health care (with all variables relating to a particular programme of care). We are particularly interested in the size of the coefficient on expenditure ($\beta 1$). We do not use OLS to estimate this outcome model because expenditure (x) is endogenous and, in the presence of an endogenous regressor, OLS will provide both a biased and an inconsistent estimator of $\beta 1$. Instead, we use instrumental variable (IV) techniques. Unlike OLS, IV will provide a consistent estimator of $\beta 1$ and, although in finite samples the IV estimator will be biased, the belief is that (providing certain assumptions are met) this bias will be less than that associated with OLS.

IV estimation involves finding variables (instruments) that are good predictors of expenditure (x) but which are appropriately excluded from the equation of interest (that is, equation 1). The assumption is that the instruments impact upon mortality (y) through their impact on expenditure (x) only, and that they do not have a direct effect on mortality (y). If, on the other hand, an instrument reflects unobserved factors that affect both expenditure and mortality directly, the use of this instrument will lead to a biased and inconsistent estimate of the coefficient on expenditure. Such an instrument is said to be 'invalid' because it belongs in the equation of interest in its own right.

In our outcome models we typically employ two instruments (call these z1 and z2) for expenditure. IV assumes that these instruments do not belong in the outcome equation (9.1). In other words, IV assumes that the coefficients $\gamma 1$ and $\gamma 2$ in the outcome model

$$y = \propto + \beta 1x + \beta 2n + \gamma 1z1 + \gamma 2z2 + \epsilon \qquad (9.2)$$

are identically zero. Such exclusion restrictions can be debatable and researchers who employ IV techniques often devote considerable effort towards convincing the reader that their assumed exclusion restrictions are a good approximation[35, 36]. These efforts usually take two forms: first, researchers often offer a strong theoretical economic argument why their instruments do not belong in the equation of interest; and, second, statistical tests for the validity of the exclusion restrictions (Sargan 2SLS, Hansen J-test GMM) are routinely reported as part of the results for any study that employs IV techniques.

It is difficult for us to identify clear theoretical reasons why our instruments (such as the proportion of lone pensioner households, the provision of unpaid care, and an index of multiple deprivation) do not belong in the equation of interest (that is, that they will not *directly* affect mortality). Of necessity, therefore, we must be guided by the available statistical tests for the validity of the exclusion restrictions. However, although our outcome models 'pass' the relevant statistical test, some commentators have argued that the Sargan/Hansen test may have weak power and may fail to reject the null hypothesis of instrument validity even when an exclusion restriction is not valid. Given our reliance on this test, it is important that we examine the circumstances in which this test may have weak power.

B9.2 The Sargan-Hansen test of overidentifying restrictions: when will it have low power?

As we have one endogenous variable (expenditure) in our outcome model and more than one instrument available for health care expenditure, our estimating equation is said to be 'overidentified'. With more instruments than endogenous regressors, there is more than one way of using the instruments to estimate the parameter $\beta 1$ on the endogenous variable. The Sargan-Hansen J test of overidentifying (OID) restrictions calculates whether different instruments or different combinations of instruments generate significantly different values for the coefficient ($\beta 1$) on the endogenous variable in the equation of interest. If significant differences are detected then the test will reject the null hypothesis that all instruments are jointly valid. Of course, the test does not reveal *which* instrument(s) is(are) invalid; instead, the test uses the fact that different instruments (or combinations thereof) generate different estimates of $\beta 1$ to infer that something is wrong with the set of instruments. Even if all of the instruments are invalid in the sense that they are all correlated with the error term in the equation of interest (and thus belong in the outcome equation as regressors), the test can detect this failure if the induced biases in the estimates of $\beta 1$ differ across instruments. This 'vector-of-contrasts' interpretation of the Sargan-Hansen test makes it clear when the J test will lack power to reject the null hypothesis when it is false. The J statistic will be small when the null hypothesis of valid instruments is correct; but it will also be small if the biases induced in $\widehat{\beta 1}$ by invalid instruments all coincide (i.e., the instruments all identify the *same* wrong parameter). [37]

Most of our estimated models involve the use of two instruments. Kovandic, Schaffer and Kleck[37] point that when there are only two instruments '…the J test statistic is numerically identical to a Hausman test statistic that contrasts the estimator using both instruments with an estimator using just one instrument. The intuition [behind this result] is…straightforward: a Hausman test will reject the null hypothesis that the two estimators being contrasted are both consistent so long as the estimators converge to different values. It is not a requirement for one of the two estimators to be consistent for the Hausman test (and therefore the J test) to have power to reject the null.' One implication of this observation is that misspecification, in the conditional mean of the model, need not necessarily cause the Hansen-Sargan test to fail.

Kovandic, Schaffer and Kleck point out that these arguments suggest '…that the more unrelated the instruments are to each other, the more credible is a failure to reject the null that the instruments are exogenous, since a failure to reject would require that two unrelated instruments generate the same asymptotic bias in $\widehat{\beta 1}$' (p19).

Schaffer [40] argues that '[d]ifferent sets of instruments are likely to have more or less power depending on where they come from. If all the instruments are minor variations on the same variable -- e.g., they are the same variable but lagged a few different periods -- then they are all likely to identify the same psuedo-parameter. The critique of low power is going to be fairly convincing here.'

On the other hand, if the instruments are very different and, even better, there are *ex ante* reasons for thinking that if they are invalid, they are invalid in 'different ways', the J test will have more power. For example, suppose two instruments are available and it is thought that, if one is invalid, it will bias the estimated parameter upwards but, if the other instrument is invalid, it will bias the estimated parameter downwards. If the Hansen-Sargan J test fails to reject in this setting, it is a convincing result.[40]

In this study we typically use any two from three available instruments when estimating our outcome equations. These three instruments are:

(a) the proportion of households that are lone pensioner households (from the 2001 Census);
(b) the proportion of residents providing more than one hour of unpaid care per week (from the 2001 Census); and
(c) the index of multiple deprivation (IMD) 2007.

For the Hansen-Sargan J test to have low power the use of any two of these instruments should generate the *same* asymptotic bias in $\widehat{\beta 1}$. However, it is far from obvious that this will be the case, particularly given that our outcome equation already includes a measure of the need for health care.

Nevertheless, we must admit that it is possible that our instruments are correlated with both expenditure and some unobserved factor which is directly influencing the mortality rate, and that the induced bias in $\widehat{\beta 1}$ is the same for both instruments. In the next section we therefore examine the sensitivity of the estimated outcome elasticity to the validity of the exclusion restrictions.

In summary:

- the Sargan-Hansen J test of overidentifying restrictions calculates whether different instruments or different combinations of instruments generate significantly different values for the coefficient ($\beta 1$) on the endogenous variable in the equation of interest. If significant differences are detected then the test will reject the null hypothesis that all instruments are jointly valid.

- the J test uses the fact that different instruments (or combinations thereof) generate different estimates of $\beta 1$ to infer that something is wrong with the chosen set of instruments.

- even if all of the instruments are invalid in the sense that they are all correlated with the error term in the equation of interest, the test can detect this failure if the induced biases in the estimates of $\beta 1$ differ across instruments.

- this 'vector-of-contrasts' interpretation of the Sargan-Hansen test also makes it clear when the J test will lack power to reject the null hypothesis when it is false. The J statistic will be small when the null hypothesis of valid instruments is correct; but it will also be small if the biases induced in $\widehat{\beta 1}$ by invalid instruments all coincide (i.e., the instruments all identify the *same* wrong parameter).

- most of our outcome models use two from the following three instruments: lone pensioners, multiple deprivation and unpaid carers. Thus our Hansen-Sargan test statistics are likely to have low power if our selected pair of instruments are both inducing the same bias in $\widehat{\beta 1}$. It is far from obvious that these instruments will induce the same bias in the coefficient on expenditure.

- however, in case our instruments are imparting the same bias to $\widehat{\beta 1}$, the next section examines the sensitivity of the estimated outcome elasticity to the validity of the exclusion restrictions.

B9.3 The value selection problem

Given that the Hansen-Sargan J test might be unable to detect the presence of invalid instruments in some (rather restrictive) circumstances, several studies have suggested that researchers using IV techniques should subject the estimated coefficient on the endogenous variable to a sensitivity analysis (e.g., Conley, Hansen and Rossi, 2012; Small, 2007). Recall that IV estimation involves the assumption that the instruments do not belong in the equation of interest (i.e., in the outcome equation). In other words, the assumption is that the coefficients $\gamma 1$ and $\gamma 2$ on the instruments z1 and z2 in the outcome model

$$y = \propto + \beta 1x + \beta 2n + \gamma 1z1 + \gamma 2z2 + \epsilon \qquad (9.3)$$

are identically zero (where y is mortality, x is expenditure, and n is a measure of the need for health care). One suggestion is that investigators should relax the assumption that $\gamma 1$ and $\gamma 2$ are identically zero and examine the impact of this relaxation on the estimated value for $\beta 1$. This proposal, however, raises the issue of which non-zero values should be imposed upon $\gamma 1$ and $\gamma 2$.

Proponents of this approach suggest that prior information about the extent of deviations from the exact exclusion restriction might be drawn from other research studies or from subject matter experts[35, 36]. In the present context, however, we have no prior beliefs about the likely values for, or even the signs on, $\gamma 1$ and $\gamma 2$.

As a starting point we re-estimated the outcome model for the 2006/7 cancer programme 420 times, assuming a uniform distribution between -1 and 1 for both $\gamma 1$ and $\gamma 2$.[32] [33] Table B9.1 shows the estimated coefficients on expenditure ($\widehat{\beta 1}$) in our cancer outcome equation associated with the various pairs of values imposed upon $\gamma 1$ and $\gamma 2$. The coefficients in this table indicate that the outcome elasticity is rather sensitive to the precise values assigned to $\gamma 1$ and $\gamma 2$. However, in the absence of any guidance from other research studies or from subject matter experts, we require a method that will identify a plausible range of values for both $\gamma 1$ and $\gamma 2$, and which we can use as the basis for our sensitivity analysis.

[32] See column 1 of Table 8.19 for the estimated IV cancer outcome model.

[33] We used a symmetric distribution about zero because we have no priors about the signs of the coefficients on the instruments. The use of a uniform distribution is arbitrary but of no significance.

Table B9.1: table showing the impact of weakening the exclusion restrictions on the instruments in the cancer outcome equation

Coefficients on expenditure ($\widehat{\beta 1}$) ↘	Imposed coefficient on IMD variable (γ_2)																				
	-1	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.00	-4.47	-4.22	-3.96	-3.71	-3.45	-3.20	-2.95	-2.70	-2.45	-2.20	-1.96	-1.71	-1.47	-1.23	-0.99	-0.75	-0.50	-0.24	0.03	0.30	0.57
0.90	-4.31	-4.06	-3.80	-3.54	-3.29	-3.04	-2.79	-2.53	-2.29	-2.04	-1.80	-1.55	-1.31	-1.07	-0.83	-0.59	-0.33	-0.07	0.20	0.47	0.74
0.80	-4.15	-3.90	-3.64	-3.38	-3.13	-2.88	-2.62	-2.37	-2.13	-1.88	-1.64	-1.40	-1.16	-0.92	-0.67	-0.42	-0.16	0.10	0.37	0.65	0.92
0.70	-3.99	-3.74	-3.48	-3.22	-2.97	-2.72	-2.46	-2.21	-1.96	-1.72	-1.48	-1.24	-1.00	-0.76	-0.51	-0.26	0.01	0.28	0.55	0.82	1.10
0.60	-3.83	-3.58	-3.32	-3.06	-2.81	-2.55	-2.30	-2.05	-1.80	-1.56	-1.32	-1.08	-0.84	-0.60	-0.35	-0.09	0.18	0.45	0.73	1.00	1.28
0.50	-3.67	-3.41	-3.16	-2.90	-2.65	-2.39	-2.14	-1.89	-1.64	-1.40	-1.16	-0.92	-0.68	-0.44	-0.18	0.08	0.35	0.63	0.91	1.18	1.45
0.40	-3.51	-3.25	-3.00	-2.74	-2.49	-2.23	-1.98	-1.73	-1.48	-1.23	-0.99	-0.76	-0.52	-0.27	-0.01	0.26	0.53	0.81	1.09	1.36	1.63
0.30	-3.35	-3.09	-2.84	-2.58	-2.32	-2.07	-1.82	-1.57	-1.32	-1.07	-0.83	-0.60	-0.35	-0.10	0.16	0.44	0.71	0.99	1.26	1.53	1.80
0.20	-3.19	-2.93	-2.68	-2.42	-2.16	-1.91	-1.65	-1.40	-1.15	-0.91	-0.67	-0.43	-0.19	0.07	0.34	0.62	0.90	1.17	1.44	1.71	1.97
0.10	-3.03	-2.77	-2.51	-2.26	-2.00	-1.75	-1.49	-1.24	-0.99	-0.75	-0.51	-0.27	-0.02	0.25	0.52	0.80	1.07	1.34	1.61	1.88	2.14
0.00	-2.87	-2.61	-2.35	-2.10	-1.84	-1.58	-1.33	-1.07	-0.83	-0.58	-0.34	-0.10	0.15	0.43	0.70	0.98	1.25	1.52	1.78	2.04	2.31
-0.10	-2.71	-2.45	-2.19	-1.93	-1.68	-1.42	-1.16	-0.91	-0.66	-0.42	-0.18	0.07	0.33	0.61	0.88	1.15	1.42	1.68	1.95	2.21	2.47
-0.20	-2.55	-2.29	-2.03	-1.77	-1.51	-1.26	-1.00	-0.75	-0.50	-0.25	-0.01	0.24	0.51	0.79	1.06	1.32	1.59	1.85	2.11	2.37	2.63
-0.30	-2.39	-2.13	-1.87	-1.61	-1.35	-1.10	-0.84	-0.58	-0.33	-0.09	0.15	0.41	0.68	0.95	1.22	1.48	1.75	2.01	2.27	2.53	2.79
-0.40	-2.23	-1.97	-1.71	-1.45	-1.19	-0.93	-0.68	-0.42	-0.17	0.06	0.30	0.57	0.84	1.11	1.38	1.64	1.91	2.17	2.43	2.69	2.95
-0.50	-2.07	-1.81	-1.55	-1.29	-1.03	-0.78	-0.52	-0.27	-0.03	0.20	0.44	0.72	1.00	1.27	1.54	1.80	2.06	2.32	2.58	2.84	3.10
-0.60	-1.91	-1.65	-1.39	-1.13	-0.87	-0.62	-0.37	-0.13	0.10	0.32	0.58	0.87	1.15	1.42	1.69	1.95	2.22	2.48	2.74	3.00	3.26
-0.70	-1.75	-1.49	-1.23	-0.97	-0.72	-0.47	-0.22	0.01	0.22	0.46	0.73	1.02	1.30	1.57	1.84	2.11	2.37	2.63	2.89	3.16	3.42
-0.80	-1.59	-1.33	-1.07	-0.82	-0.57	-0.32	-0.09	0.14	0.35	0.60	0.88	1.17	1.45	1.72	1.99	2.26	2.52	2.79	3.05	3.31	3.57
-0.90	-1.43	-1.17	-0.92	-0.67	-0.42	-0.18	0.05	0.27	0.50	0.76	1.04	1.32	1.60	1.88	2.15	2.41	2.68	2.94	3.20	3.47	3.73
-1.00	-1.27	-1.02	-0.76	-0.52	-0.27	-0.04	0.18	0.41	0.65	0.92	1.20	1.48	1.76	2.03	2.30	2.57	2.83	3.09	3.36	3.62	3.88

Imposed coefficient on lone pensioner households variable (γ_1)

Notes: This spreadsheet shows the value of the coefficient on expenditure ($\widehat{\beta 1}$) when estimating the cancer outcome equation [$y^* = (y - \gamma_1 z_1 - \gamma_2 z_2) = \propto + \beta 1x + \beta 2n + \epsilon$] using IV having imposed different pairs of values for $\gamma 1$ and $\gamma 2$ between -1 and 1. Cells in the top left-hand quadrant contain negative values for the outcome elasticity. The outcome elasticity associated with our standard IV model (-0.34) is shown in the central square where $\gamma 1$ and $\gamma 2$ are, of course, zero.

B9.4 The identification of values to be imposed on the coefficients on the excluded instruments

Our outcome equations typically involve two instruments and one endogenous regressor. With this structure we can re-estimate our outcome model twice, each time including one of the previously excluded instruments to the equation of interest. In particular, we can estimate

$$y = \propto + \beta 1x + \beta 2n + \gamma 1z1 + \qquad \epsilon \qquad (9.4)$$

and then

$$y = \propto + \beta 1x + \beta 2n \qquad + \gamma 2z2 + \epsilon \qquad (9.5)$$

with the same set of (included and excluded) instruments (n, z1, and z2) being used to instrument x1 in both cases. This provides us with coefficient and variance estimates for $\gamma 1$ and $\gamma 2$ and we can sample from these point estimates and their distributions to examine the impact of different (non-zero) values for $\gamma 1$ and $\gamma 2$ on the outcome elasticity ($\widehat{\beta 1}$).

The sampling procedure is straightforward. We sample from these estimates and their distributions by drawing two random numbers from a standard normal distribution and we form the product of these numbers and the standard errors associated with our estimates of $\gamma 1$ and $\gamma 2$. Our sampled pair of values of $\gamma 1$ and $\gamma 2$ (call these sampled values $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$) are then the sum of these products and the respective coefficient estimates of $\gamma 1$ and $\gamma 2$. Table B9.2 shows the relevant coefficient and variance estimates for $\gamma 1$ and $\gamma 2$ that are used as part of this sampling procedure.

Table B9.2: table showing various estimates associated with the excluded instruments from the outcome equation for the big four programmes

Programme	Instrument	coefficient	Std error	variance	covariance
Cancer	z1: lone pensioner households	-0.2074942	0.1773647	0.0314582	0.00494454
	z2: IMD2007	-0.0827677	0.1141054	0.0130200	
Circulatory disease	z1: lone pensioner households	-0.2606290	0.2441101	0.059590	0.01122591
	z2: IMD2007	-0.2105334	0.2879230	0.082900	
Respiratory problems	z1: long term unemployment rate	0.2642582	0.13273061	0.0176174	-0.02136305
	z2: limiting long-term illness rate	-1.739808	1.611403	2.5966196	
Gastro-intestinal problems	z1: unpaid carers	1.812286	2.347459	5.510564	0.08016639
	z2: IMD2007	0.5567431	0.2066839	0.0427822	

The estimation of equations (9.4) and (9.5) does not generate estimates of $\gamma 1$ and $\gamma 2$ as part of the same model and so the sampling procedure outlined above implicitly assumes a zero covariance between these estimates. If we want to incorporate a covariance term into the sampling procedure this must be obtained from elsewhere. In the absence of an obviously better approach, we obtain a covariance term from the OLS estimation of our outcome model with the previously excluded instruments both included in the regression equation. Thus we estimate

$$y = \propto + \beta 1x + \beta 2n + \gamma 1z1 + \gamma 2z2 + \epsilon \qquad (9.6)$$

where $y, x, n, z1$ and $z2$ have their usual meaning, and where $\beta 1$ is constrained to be equal to its value from the IV estimation of equation (9.1). The value for the covariance term between the estimates of $\gamma 1$ and $\gamma 2$ for each of the big four programmes is shown in the final column of Table B9.2.

The sampling procedure from our estimates of $\gamma 1$ and $\gamma 2$ with a non-zero covariance is essentially the same as that outlined above but it incorporates the presence of a covariance term for the estimates of $\gamma 1$ and $\gamma 2$. We can illustrate this procedure using data from the cancer outcome model. First, we form the implied variance-covariance matrix for the estimates of $\gamma 1$ and $\gamma 2$

	lone_pensioners	IMD2007
lone_pensioners	0.031458	0.004944
IMD2007	0.004944	0.013020.

Second, we form the product of a pair of random numbers from the standard normal distribution (r1, r2) with the Cholesky decomposition matrix from the variance-covariance matrix for the estimates of $\gamma 1$ and $\gamma 2$. The latter is given by:

	lone_pensioners	IMD2007
lone_pensioners	0.177364	0
IMD2007	0.027877	0.110647.

Finally, we add this pair of products to the respective coefficient estimates of $\gamma 1$ and $\gamma 2$ to obtain our sampled pair of values of $\gamma 1$ and $\gamma 2$ (call these $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$). In other words, for each pair of random numbers (r1, r2), we calculate the sampled values $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ where

$$\begin{bmatrix} \tilde{\gamma}_1 \\ \tilde{\gamma}_2 \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} + \left[\begin{bmatrix} a11 & a12 \\ a21 & a22 \end{bmatrix} \begin{bmatrix} r1 \\ r2 \end{bmatrix} \right]$$

sampled values for coefficients on excluded instruments	=	coefficients from IV regressions (9.4) & (9.5) above	+	(Cholesky (decomposition * (matrix | pair of random numbers from standard normal distribution)))

This sampling procedure is undertaken 1,000 times, both with a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$, and again with a non-zero covariance.

These procedures generate two sets of 1,000 pairs of values for $\hat{\gamma}_1$ and $\hat{\gamma}_2$ (one set assumes a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ and the other does not). These sets of values for $\hat{\gamma}_1$ and $\hat{\gamma}_2$ can be used to examine the sensitivity of the estimated outcome elasticity to alternative non-zero values for the coefficients on the excluded instruments.

B9.5 Obtaining the outcome elasticities associated with sampled coefficients on the excluded instruments

For each pair of sampled values of $\gamma 1$ and $\gamma 2$ ($\tilde{\gamma}_1$ and $\tilde{\gamma}_2$), we can use IV techniques to estimate the model

$$y_{new} = y - \tilde{\gamma}1z1 - \tilde{\gamma}2z2 \ = \propto + \beta 1x + \beta 2n + \epsilon \qquad (9.7)$$

with the usual instrument set (x2, z1, and z2) used to instrument the endogenous variable x1 (expenditure). For each pair of sampled the values $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, we obtain a different outcome elasticity ($\widetilde{\beta 1}$) and these different values can be plotted in a histogram. Such a plot illustrates the uncertainty

associated with our point estimate of the outcome elasticity due to doubts about the validity of our instruments; we call this type of uncertainty 'level 2' uncertainty. This 'level 2' uncertainty is in addition to what we label 'level 1' uncertainty, that is, the uncertainty about the value of the outcome elasticity assuming the validity of our exclusion restrictions (remember that our estimated outcome elasticity is only a point estimate and that it has a distribution attached to it). To illustrate this 'level 1' uncertainty, we can sample from the distribution of the point estimate for the outcome elasticity from our basic IV model (where $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are zero in equation (9.7)) and plot the sampled values.

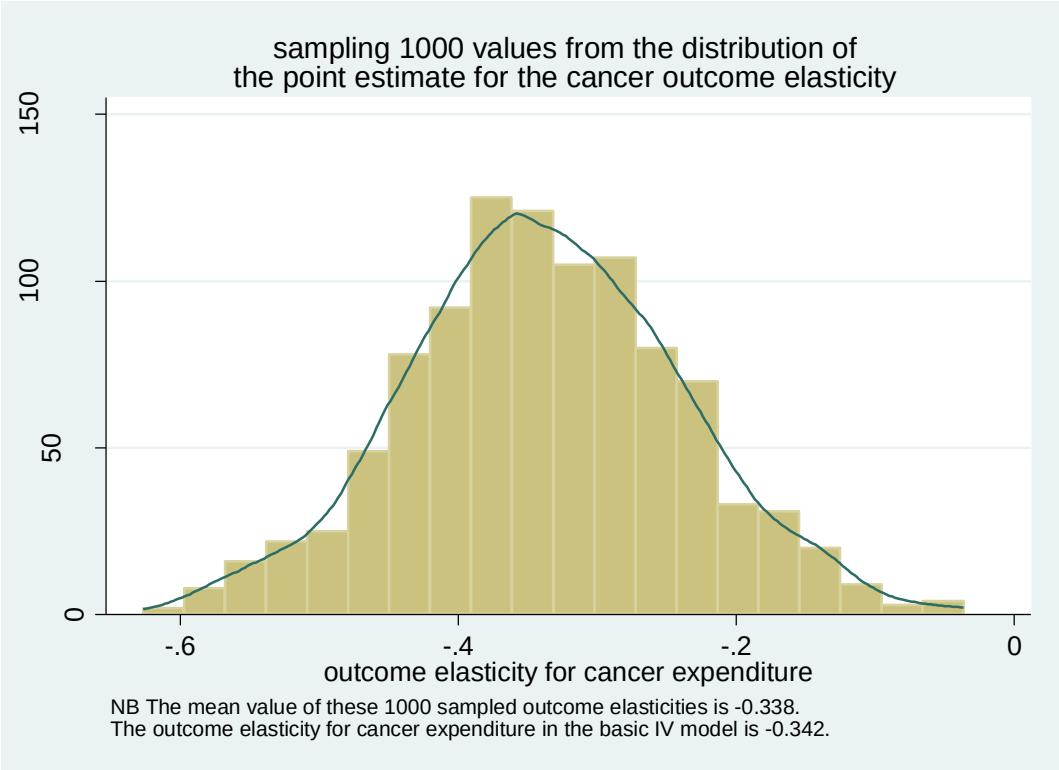
Plots illustrating the degree of level 1 uncertainty for each of the big four programmes are shown as Figures B9.1a (cancer), B9.2a (circulatory disease), B9.3a (respiratory problems), and B9.4a (gastro-intestinal problems) below. These level 1 uncertainty plots can be compared with plots of $\widehat{\beta 1}$ from the estimation of equation (9.7). The latter plots illustrate the degree of level 2 uncertainty, that is, the uncertainty associated with our point estimate of the outcome elasticity due to doubts about the validity of the instruments. Figures B9.1b (cancer), B9.2b (circulatory disease), B9.3b (respiratory problems), and B9.4b (gastro-intestinal problems) show plots of the outcome elasticity ($\widehat{\beta 1}$ from equation (9.7)) assuming a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5). Figures B9.1c (cancer), B9.2c (circulatory disease), B9.3c (respiratory problems), and B9.4c (gastro-intestinal problems) show plots of the outcome elasticity ($\widehat{\beta 1}$from equation (9.7)) assuming a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5).

Finally, the point estimates $\widehat{\beta 1}$ from the estimation of equation (9.7) also have a standard error and we can sample from these distributions. These sampled values illustrate what we term 'level 3' uncertainty, that is, the uncertainty associated with the value of the outcome elasticity due to both level 1 (sampling) and level 2 (instrument invalidity) effects.

Plots illustrating the degree of level 3 uncertainty, assuming a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5), are shown as Figures B9.1d (cancer), B9.2d (circulatory disease), B9.3d (respiratory problems), and B9.4d (gastro-intestinal problems) below. Plots illustrating the degree of level 3 uncertainty, assuming a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5), are shown as Figures B9.1e (cancer), B9.2e (circulatory disease), B9.3e (respiratory problems), and B9.4e (gastro-intestinal problems) below.

Uncertainty and the value of the cancer outcome elasticity (Figures B9.1a-B9.1e)
Figure B9.1a plots 1000 values from the distribution of the point estimate for the cancer outcome elasticity. The mean value of these sampled values is -0.338 (the outcome elasticity in the basic IV model is -0.342 and its standard error is 0.099) and virtually all of them lie between 0 and -0.6.

Figure B9.1a: graph showing histogram for the sampled cancer outcome elasticity

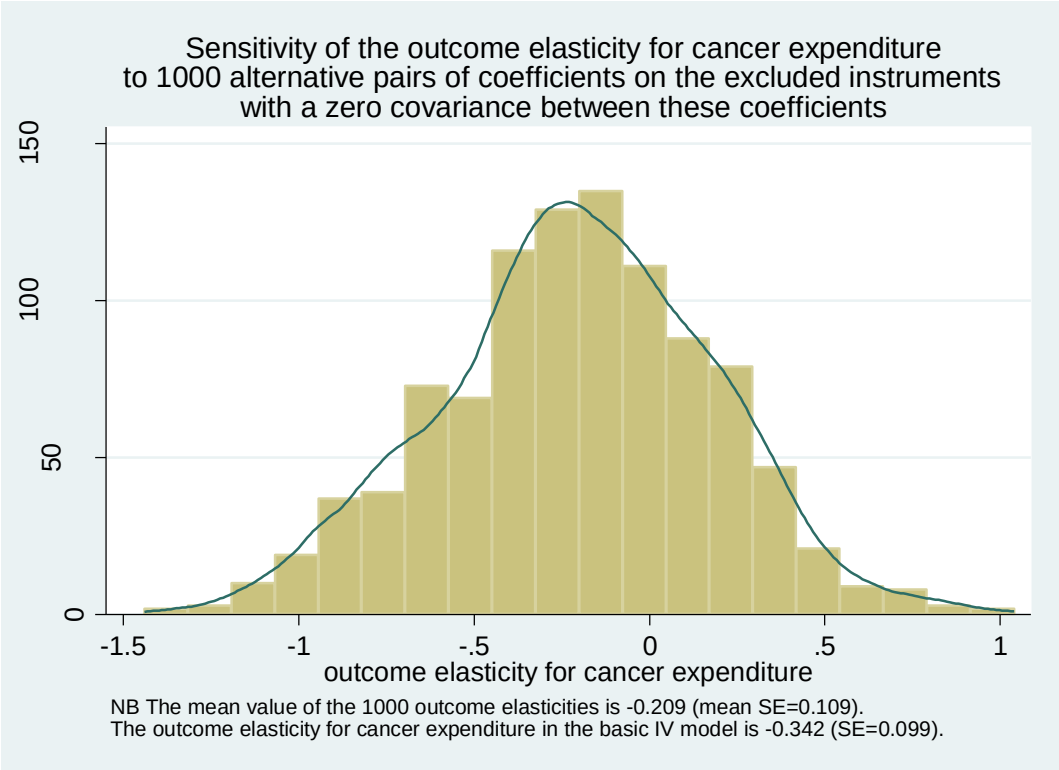


The histogram in Figure B9.1b provides a plot of 1000 point estimates for the cancer outcome elasticity if we drop the assumption that the coefficients on the excluded instruments are exactly zero (and we also assume a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). The mean value of these 1000 outcome elasticities is -0.209 and this is about one-third lower than the elasticity in the basic IV model (= -0.342). In addition, the mean value of the standard errors associated with these 1000 elasticities (0.109) is slightly greater than the standard error in the basic IV model (0.099) so that about one-quarter of the outcome elasticities in Figure B9.1b take a non-negative value.

The histogram in Figures B9.1c provides a similar plot to that in Figure B9.1b but this time we assume a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5). There is very little difference between the zero (Figure B9.1b) and non-zero (9.1c) covariance plots, with both the mean elasticity and mean standard error virtually identical in both plots.

The histograms in Figures B9.1b and B9.1c provide plots of the point estimate for the cancer outcome elasticity if we drop the assumption that the coefficients on the excluded instruments are exactly zero. Each point estimate also has a standard error and we can sample from these estimates and their distributions to obtain the histograms shown in Figures B9.1d and B9.1e. With the exception of a slight lengthening in the tail on the left hand side, these plots are very similar to the plots in Figures B9.1b and B9.1c.

Figure B9.1b: graph showing histogram for estimated outcome elasticity associated with cancer outcome model (zero covariance between the coefficients on the excluded instruments)



FigureB9.1c: graph showing histogram for estimated outcome elasticity associated with cancer outcome model (non-zero covariance between the coefficients on the excluded instruments)

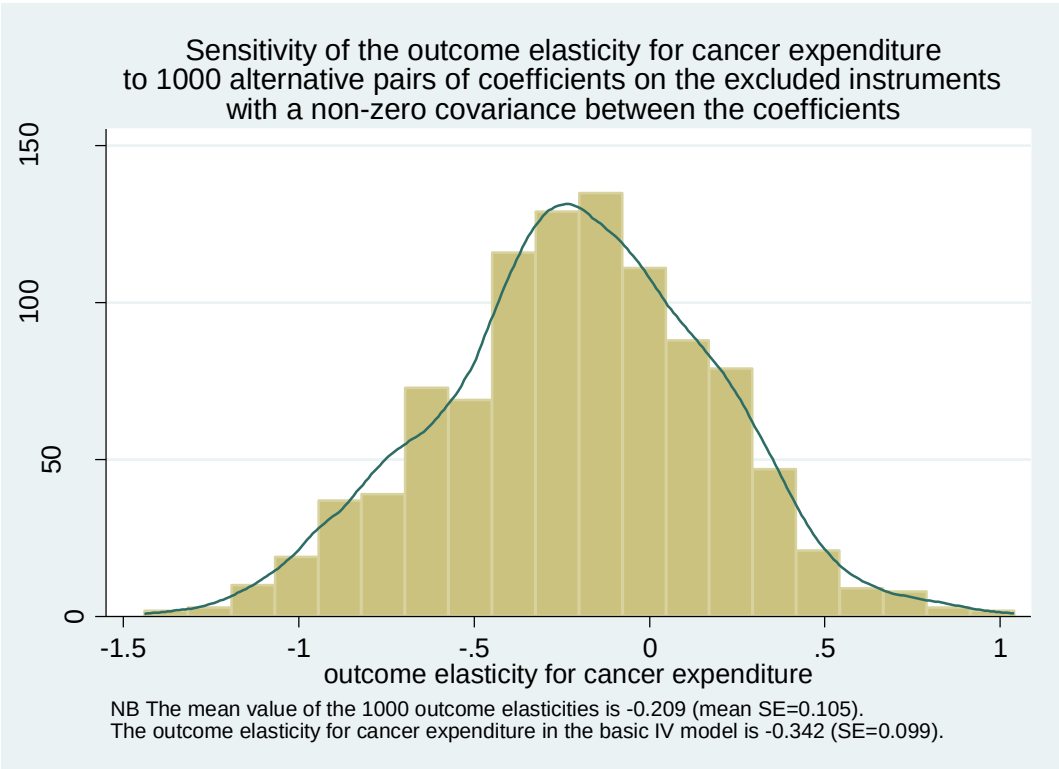


Figure B9.1d: graph showing histogram for sampled values of the cancer outcome elasticity (zero covariance between the coefficients on the excluded instruments)

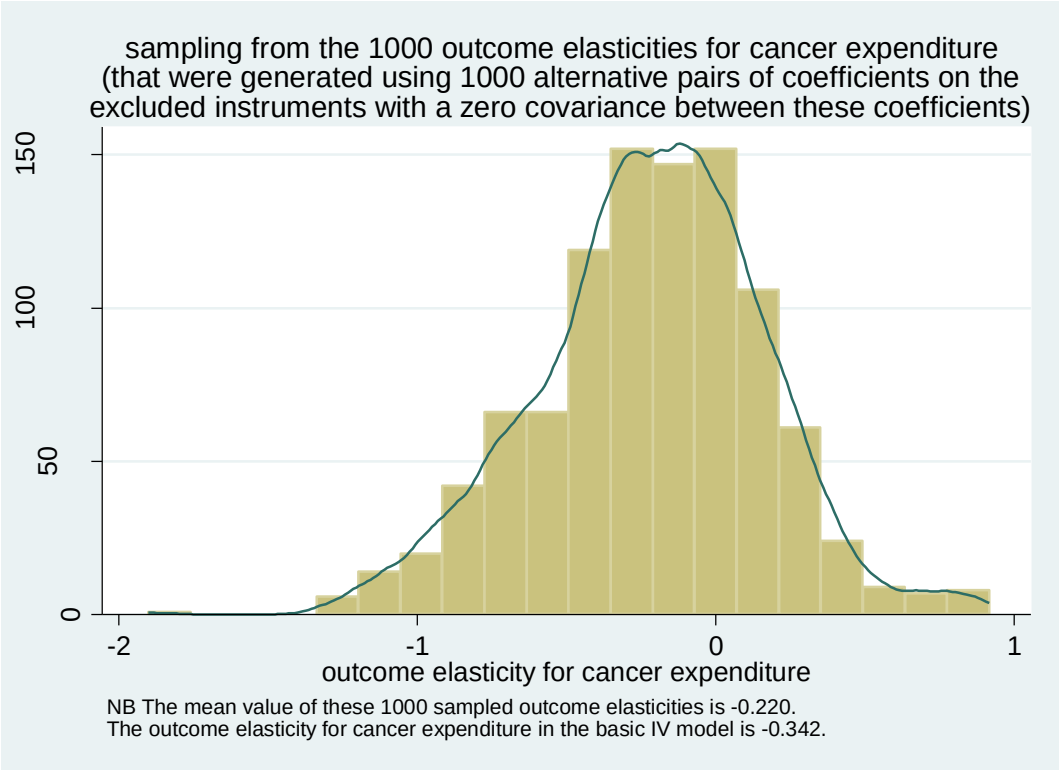


Figure B9.1e: graph showing histogram for sampled values of the cancer outcome elasticity (non-zero covariance between the coefficients on the excluded instruments)

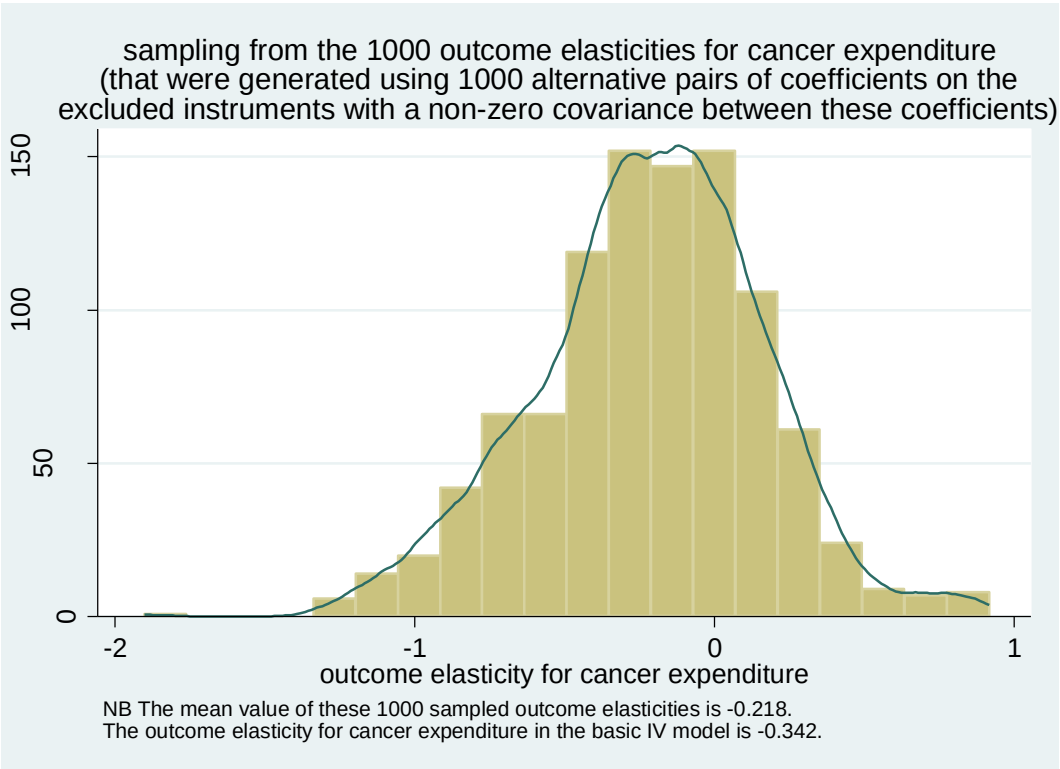


Figure B9.1f reproduces the three kernel density plots shown in Figures B9.1a, B9.1b and B9.1d (remember that Figures B9.1b and B9.1d assume a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4)

and (9.5)). Together these plots illustrate the impact of all three levels of uncertainty on our estimate of the cancer outcome elasticity. It is clear that the uncertainty induced by the instrument validity issue considerably increases the uncertainty associated with our estimate of the outcome elasticity (compare, for example, the density plot for level 1 uncertainty with those for both level 2 and level 3 uncertainty).

Figure B9.1f: graph showing comparing the three kernel density plots shown in Figures B9.1a, 9.1b and 9.1d

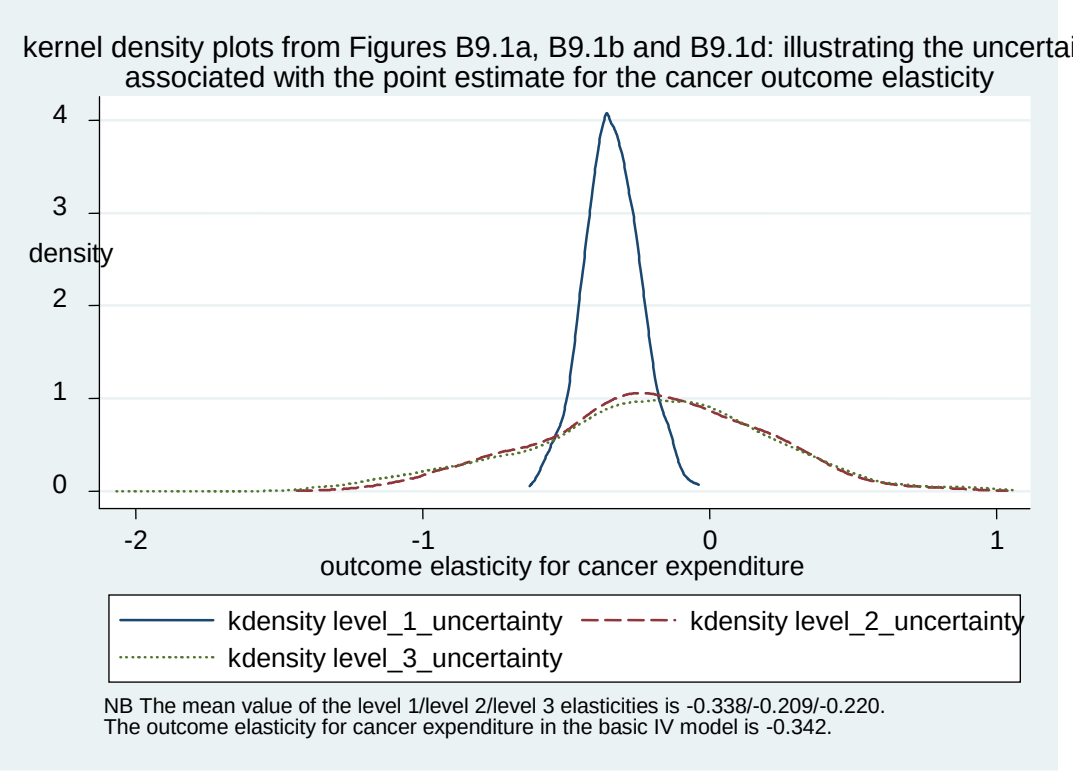
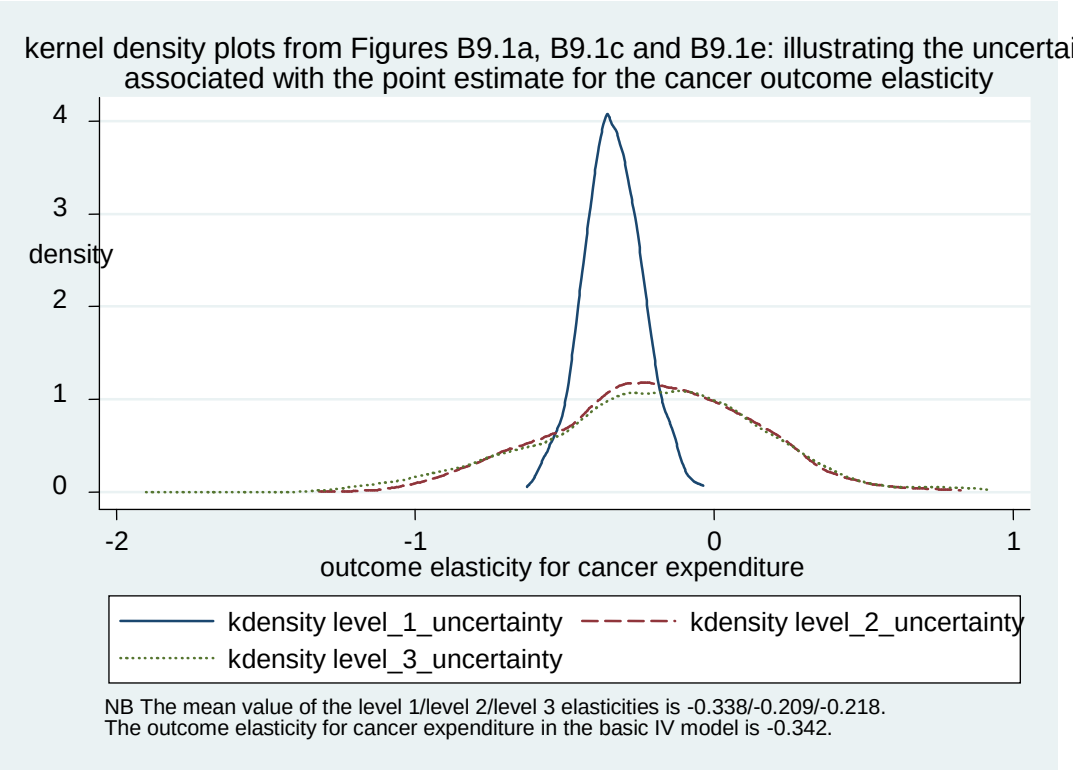


Figure B9.1g reproduces the three kernel density plots shown in Figures B9.1a, B9.1c and B9.1e (remember that Figures B9.1c and B9.1e assume a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). As is the case for Figure B9.1f, these plots illustrate the impact of all three levels of uncertainty on our estimate of the cancer outcome elasticity. And again, it is clear that it is the uncertainty induced by the instrument validity issue that considerably increases the uncertainty associated with our estimate of the outcome elasticity. For example, the standard deviation associated with the level 1 uncertainty density plot is 0.099 but the standard deviation for the level 2 (0.338) and level 3 (0.379) uncertainty density plots are both considerably larger than this.

Figure B9.1g: graph showing comparing the three kernel density plots shown in Figures B9.1a, B9.1c and B9.1e

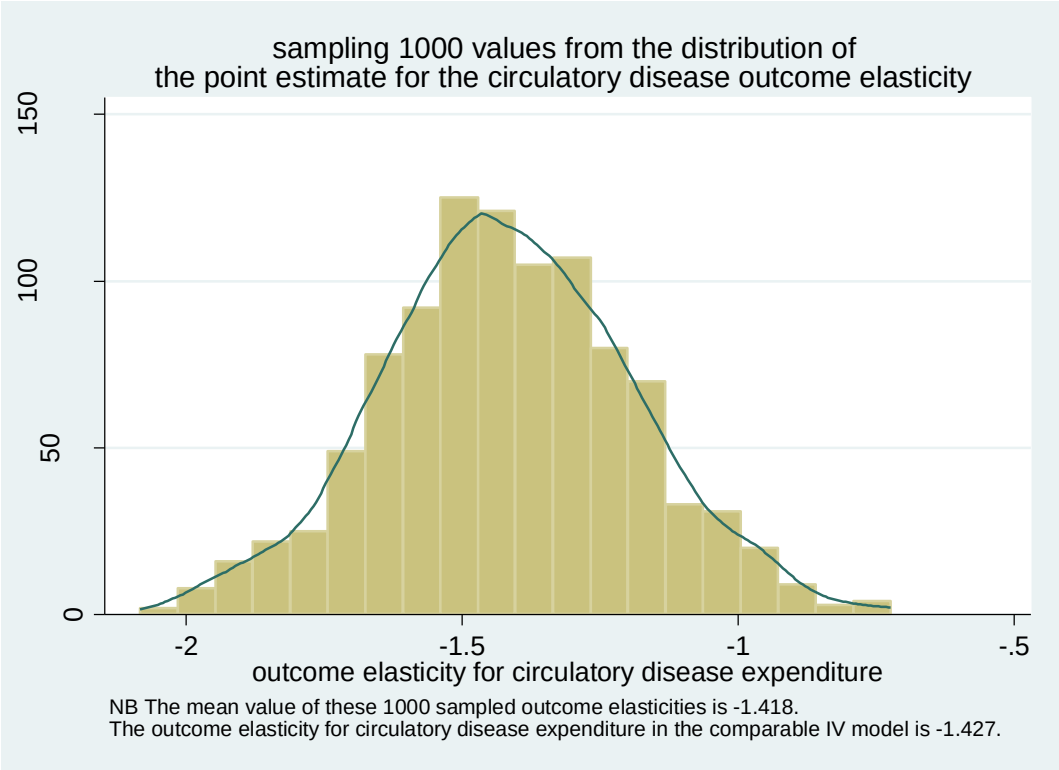


Uncertainty and the value of the circulatory disease outcome elasticity (Figures B9.2a-B9.2e)

Figure B9.2a plots 1000 values from the distribution of the point estimate for the circulatory disease outcome elasticity. The mean value of these sampled values is -1.418 and virtually all of these values lie between -2.0 and -0.75. The outcome elasticity in the comparable IV model is -1.427.[34]

[34]The outcome model for circulatory disease reported in Table B8.19 (using PB expenditure for 2006/7 and mortality data for 2006/7/8) contains four instruments. The application of the sensitivity analysis described in this section is considerably easier to implement if only two instruments are present and re-estimation of the outcome model for circulatory disease without the two least significant instruments generates very similar results to those obtained with all four instruments (for example, the coefficient on expenditure declines marginally from -1.434 to -1.427). Therefore the sensitivity analysis reported here uses the outcome model containing only two instruments.

Figure B9.2a: graph showing histogram for the sampled circulatory disease outcome elasticity



The histogram in Figure B9.2b provides a plot of 1000 point estimates for the circulatory disease outcome elasticity if we drop the assumption that the coefficients on the excluded instruments are exactly zero (and we also assume a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). The mean value of these 1000 outcome elasticities is -1.697 and this is about one-fifth larger than the elasticity in the comparable IV model (-1.427). Similarly, the mean value of the standard errors associated with these 1000 elasticities (0.269) is also about one-fifth larger than the standard error in the comparable basic IV model (0.228). Virtually all of the point estimates values lie between -4.0 and 0.0, and there are very few non-negative values.

The histogram in Figure B9.2c provides a similar plot to that in Figure B9.2b but this time we assume a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5). There is very little difference between the zero (Figure B9.2b) and non-zero (9.2c) covariance plots, with both the mean elasticity and mean standard error virtually identical in both plots.

The histograms in Figures B9.2b and B9.2c provide plots of the point estimate for the circulatory disease outcome elasticity if we drop the assumption that the coefficients on the excluded instruments are exactly zero. Each point estimate also has a standard error and we can sample (with replacement) from these estimates and their distributions to obtain the histograms shown in Figures B9.2d and B9.2e. With the exception of a slight lengthening in the tail on the left hand side (as was also observed for the cancer programme), these plots are very similar to the plots in Figures B9.2b and B9.2c.

Figure B9.2b: graph showing histogram for estimated outcome elasticity associated with the circulatory disease outcome model (zero covariance between the coefficients on the excluded instruments)

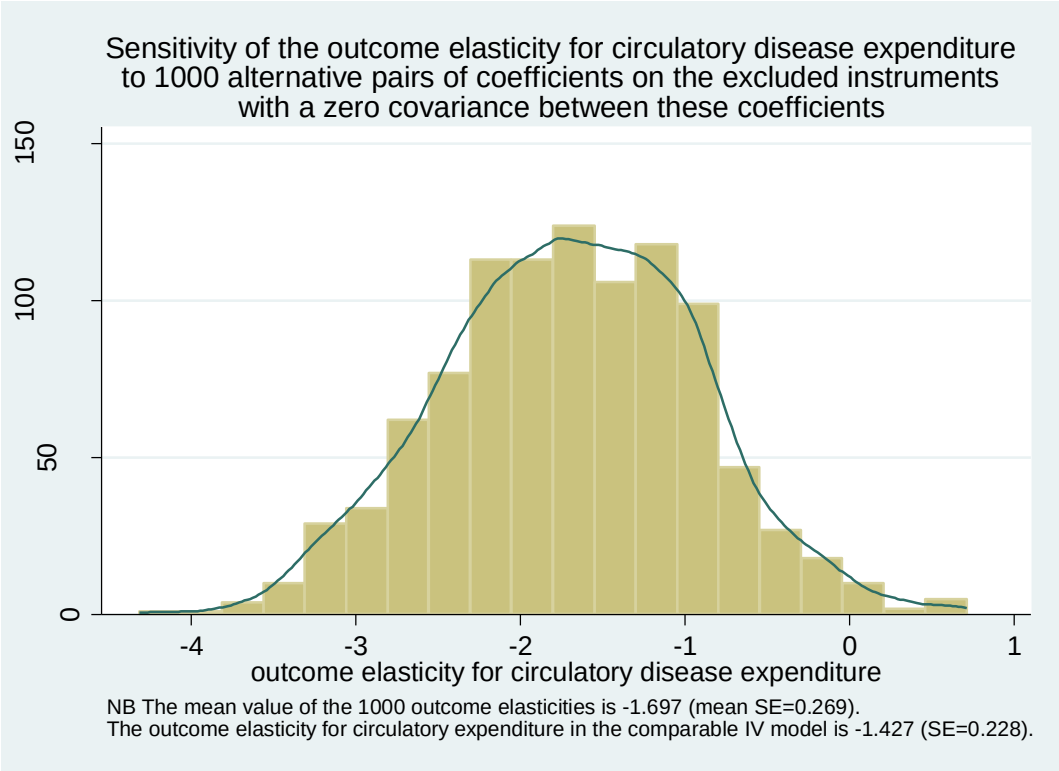


Figure B9.2c: graph showing histogram for estimated outcome elasticity associated with the circulatory disease outcome model (non-zero covariance between the coefficients on the excluded instruments)

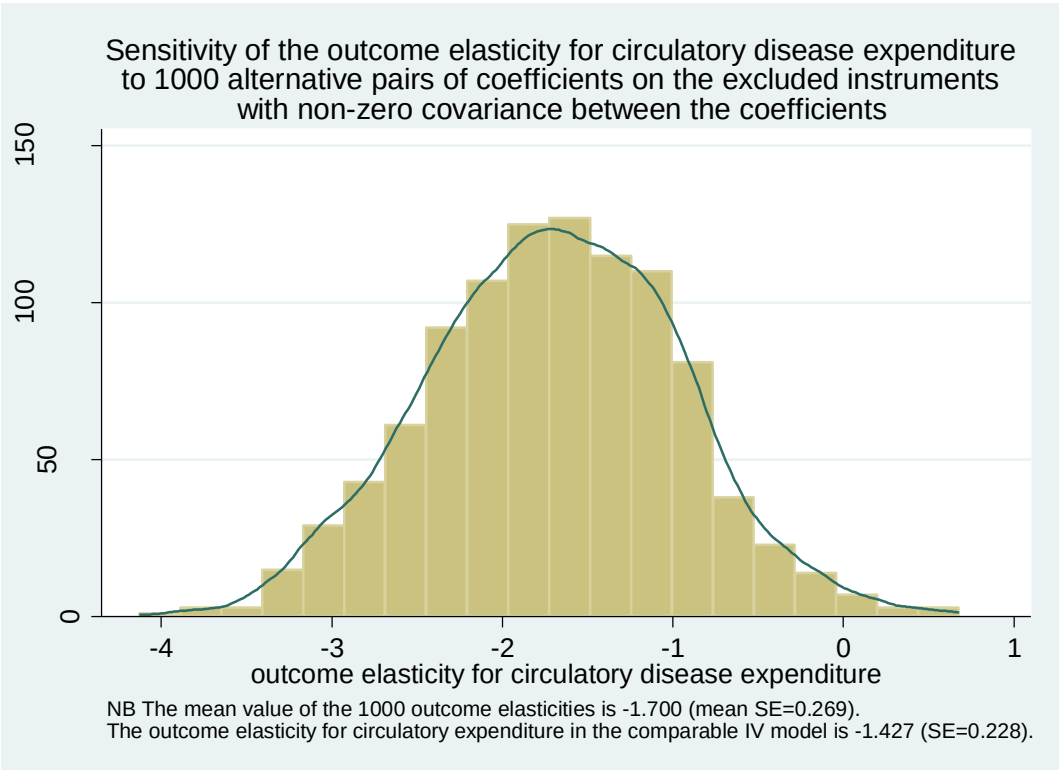


Figure B9.2d: graph showing histogram for sampled values of the circulatory disease outcome elasticity (zero covariance between the coefficients on the excluded instruments)

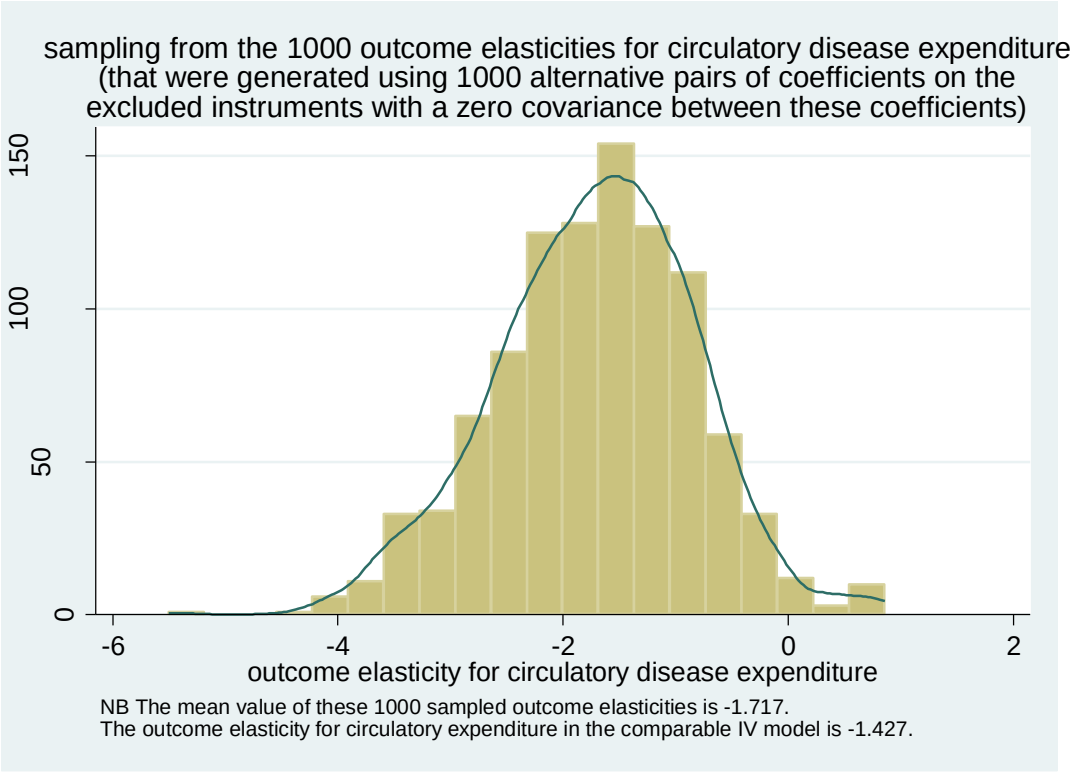


Figure 9.2e: graph showing histogram for sampled values of the circulatory disease outcome elasticity (non-zero covariance between the coefficients on the excluded instruments)

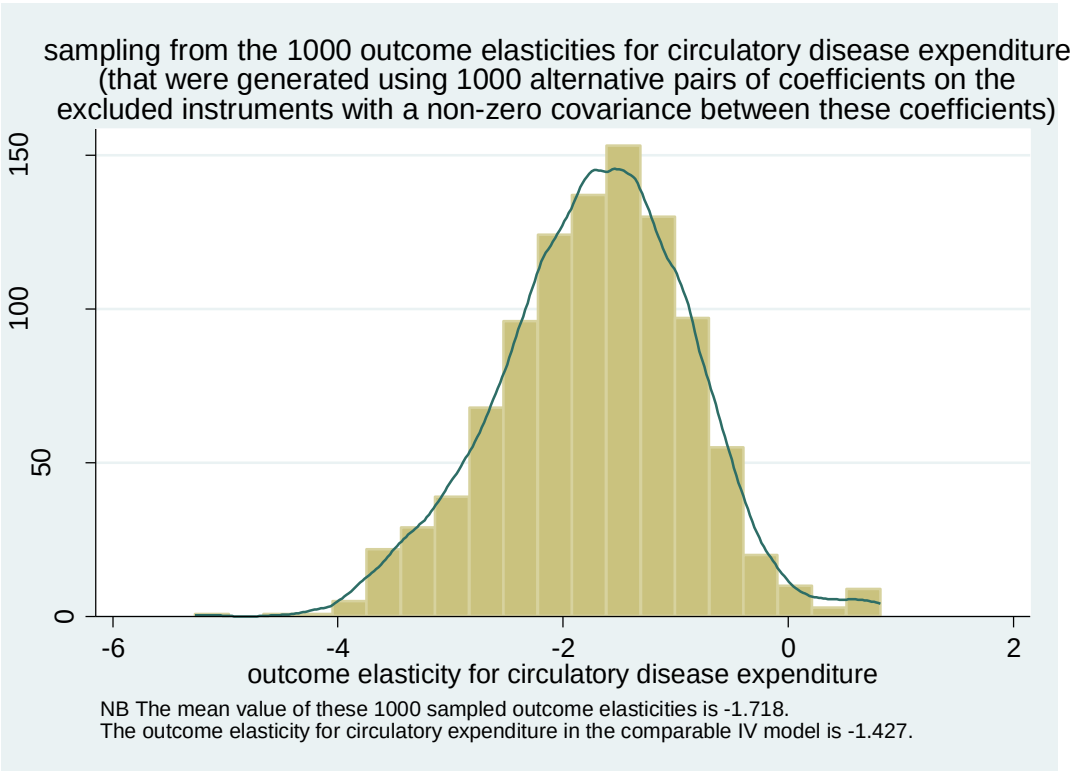


Figure B9.2f reproduces the three kernel density plots shown in Figures B9.2a, B9.2b and B9.2d (remember that Figures B9.2b and B9.2d assume a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). Together these plots illustrate the impact of all three levels of uncertainty on our estimate of the circulatory disease outcome elasticity. It is clear that the uncertainty induced by the instrument validity issue considerably increases the uncertainty associated with our estimate of the outcome elasticity (note that the range of values increases dramatically from the density plot illustrating level 1 uncertainty to that illustrating level 2 uncertainty).

Figure B9.2f: graph showing comparing the three kernel density plots shown in Figures B9.2a, B9.2b and B9.2d

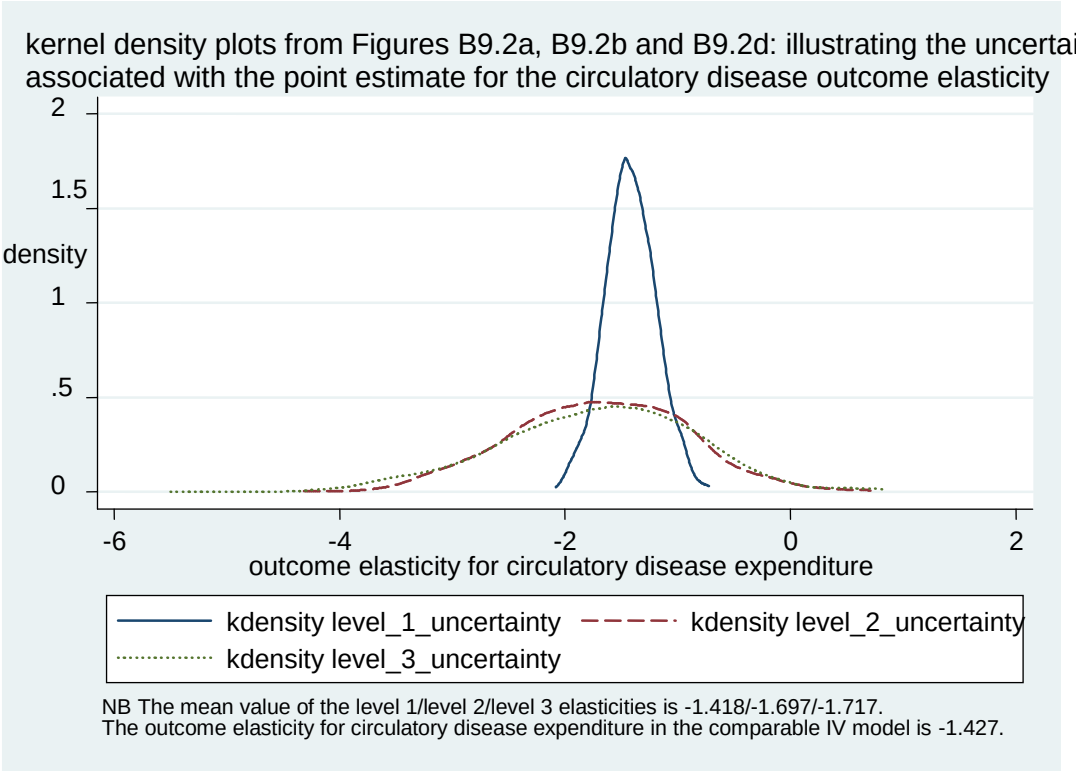
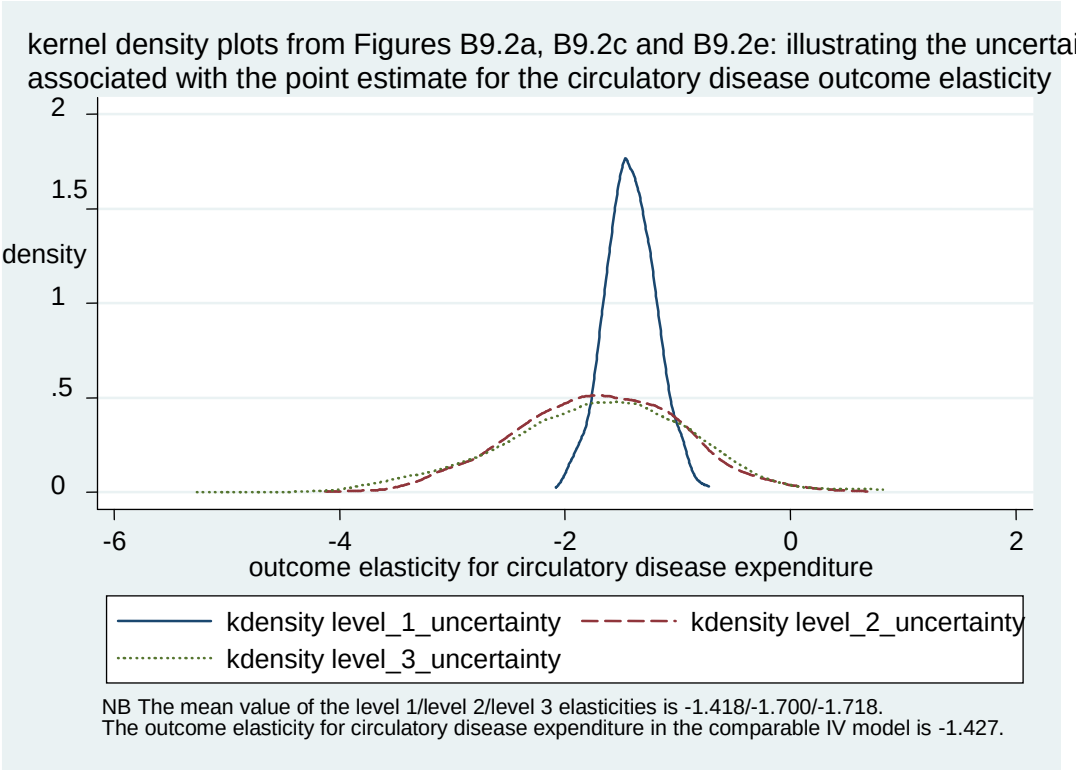


Figure B9.2g reproduces the three kernel density plots from Figures B9.2a, B9.2c and B9.2e (remember that Figures B9.2c and B9.2e assume a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). As is the case for Figure B9.2f, these plots illustrate the impact of all three levels of uncertainty on our estimate on the circulatory disease outcome elasticity. And again, it is clear that it is the uncertainty induced by the instrument validity issue that considerably increases the uncertainty associated with our estimate of the outcome elasticity. For example, the standard deviation associated with the level 1 uncertainty density plot is 0.228 but the standard deviation for the level 2 (0.735) and level 3 (0.843) uncertainty density plots are both considerably larger than this.

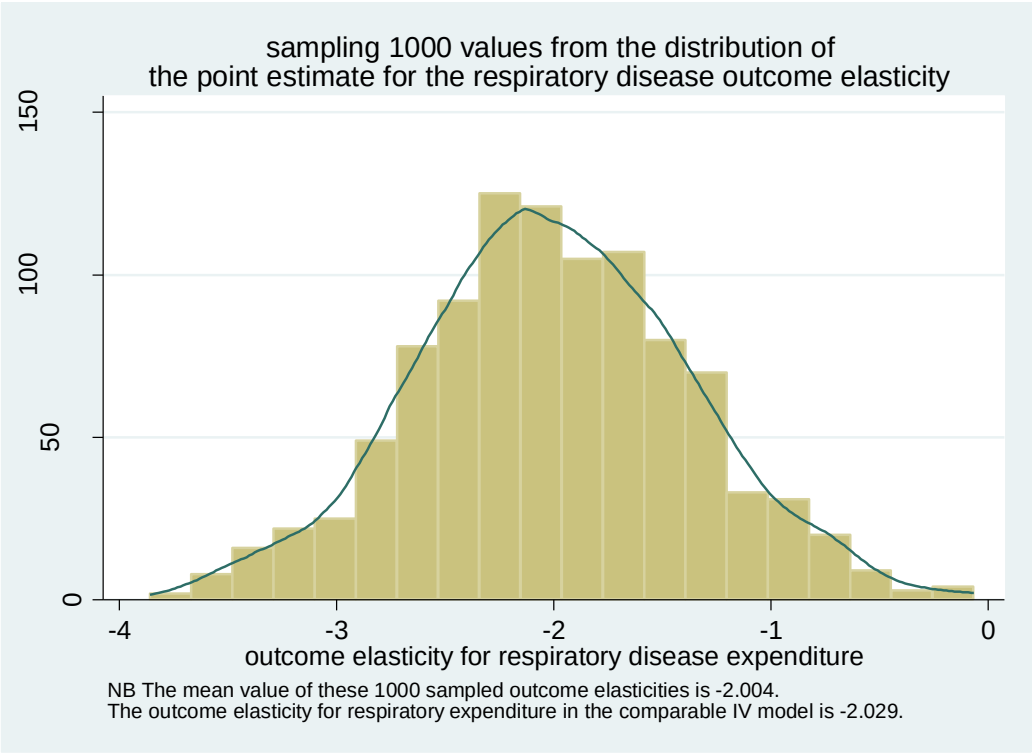
Figure B9.2g: graph showing comparing the three kernel density plots shown in Figures B9.2a, B9.2c and B9.2e



Uncertainty and the value of the respiratory disease outcome elasticity (Figures B9.3a-B9.3e)

Figure B9.3a plots 1000 values from the distribution of the point estimate for the respiratory disease outcome elasticity (see column 5 of Table B8.19). The mean value of these sampled values is -2.004 (the outcome elasticity in the comparable IV model is -2.029) and all of these values lie between -4.0 and 0.

Figure B9.3a: graph showing histogram for the sampled respiratory disease outcome elasticity



The histogram in Figure B9.3b provides a plot of 1000 point estimates for the respiratory disease outcome elasticity if we drop the assumption that the coefficients on the excluded instruments are exactly zero (and we also assume a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). The mean value of these 1000 outcome elasticities from the respiratory disease outcome model (-1.145) is almost one-half of the size of the elasticity in the comparable basic IV model (-2.029). And the mean value of the standard errors associated with these 1000 elasticities (0.489) is about one-quarter less than the standard error in the comparable basic IV model (0.636).

The histogram in Figure B9.3c provides a similar plot to that in Figure B9.3b but this time we assume a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5). However, there is very little difference between the zero (Figure B9.3b) and non-zero (B9.3c) covariance plots, with both the mean elasticity and mean standard error virtually identical in these two plots.

The histograms in Figures B9.3b and B9.3c provide plots of the point estimate for the respiratory disease outcome elasticity if we drop the assumption that the coefficients on the excluded instruments are exactly zero. Each point estimate also has a standard error and we can sample from these estimates and their distributions to obtain the histograms shown in Figures B9.3d and B9.3e. With the exception of a slight lengthening of the tail on the left hand side, these plots are similar to those in Figures B9.3b and B9.3c so the sampling procedure would appear to have little impact on the distribution of the point elasticities.

Figure B9.3b: graph showing histogram for estimated outcome elasticity associated with respiratory disease outcome model (zero covariance between the coefficients on the excluded instruments)

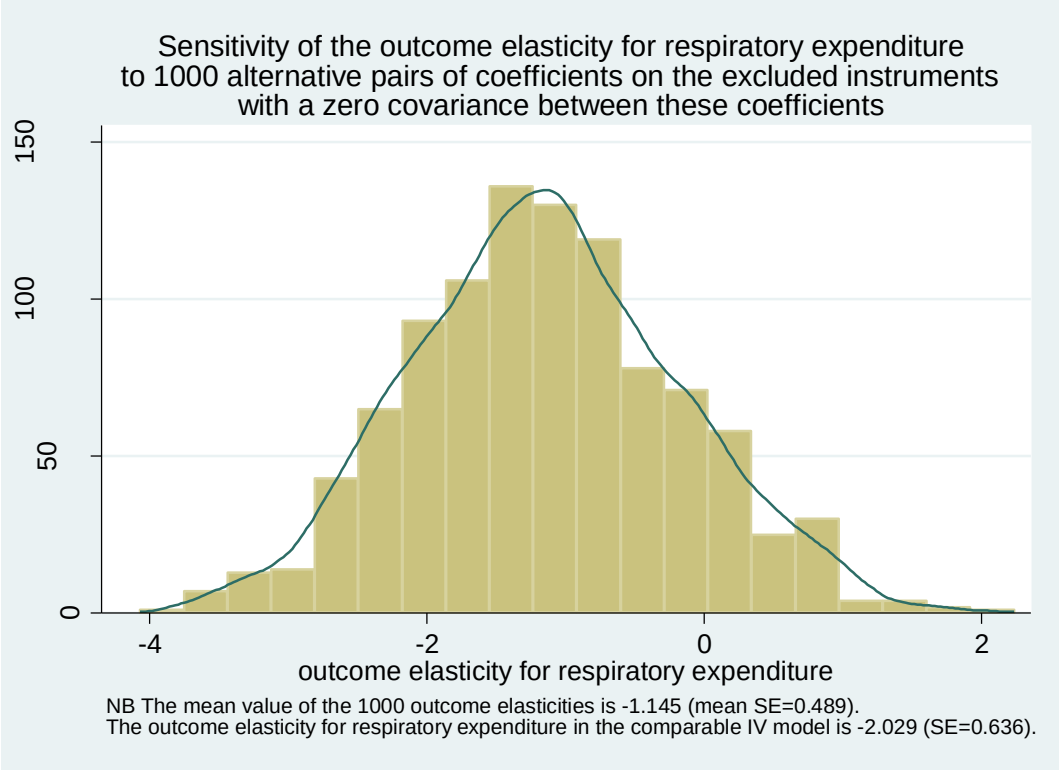


Figure B9.3c: graph showing histogram for estimated outcome elasticity associated with respiratory disease outcome model (non-zero covariance between the coefficients on the excluded instruments)

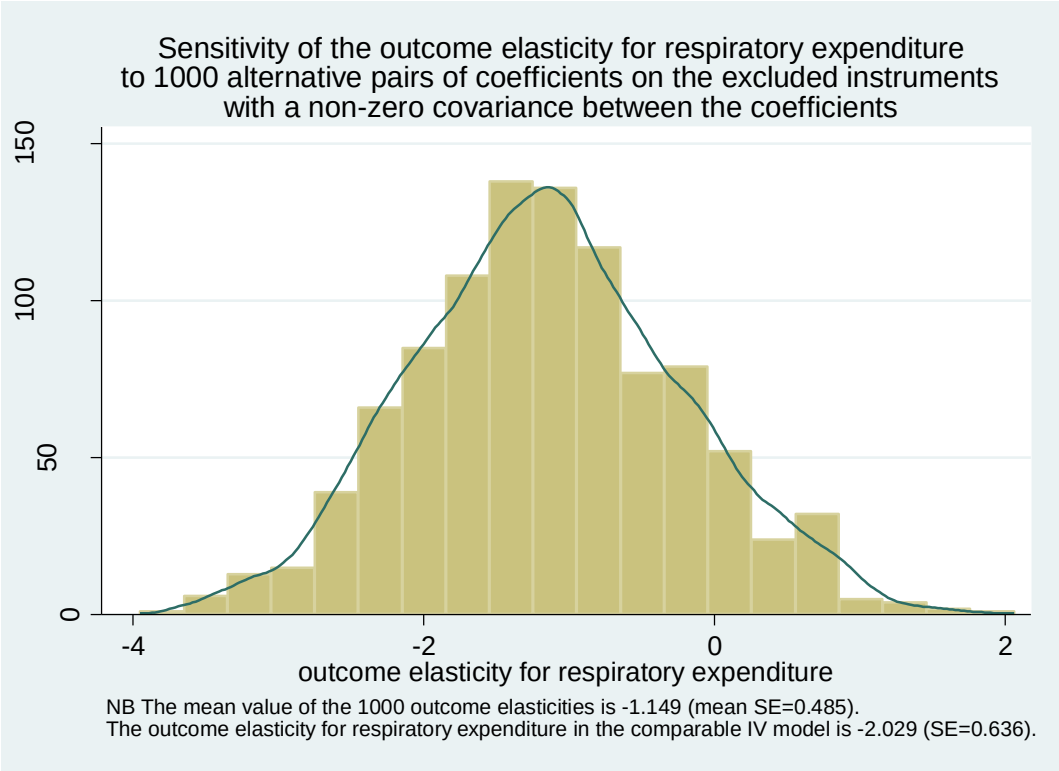


Figure B9.3d: graph showing histogram for sampled values of the respiratory disease outcome elasticity (zero covariance between the coefficients on the excluded instruments)

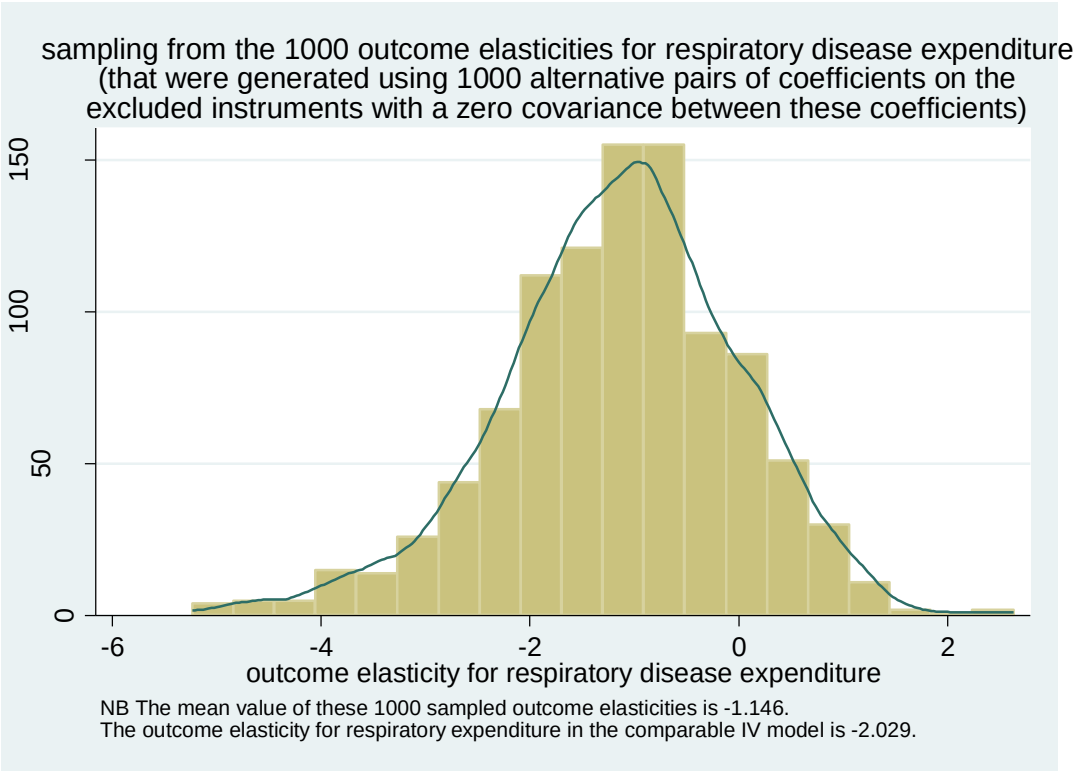


Figure B9.3e: graph showing histogram for sampled values of the respiratory disease outcome elasticity (non-zero covariance between the coefficients on the excluded instruments)

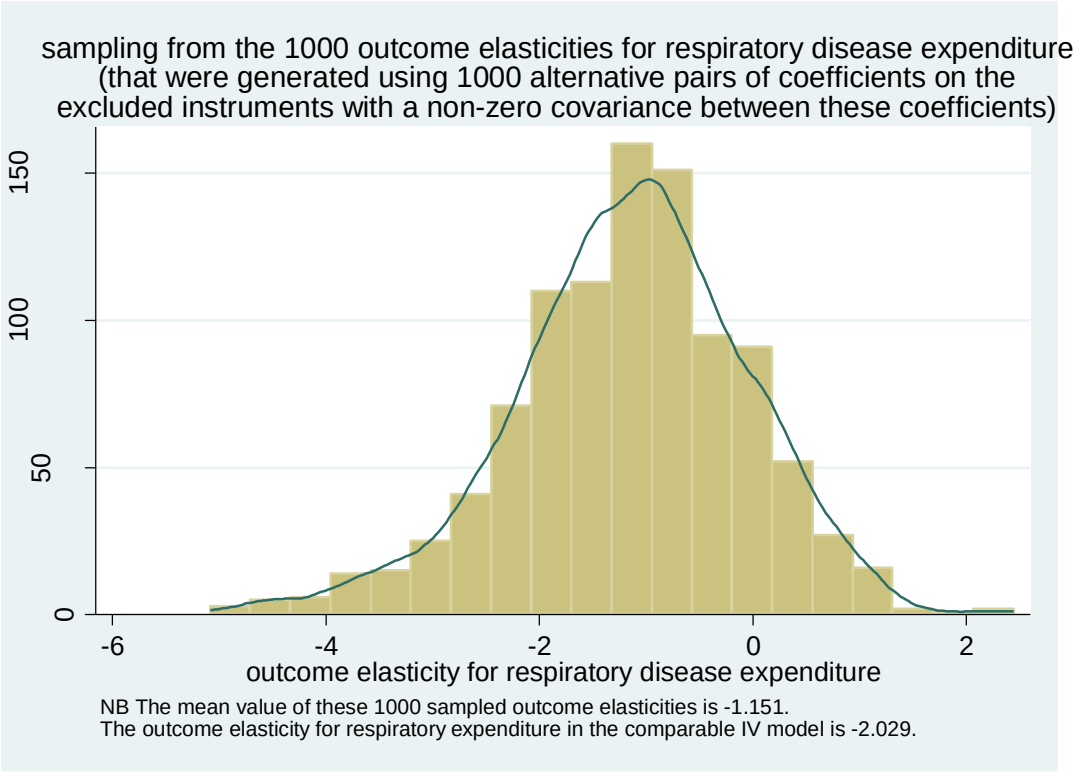


Figure B9.3f reproduces the three kernel density plots shown in Figures B9.3a, B9.3b and B9.3d (remember that Figures B9.3b and B9.3d assume a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). Together these plots illustrate the impact of all three levels of uncertainty on our estimate of the respiratory disease outcome elasticity. It is clear that the uncertainty induced by the instrument validity issue both shifts the density plot to the right and increases the uncertainty associated with our estimate of the outcome elasticity (e.g., the range of values increases from -4.0 to 0.0 at level 1 to -5.0 to 2.5 at level 3).

Figure B9.3f: graph showing comparing the three kernel density plots shown in Figures B9.3a, B9.3b and B9.3d

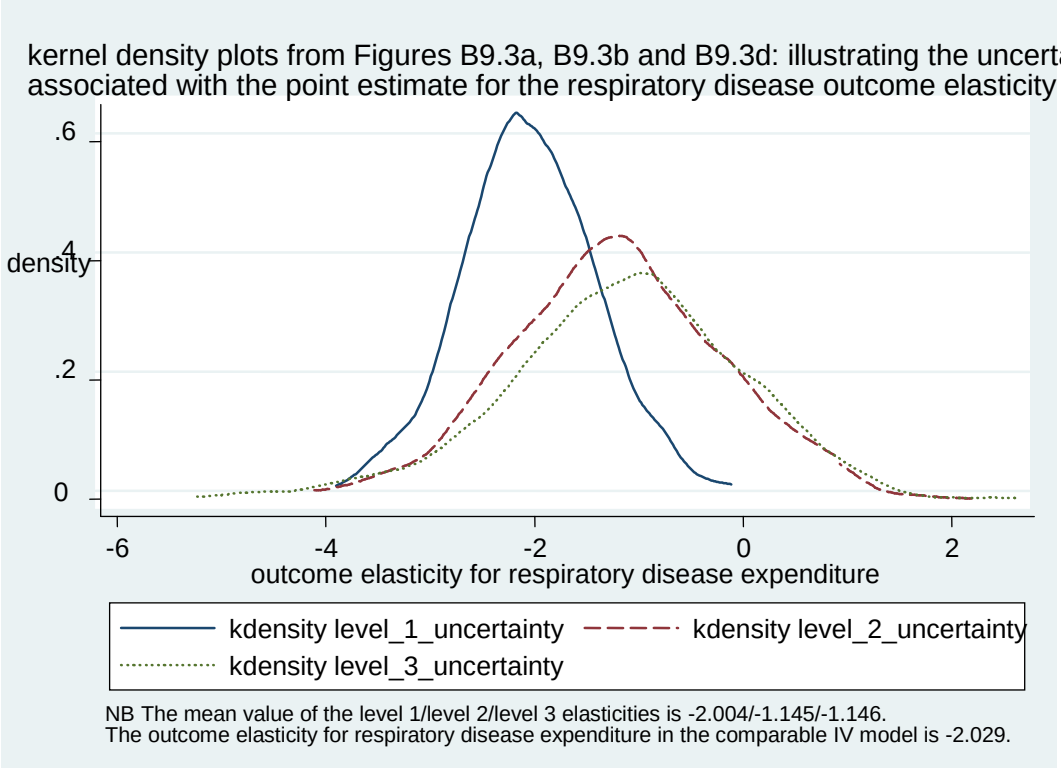
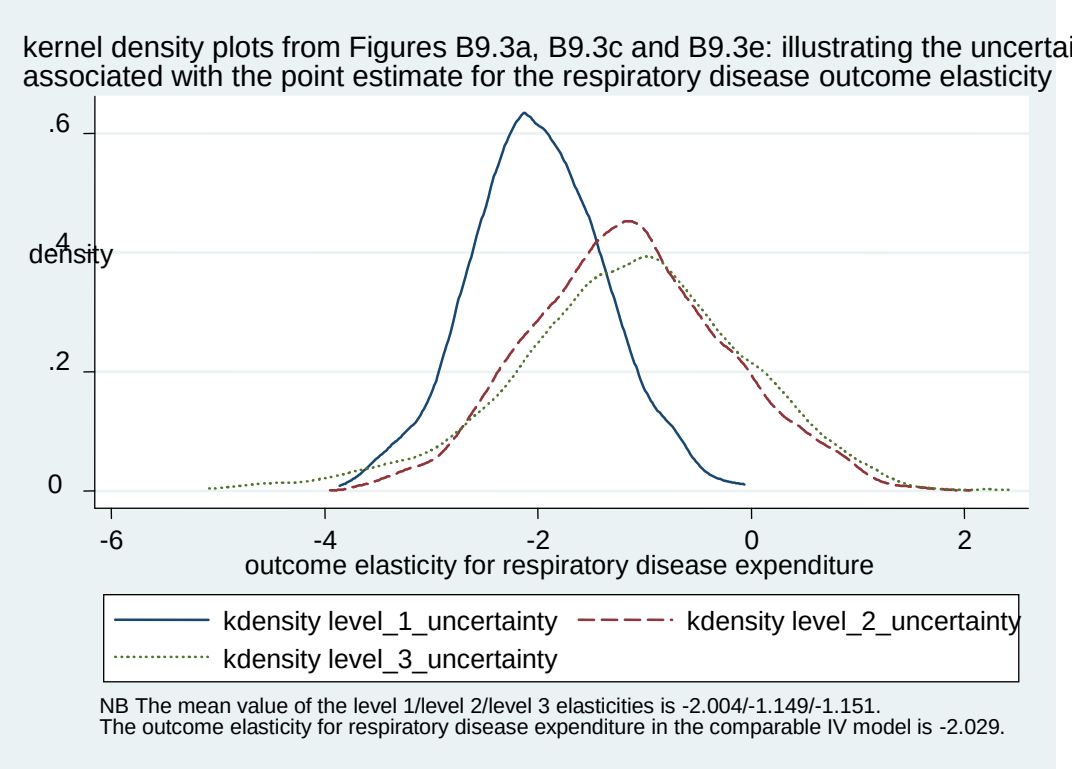


Figure B9.3g reproduces the three kernel density plots shown in Figures B9.3a, B9.3c and B9.3e (remember that Figures B9.3c and B9.3e assume a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). As is the case for Figure B9.3f, these plots illustrate the impact of all three levels of uncertainty on our estimate of the respiratory disease outcome elasticity. And again, it is clear that the uncertainty induced by the instrument validity issue both shifts the density plot to the right and considerably increases the uncertainty associated with our estimate of the outcome elasticity. More precisely, the standard deviation associated with the level 1 uncertainty density plot is 0.636 but the standard deviation for the level 2 (0.919) and level 3 (1.098) uncertainty density plots are both considerably larger than this.

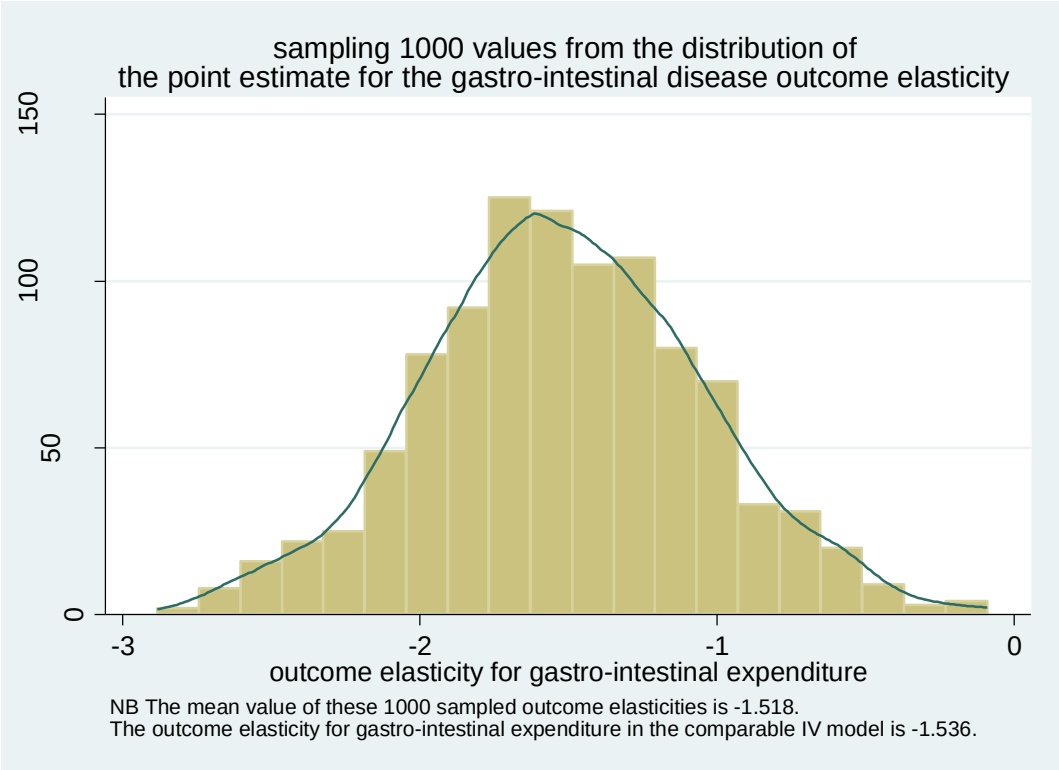
Figure B9.3g: graph showing comparing the three kernel density plots shown in Figures B9.3a, B9.3c and B9.3e



Uncertainty and the value of the gastro-intestinal disease outcome elasticity (Figures B9.4a-B9.4e)

Figure B9.4a plots 1000 values from the distribution of the point estimate for the gastro-intestinal disease outcome elasticity (see column 7 of Table B8.19). The mean of these sampled values is -1.518 (the outcome elasticity in the comparable IV model is -1.536) and all of these values lie between -3.0 and 0.0.

Figure B9.4a: graph showing histogram for the sampled gastro-intestinal disease outcome elasticity



The histogram in Figure B9.4b provides a plot of 1000 point estimates for the respiratory disease outcome elasticity if we drop the assumption that the coefficients on the excluded instruments are exactly zero (and we also assume a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). The mean value of these 1000 outcome elasticities (-2.365) is 50% larger than the size of the elasticity in the comparable IV model (-1.536). And the mean value of the standard errors associated with these 1000 elasticities (0.853) is about 80% larger than the standard error in the basic IV model (0.468).

The histogram in Figure B9.4c provides a similar plot to that in Figure B9.4b but this time we assume a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5). However, there is very little difference between the zero (Figure B9.4b) and non-zero (B9.4c) covariance plots, with both the mean elasticity and mean standard error virtually identical in these plots.

The histograms in Figures B9.4b and B9.4c provide plots of point estimates for the gastro-intestinal problems outcome elasticity if we drop the assumption that the coefficients on the excluded instruments are exactly zero. Each point estimate also has a standard error and we can sample from these estimates and their distributions to obtain the histograms shown in Figures B9.4d and B9.4e. With the exception of a slight extension to both tails, these plots are similar to the plots in Figures B9.4b and B9.4c.

Figure B9.4b: graph showing histogram for estimated outcome elasticity associated with gastro-intestinal outcome model (zero covariance between the coefficients on the excluded instruments)

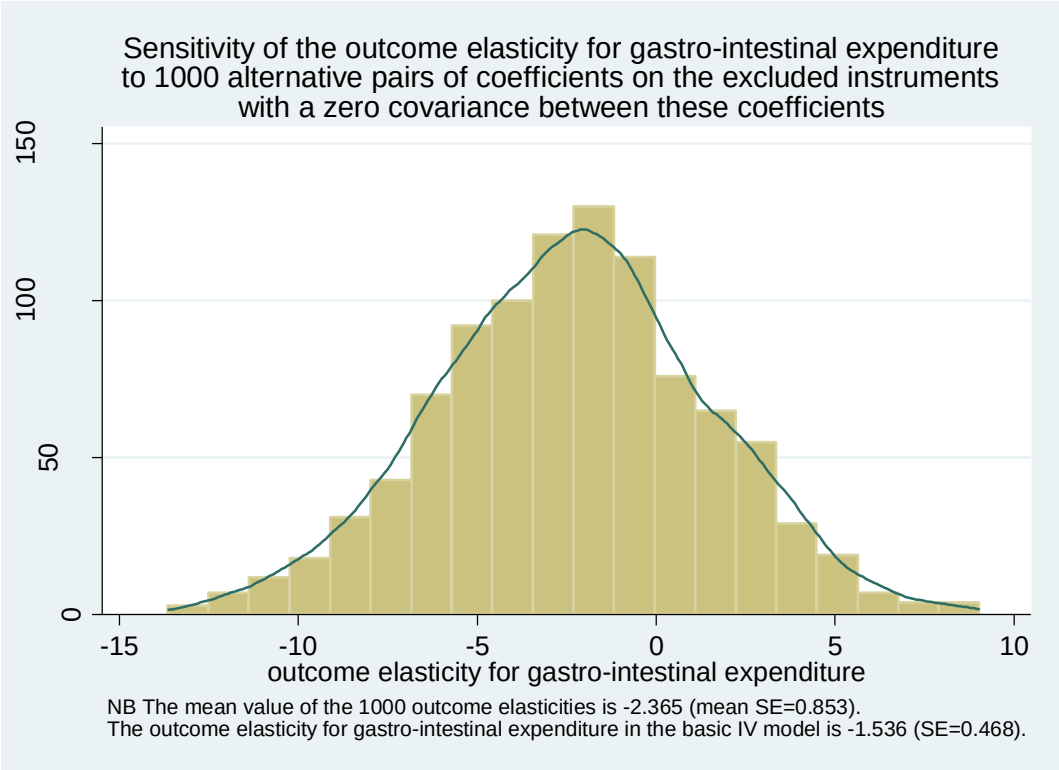


Figure B9.4c: graph showing histogram for estimated outcome elasticity associated with gastro-intestinal outcome model (non-zero covariance between the coefficients on the excluded instruments)

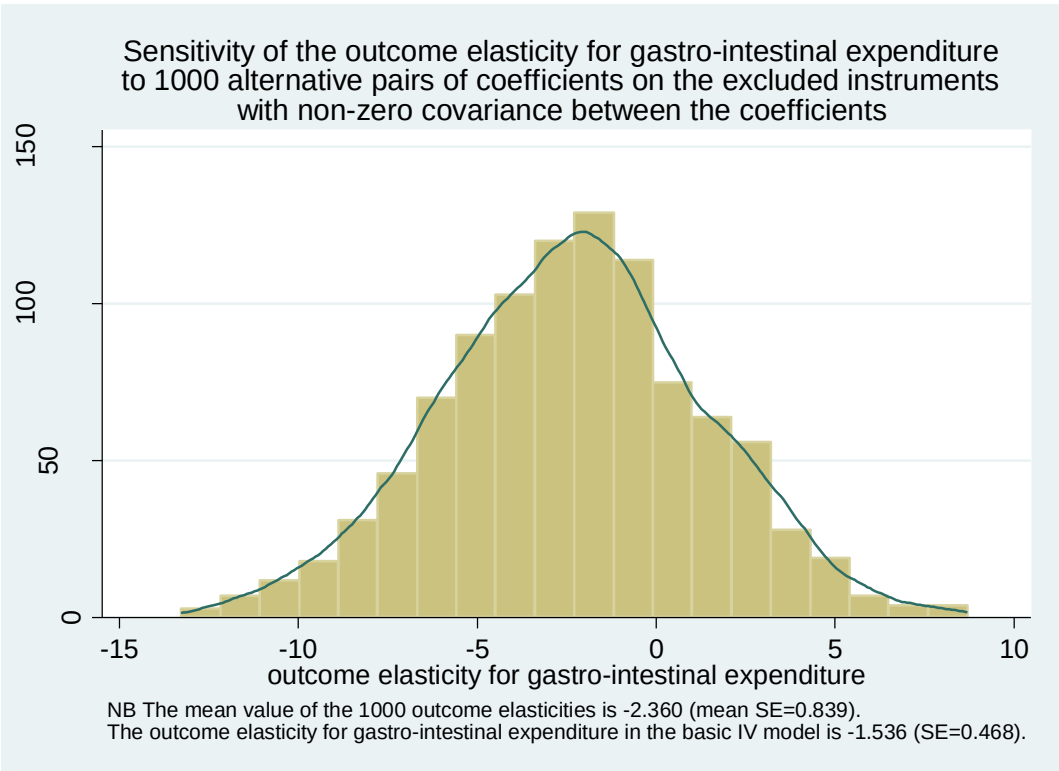


Figure B9.4d: graph showing histogram for sampled values of the gastro-intestinal problems outcome elasticity (zero covariance between the coefficients on the excluded instruments)

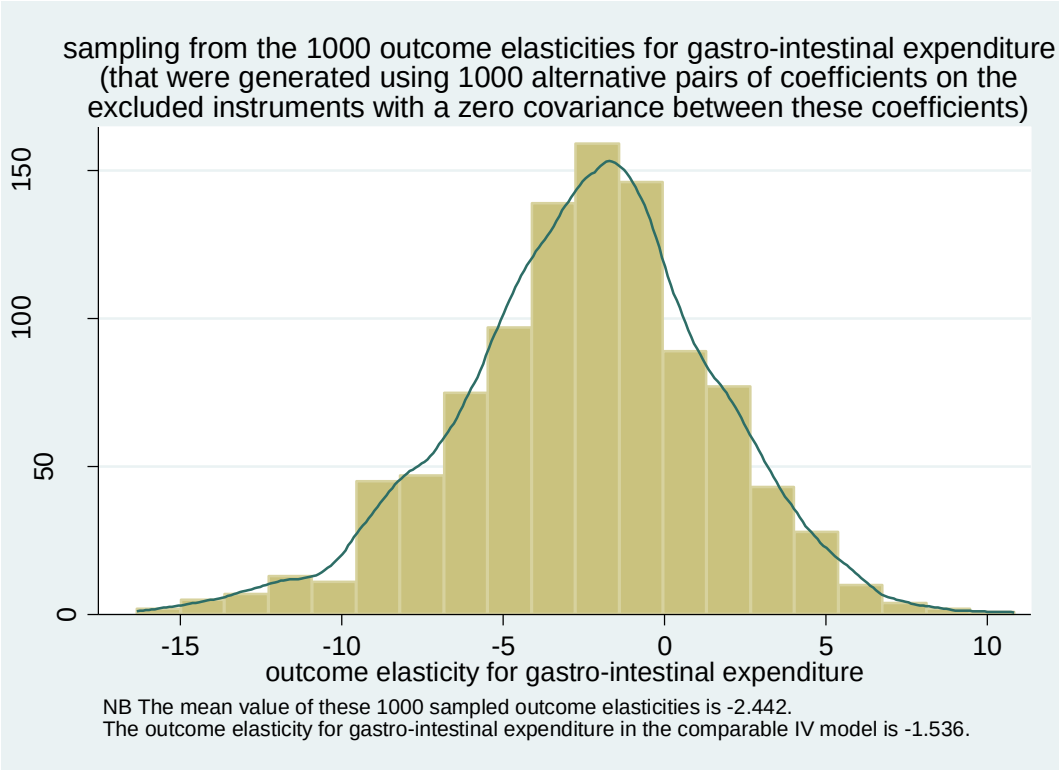


Figure B9.4e: graph showing histogram for sampled values of the gastro-intestinal problems outcome elasticity (non-zero covariance between the coefficients on the excluded instruments)

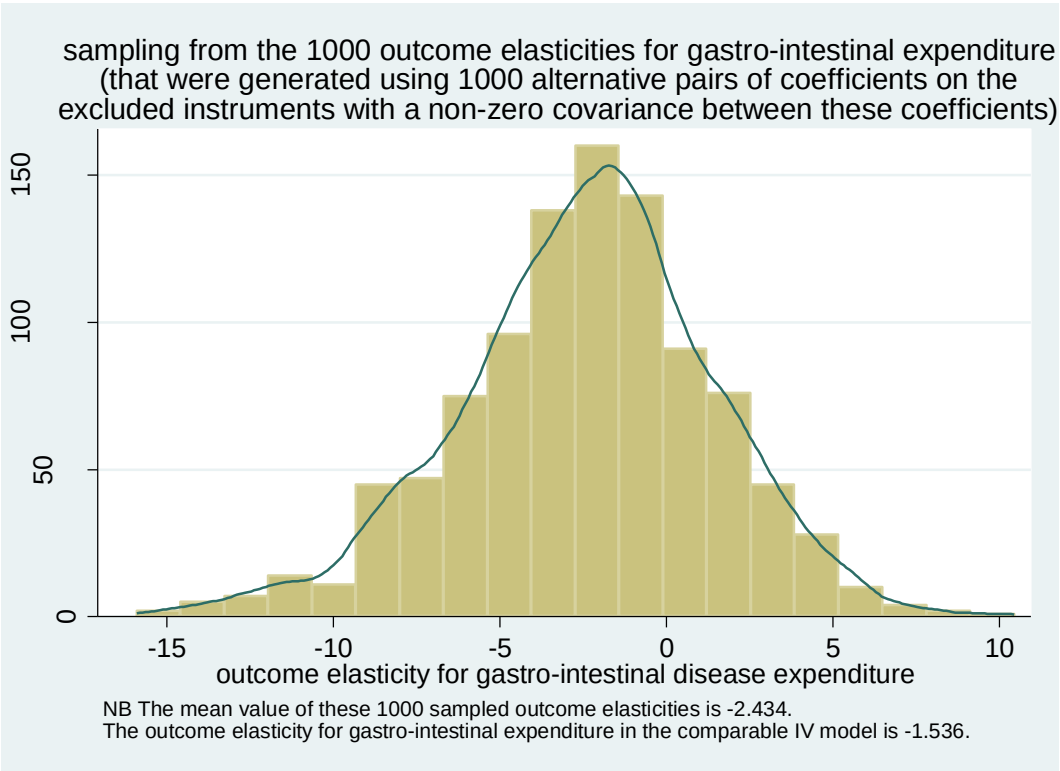


Figure B9.4f reproduces the three kernel density plots shown in Figures B9.4a, B9.4b and B9.4d (remember that Figures B9.4b and B9.4d assume a zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4)

and (9.5)). These plots illustrate the impact of all three levels of uncertainty on our estimate of the gastro-intestinal outcome elasticity. It is clear that the uncertainty induced by the instrument validity issue both shifts the density plot to the left slightly and dramatically increases the uncertainty associated with our estimate of the outcome elasticity (e.g., the range of values increases from -3 to 0 at level 1, from -13 to 9 at level 2, and then further from -16 to 11 at level 3).

Figure B9.4f: graph showing comparing the three kernel density plots shown in Figures B9.4a, B9.4b and B9.4d

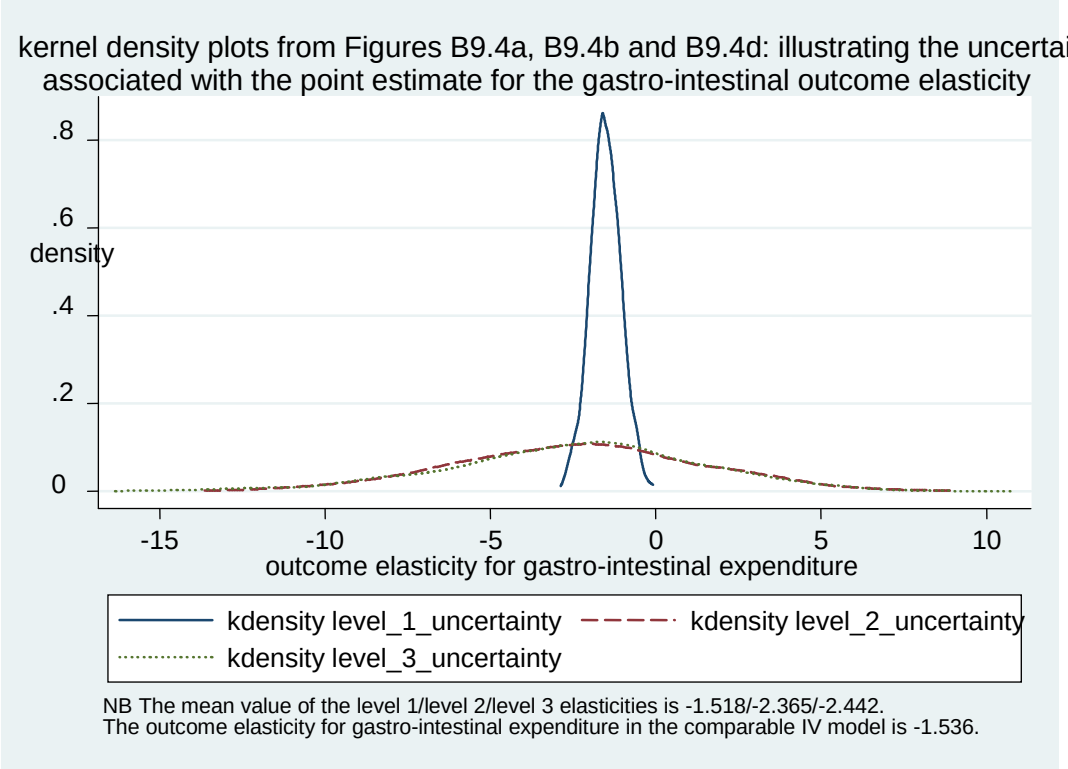
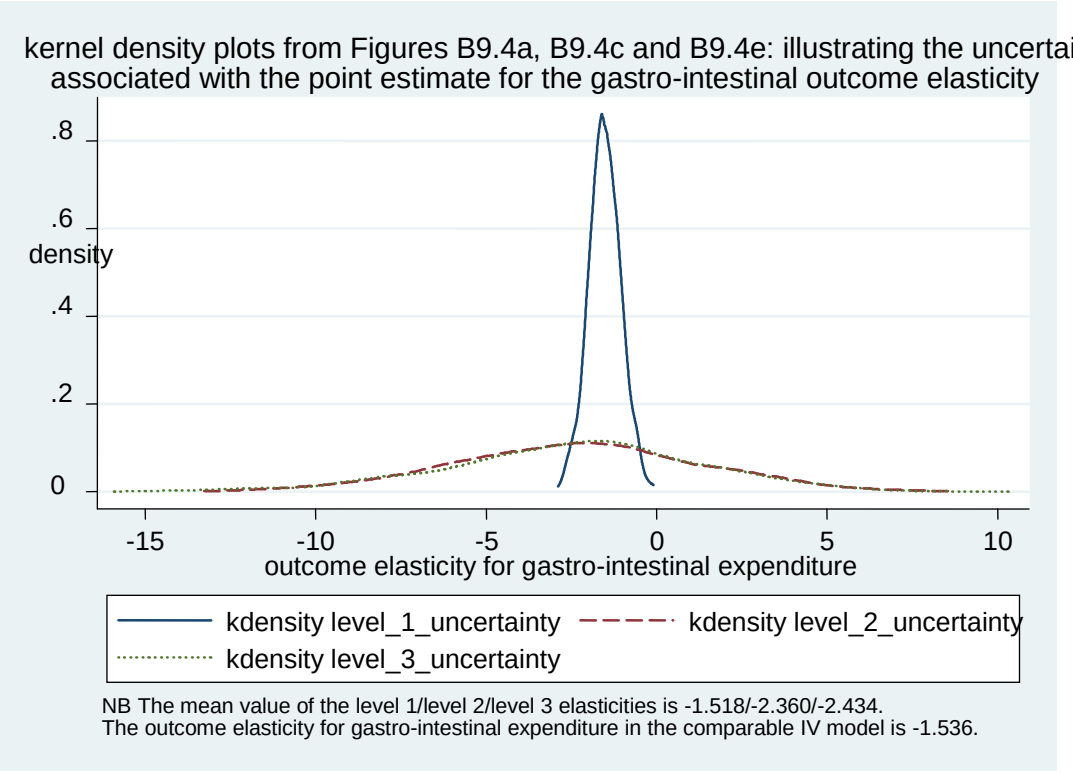


Figure B9.4g reproduces the three kernel density plots shown in Figures B9.4a, B9.4c and B9.4e (remember that Figures B9.4c and B9.4e assume a non-zero covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in equations (9.4) and (9.5)). As is the case for Figure B9.4f, these plots illustrate the impact of all three levels of uncertainty on our estimate of the gastro-intestinal outcome elasticity. And again, it is clear that the uncertainty induced by the instrument validity issue both shifts the density plot to the left slightly and considerably increases the uncertainty (range) associated with our estimate of the outcome elasticity. More precisely, the standard deviation associated with the level 1 uncertainty density plot is 0.468 but the standard deviation for the level 2 (3.658) and level 3 (3.834) uncertainty density plots are both eight times larger than this.

Figure B9.4g: graph showing comparing the three kernel density plots shown in Figures B9.4a, B9.4c and B9.4e



B9.6 Implications of uncertainty for the estimate of the cost of a life year

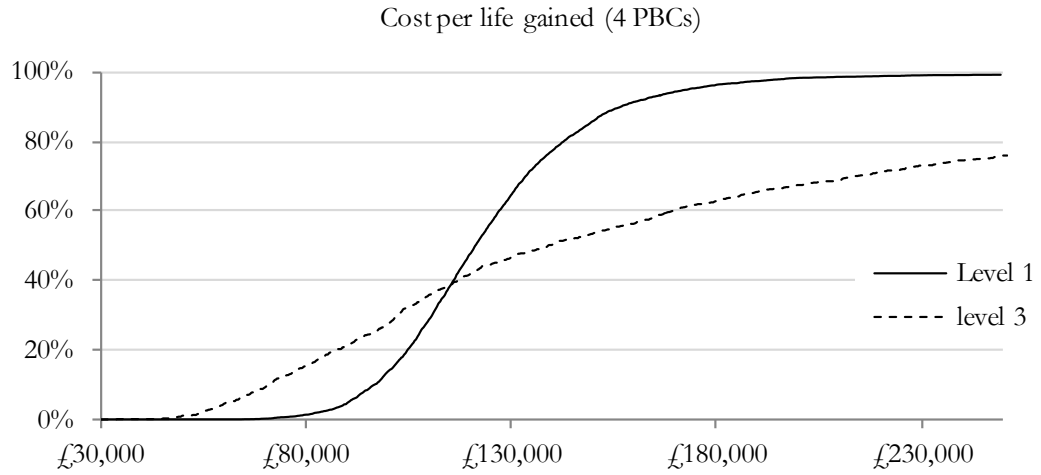
In the previous subsection, we have evaluated the outcome equation elasticities when uncertainty over the validity of instrument variables is considered ('level 3' uncertainty), in contrast to assuming the instruments are valid ('level 1' uncertainty). This analysis showed that including level 3 uncertainty affects the central value of the outcome elasticities; however, it is difficult to predict its effect on the expectation of the threshold given the impact of expenditure on mortality appears reduced in some programmes but increased in others. In Table B9.3, the mean estimates of the outcome elasticities under level 3 uncertainty were used to calculate the threshold for the big four programmes of health. The results show that relaxing the assumption of validity of the instruments has little impact on the expectation of the threshold for the big 4 PBCs [the cost per life year gained threshold changed from £10,604 (Table B8.22) to £11,009 in Table B9.3].

Table B9.3: Cost of life and life year estimates for the big four programmes using expenditure data for 2006 and outcome data for 2006/7/8 adjusted for the ICD10 coverage of the expenditure and outcome data

PBC description	Spend (£m) 2006/7	Spend elasticity	Change in spend (£m)	Annual mortality, <75years, 2006/08	Outcome elasticity (without negative sign)	Coverage of mortality data relative to spend data	Change in annual mortality adj for coverage	Cost per life gained (£) adj for coverage	Total life years lost, <75years, 2006/08	Coverage of mortality data relative to spend data	Change in annual life years lost adj for YLL	Cost per life year gained (£)	Cost per life year gained adj for YLL coverage (£)
1 Cancer	£4,122	0.465	£19.17	61,961	0.218	0.984	63.90	£299,975	2,207,021	0.984	759	£16,383	£25,265
2 Circulatory problems	£6,161	0.540	£33.27	41,106	1.718	0.992	384.42	£86,544	1,361,634	0.992	4,245	£9,466	£7,838
3 Respiratory problems	£3,285	0.679	£22.31	11,574	1.151	0.773	116.99	£190,666	324,223	0.773	1,092	£11,593	£20,419
4 Gastro-intestinal problems	£3,700	0.446	£16.50	6,160	2.434	0.571	117.11	£140,906	345,908	0.571	2,192	£20,892	£7,528
Big four programmes summary:													
5 Spend 2006 & mortality 2006/8	£17,268		£91.24	120,801			682.42	£133,707	4,238,786		8,288	£12,333	£11,009

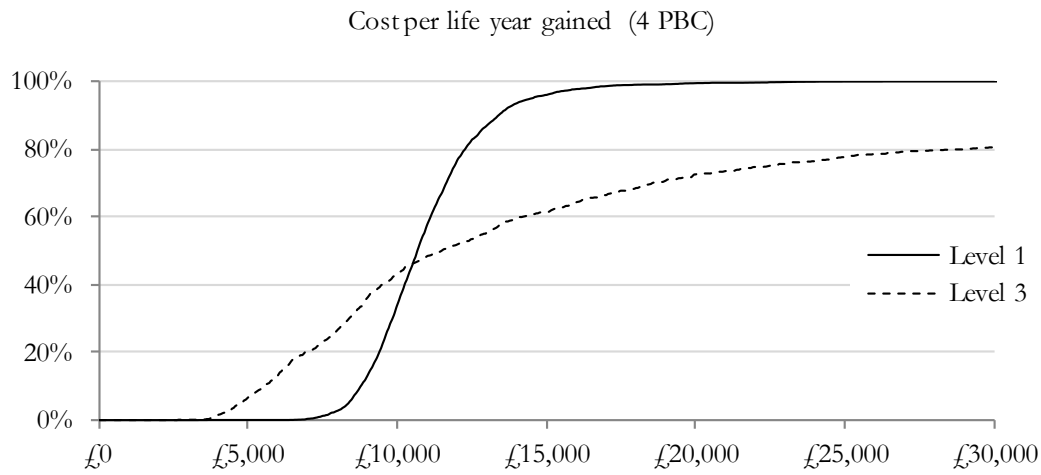
The assumption of validity of instruments is expected to affect significantly the level of uncertainty over the cost effectiveness threshold estimate. Illustrations of this source of uncertainty were presented in the previous section (B9.5) using empirical distributions derived from the sampling procedure implemented; these illustrations represent the uncertainty in the mean estimate for each of the elasticities. To characterise the effect of levels 1 and 3 uncertainty on the overall threshold we used the sets of simulated elasticities (one for each of the 4 programmes of care) to compute a threshold value; in doing so for all simulated sets, a sample of threshold values was obtained. In this way, uncertainty was propagated from the outcome elasticities to the threshold estimates, and an empiric distribution describing uncertainty over threshold estimates obtained. The cumulative density function can be used to display such uncertainty; this plots the probability (y-axis) of the threshold being below certain values (x-axis) in the simulated sample (this corresponds to a Bayesian interpretation of uncertainty). Figure B9.5a plots the cumulative density curve for the cost per life gained threshold when level 1 and level 3 uncertainty are considered in turn, and B9.5b for the cost per life year gained threshold.

Figure B9.5a Cumulative density plot for the cost per life gained threshold for the big 4 PBCs (considers covariance between the coefficients on the excluded instruments).



In drawing the cumulative density function, negative threshold values were dealt with by evaluating whether it was the health component or the cost component that was negative. For simulations where health change was negative (0% were observed for both levels 1 and 3), the threshold was left as a negative value. Simulations showing a negative change in spend were assigned a very high positive threshold value – in this was an asymptote is generated in the plot (respectively 0% and 5.6% were observed for levels 1 and 3).

Figure B9.5b Cumulative density plot for the cost per life year gained threshold for the big 4 PBCs (considers covariance between the coefficients on the excluded instruments)



In drawing the cumulative density function, negative threshold values were dealt with by evaluating whether it was the health component or the cost component that was negative. For simulations where health change was negative (0% were observed for both levels 1 and 3), the threshold was left as a negative value. Simulations showing a negative change in spend were assigned a very high positive threshold value – in this was an asymptote is generated in the plot (respectively 0.04% and 7.7% were observed for levels 1 and 3).

The probability that the overall threshold is less than £7,500 per life year is around 0.2 when uncertainty over the validity of instruments in considered (level 3), whereas when the instruments are assumed valid (level 1) this probability is 0. Under level 1 uncertainty, we would be confident that the threshold is less than £30,000 (probability of 1), but when considering level 3 uncertainty there is some chance that the threshold is higher than £30,000 (probability of 0.2). These plots show that uncertainty on the validity of the instruments generates significant uncertainty over the threshold value.

B9.7 Summary and conclusion

One of the crucial elements in the calculation of the cost of a life year for any care programme is the coefficient on the expenditure variable in the outcome equation. The endogenous nature of expenditure in our model means that OLS estimation is inappropriate and that instead IV techniques must be used. The application of these techniques requires the identification of variables that are good predictors of the endogenous variable (expenditure) but which do not have a direct effect on the dependent variable (mortality).

It is difficult to provide theoretical arguments why our selected instruments will not affect mortality directly. Instead, we rely on the widely used Hansen-Sargen test of instrument validity. Although our models 'pass' this test, some commentators have argued that this test has weak power and may fail to reject the null hypothesis of instrument validity even when an exclusion restriction is not valid. Given our reliance on this test, we noted that this test will only lack power if the biases induced in the coefficient on the endogenous variable by invalid instruments all coincide (i.e., the instruments all identify the *same* wrong parameter). However, it is far from obvious that this will be so in this case, particularly given that our outcome equation already includes a measure of the need for health care.

Nevertheless, it is possible that our instruments are correlated with both expenditure and some unobserved factor which is directly influencing the mortality rate, and that the induced bias in $\widehat{\beta 1}$ is the same for both instruments.

We therefore undertook an extensive sensitivity analysis of the estimated outcome elasticity to the validity of the exclusion restrictions. In summary, we found that both the central value and distribution of the outcome elasticity may change if we drop the assumption that the coefficients on the excluded instruments are identically zero.

This change in the central value of the outcome elasticity reduces the impact of expenditure on mortality in some programmes (e.g., for cancer the 'average' outcome elasticity falls from -0.338 to -0.210, and for respiratory disease it falls from -2.004 to -1.151). However, in other programmes this change in the central value increases the impact of expenditure on mortality (e.g., for circulatory disease the 'average' outcome elasticity increases from -1.418 to -1.718, and for gastro-intestinal problems it increases from -1.518 to -2.434).

However, in all four programmes the standard deviation associated with the distribution of the value for the outcome elasticity increased: for cancer it increased from 0.099 to 0.379; for circulatory disease it increased from 0.228 to 0.843; for respiratory disease it increased from 0.636 to 1.098; and for gastro-intestinal disease it increased from 0.468 to 3.834.

B10 Analysis of programme budgeting expenditure for 2007/8 and mortality data for 2007/2009

Outcome and expenditure models were estimated using updated data for expenditure (from 2006/7 to 2007/8) and updated mortality data (from 2006/2007/2008 to 2007/2008 /2009). Results for the outcome model are shown in Table B10.1 and results for the expenditure model are in Table B10.2. First stage regressions for these IV models can be found in Tables BA.7 and BA.8 in the annex.

B10.1 Outcome models

Some of the outcome models in Table B10.1 contain just two variables: own programme expenditure and a measure of the need for health care. The latter is usually the measure of need as employed by the Department of Health for resource allocation purposes and this incorporates the CARAN formula for acute services. For the respiratory programme we have added the square of this need measure to improve the model fit. In other PBCs we found that the all service measure of need performed poorly and we have replaced or supplemented it with either a more programme specific measure (e.g., the epilepsy prevalence rate for neurological mortality) or with a better performing proxy for need (e.g., the percentage of residents born outside the EU for maternity/neonate mortality).

Two results are reported for three of the big four programmes. One of these two results uses two instruments and so we report the instrument validity test statistic. We cannot reject the null hypothesis of instrument validity in all three cases. However, there is some evidence of weak instruments (at least in the respiratory and gastro-intestinal programmes) but, if we drop one instrument and re-estimate the model, the evidence of instrument weakness disappears (but of course there is no instrument validity test statistic with this re-estimation). The removal of one instrument has little impact on the coefficient on expenditure and it is this coefficient from this one instrument model that we use below in our cost of a life year calculations.

The first seven results in Table B10.1 show the outcome model for the big four programmes (i.e., for cancer, circulatory disease, respiratory problems and gastro-intestinal problems). In all four programmes the need variable has a positive and significant effect on mortality, and expenditure has the anticipated negative effect. The diagnostic statistics reveal that, in all four PBCs, own programme expenditure is endogenous and that the instruments are valid. They also suggest that the instruments are relevant and there is no evidence that the instruments are weak in the models with one excluded instrument. The Pesaran-Taylor test suggests that there is no evidence of model mis-specification.

The outcome results for the other programmes (in columns 8 - 13) are similar to but more diverse than those for the big four programmes. This is to be anticipated because mortality is a much rarer outcome in these programmes than it is in the big four programmes. Own programme expenditure is not endogenous in four of these programmes but we retain the IV estimator for three of these four because this yields more plausible results than the OLS estimator (the results are more plausible in the sense that the signs on the coefficients are more in line with our prior expectations).

Expenditure has the anticipated negative effect on mortality in the endocrine problems programme but this is not statistically significant. The all service measure of need is not relevant for this PBC; instead, we find that the diabetes prevalence rate is positively associated with mortality, as is a measure of deprivation (the IMD2007).

Mortality from epilepsy is negatively and significantly associated with expenditure in the neurological programme. Both the all service need for health care and the epilepsy prevalence rate are positively and significantly associated with mortality in this programme.

Expenditure has a negative and statistically significant effect on mortality (from renal problems) in the genitor-urinary problems programme. The prevalence of lone parent households is positively associated with mortality.

Expenditure has the anticipated negative effect on mortality in the infectious disease programme and this is statistically significant. The all service measure of need is not relevant for this PBC; instead, we find that a measure of need associated with HIV is positively associated with mortality, as is a measure of deprivation (the IMD2007).

Expenditure has the anticipated negative effect on mortality in the maternity & neonates programme but the estimated coefficient is not statistically significant. In this PBC the generic all service measure of need has been replaced with two other indicators of deprivation – the proportion of residents born outside the EU and the proportion of those aged 16-74 without any qualifications – and both of these are positively associated with mortality.

Finally, expenditure and need have the anticipated effects on mortality in the trauma and injuries programme. In addition, the proportion of households without access to a car is negatively associated with mortality from fractures (perhaps access to a car facilitates involvement in serious road traffic accidents), and the proportion of residents that are students is positively associated with mortality from fractures.

The relevant statistical test suggests that expenditure is endogenous in six of the ten programmes but we have retained the IV estimates for three of the other four programmes because they provide plausible results. The Hansen-Sargen test suggests that the selected instruments are valid, and the Kleibergen-Paap LM statistic suggests that they are relevant (i.e., correlated with the endogenous regressor). With the possible exception of the trauma and injuries programme, the Kleibergen-Paap F statistic suggests that we do not have a problem with weak instruments.[35] Finally, the Pesaran-Taylor/Ramsey reset test statistics reveal no evidence of mis-specification.

B10.2 Expenditure models

Most of the expenditure models in Table B10.2 contain just three variables: the PCT budget, a proxy for the own programme need for health care, and a proxy for the need for health care in other programmes.

The budget term is positive in all eleven models and it is statistically significant in eight of these eleven models.

The usual proxy for the own programme need for health care (i.e., the all service measure of need) is present in six of the models and it is significant in five of them. Its presence is supplemented with the addition of its squared value to improve model fit in the respiratory problems programme.

In some programmes (e.g., the endocrine, metabolic & nutritional programme and the neurological programme), we have replaced and/or supplemented the all service measure of need with a more programme specific measure (e.g., the diabetes prevalence rate and the epilepsy prevalence rate) and these measures of need have the anticipated positive impact on expenditure.

In addition, in a couple of other programmes we have used alternative proxies for the own programme need (e.g., with the use of the Department of Health's measure of maternity need in the maternity/neonates expenditure equation).

For eight of the eleven programmes we have used the all cause mortality rate less the own programme mortality rate as the proxy for the need for health care in other programmes, and the coefficient on this term is negative in seven programmes and statistically significant in six of the seven. In three programmes -- maternity/neonates, GMS/PMS and trauma & injuries programmes -- we have used the all cause mortality rate as the proxy for the need for health care in other programmes due to difficulties associated with the measurement of the own programme mortality rate. The coefficient on this term is not significant in any of the three models.

[35] The Kleibergen-Paap F statistic is very close to the target value of ten for both the genitor-urinary and infectious diseases outcome models.

The relevant statistical test suggests that expenditure is endogenous in six of the eleven programmes but we have retained the IV estimates for two other programmes (GMS/PMS and trauma & injuries) because the IV estimator provides more plausible results. In the other three programmes we report OLS results.

The Hansen-Sargen test suggests that the selected instruments are valid, and the Kleibergen-Paap LM statistic suggests that they are relevant (i.e., correlated with the endogenous regressor). The Kleibergen-Paap F statistic suggests that we do not have a problem with weak instruments. Finally, the Pesaran-Taylor reset test statistics and the Ramsey reset F statistics reveal no evidence of model mis-specification.

B10.3 Calculation of the cost of a life and life year

Expenditure and outcome elasticities for our preferred models are shown in Table B10.3 (see columns D and H) and these are used to calculate the cost of a life and the cost of a life year, both for individual programmes and for all programmes collectively.

Column N reports the cost per life gained and column U reports the cost per life year gained. From the latter we can see that the cost per life year gained is £13,830 for the big four programmes and £28,983 for all ten programmes with a mortality-based outcome indicator. These represent 30% and 45% increases on the respective costs for the previous year (i.e., using expenditure data for 2006/7 and mortality data for 2006/2007/2008).

If we assume that the other 13 programmes (all without a mortality based outcome indicator) offer no health gain, then the cost per life year across all PCT expenditure is £82,765. This is up from £73,457 using data for the previous year (an increase of 13%).

In addition, Table B10.4 shows that if we assume that PBC23 generates a zero health gain and that the gain attributable to the remaining 12 programmes is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes is £31,846 (it was £22,565 using data for the previous year).

B10.4 Summary and conclusion

In this section we have estimated outcome and expenditure models using PB data for 2007/8 and mortality data for 2007/8/9. The cost of an additional life year for all ten programmes with a mortality based outcome is £28,983. This is a 45% increase on the cost (£19,965) for the previous year (i.e., using expenditure data for 2006/7 and mortality data for 2006/2007/2008). The next section presents outcome and expenditure models using PB data for 2008/9 and mortality data for 2008/9/10, and it explores the reasons for the increase in the cost of an additional life year identified in this section.

Table B10.1: table showing outcome models using spend data for 2007/8 (two MFFs) and mortality data for 2007/8/9

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	PBC 2	PBC 2	PBC 10	PBC 11	PBC 11	PBC 13	PBC 13	PBC 4	PBC 7	PBC 17	PBC 1	PBC 1819	PBC 16
	cancer	cancer	circulation	respiratory	respiratory	gastro-intestinal	gastro-intestinal	endocrine	neurological	genito-urinary	infectious diseases	maternity & neonates	trauma & injuries
	2007/8	2007/8	2007/8	2007/8	2007/8	2007/8	2007/8	2007/8	2007/8	2007/8	2007/8	2007/8	2007/8
	outcome model instrument spend	outcome model instrument spend	outcome model instrument spend	outcome model instrument spend	outcome model instrument spend	outcome model instrument spend	outcome model instrument spend	outcome model instrument spend	outcome model o/need exogenous	outcome model instrument spend	outcome model instrument spend	outcome model instrument spend	outcome model instrument spend
	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted
VARIABLES	second stage	second stage	second stage	second stage	second stage	second stage	second stage	second stage	OLS	second stage	second stage	second stage	second stage
own programme spend per head	-0.365***	-0.365***	-1.277***	-2.205***	-2.211***	-1.292***	-1.328**	-0.566	-0.339**	-1.898**	-0.546*	-0.110	-0.369
	[0.106]	[0.107]	[0.206]	[0.705]	[0.739]	[0.497]	[0.519]	[0.550]	[0.144]	[0.921]	[0.300]	[0.139]	[0.353]
need CARAN per head	0.984***	0.985***	2.818***	5.119***	5.113***	3.908***	3.947***		0.853**				3.029***
	[0.108]	[0.110]	[0.256]	[1.052]	[1.105]	[0.633]	[0.658]		[0.344]				[0.717]
need CARAN per head squared				4.085**	3.982**								
				[1.721]	[1.774]								
IMD 2007								0.517***			0.481***		
								[0.109]			[0.098]		
diabetes prevalence rate 2007/8								0.820**					
								[0.359]					
epilepsy prevalence rate 2007/8									0.652***				
									[0.231]				
lone parent households										1.767***			
										[0.430]			
HIV need per head squared											0.143**		
											[0.064]		
HIV need per head											0.487***		
											[0.120]		
born outside the EU												0.152***	
												[0.028]	
no qualifications aged 16 to 74												0.990***	
												[0.115]	
no car households													-0.658***
													[0.221]
full-time students													0.528***
													[0.128]
Constant	6.635***	6.637***	10.643***	12.244***	12.269***	8.688***	8.845***	0.512	3.072***	12.110**	2.176***	3.303***	2.654**
	[0.480]	[0.483]	[0.996]	[2.947]	[3.090]	[2.142]	[2.237]	[1.349]	[0.614]	[4.852]	[0.675]	[0.762]	[1.346]
Observations	151	151	151	151	151	151	151	151	151	147	151	151	151
Endogeneity test statistic	17.288	16.323	39.948	21.368	28.333	18.871	17.769	1.293		3.916	3.603	0.551	1.375
Endogeneity p-value	3.21e-05	5.34e-05	1.42e-05	3.79e-06	1.02e-07	1.40e-05	2.49e-05	0.255		0.0478	0.0577	0.458	0.241
Hansen-Sargan test statistic	0.00124	n/a	0.056	n/a	0.163	0.120	n/a	n/a		6.710	0.583	0.675	5.001
Hansen-Sargan p-value	0.972		0.814		0.686	0.729				0.0349	0.747	0.411	0.0820
Shea's partial R-squared	0.162	0.162	0.323	0.0832	0.0977	0.126	0.112	0.133		0.160	0.104	0.201	0.137
Kleibergen-Paap LM test statistic	19.52	19.44	20.71	8.807	8.840	10.76	10.53	20.71		20.01	16.45	30.58	16.82
Kleibergen-Paap p-value	5.76e-05	1.04e-05	0.0000	0.00300	0.0120	0.00462	0.00117	5.36e-06		0.000169	0.000917	2.29e-07	0.000770
Kleibergen-Paap F statistic	14.50	29.13	34.54	12.26	6.533	7.809	14.70	25.56		9.624	9.688	23.31	7.835
Pesaran-Taylor/Ramsey test statisti	0.00606	0.0115	2.06	2.839	2.850	0.418	0.106	0.00725	0.469	0.393	2.251	0.00684	0.0128
Pesaran-Taylor/Ramsey p-value	0.938	0.915	0.1515	0.0920	0.0914	0.518	0.744	0.932	0.704	0.531	0.134	0.934	0.910

Note: (i) robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1;
(ii) the addition of unpaid carers as an instrument for the endocrine outcome model generates a Hansen-Sargen test statistic of 0.372 (p-value 0.5418) and the coefficient on expenditure is -0.423.

Table B10.2: table showing expenditure models using spend data for 2007/8 (two MFFs) and mortality data for 2007/8/9

VARIABLES	(1) PBC 2 cancer 2007/8 spend model second stage	(2) PBC 10 circulation 2007/8 spend model second stage	(3) PBC 11 respiratory 2007/8 spend model second stage	(4) PBC 13 gastro-intestinal 2007/8 spend model second stage	(5) PBC 1 infectious disease 2007/8 spend model OLS	(6) PBC 4 endocrine 2007/8 spend model second stage	(7) PBC 7 neurological 2007/8 spend model second stage	(8) PBC 17 genito-urinary 2007/8 spend model OLS	(9) PBC 1819 maternity/ neonates 2007/8 spend model OLS	(10) PBC 23 GMS/PMS etc 2007/8 spend model second stage	(11) PBC 16 trauma/injuries 2007/8 spend model second stage
all cause SYLLR excluding cancer	-1.227***										
	[0.220]										
PCT budget per head	0.890**	0.293	0.536*	0.622*	1.435***	0.264	1.036***	1.004***	0.514*	0.563	1.686***
	[0.431]	[0.350]	[0.298]	[0.321]	[0.258]	[0.206]	[0.307]	[0.356]	[0.264]	[0.344]	[0.384]
need CARAN per head	1.659***	3.117***	1.786***	1.982***		0.925***		0.029			
	[0.430]	[0.535]	[0.334]	[0.422]		[0.305]		[0.371]			
all cause SYLLR exc circulatory		-2.115***									
		[0.397]									
all cause SYLLR exc respiratory			-0.781***								
			[0.236]								
need CARAN per head squared			1.687***								
			[0.446]								
all cause SYLLR exc gastro				-1.279***							
				[0.333]							
HIV need per head					0.440***						
					[0.025]						
all cause SYLLR exc infect diseases					-0.543**						
					[0.249]						
HIV need per head squared					0.183***						
					[0.021]						
all cause SYLLR exc diabetes						-0.384*					
						[0.218]					
diabetes prevalence rate 2007/8						0.332***					
						[0.123]					
all cause SYLLR exc epilepsy							-0.259				
							[0.223]				
epilepsy prevalence rate 2007/8							0.571***				
							[0.072]				
all cause SYLLR exc renal								-0.072			
								[0.168]			
maternity need per head									0.582***		
									[0.098]		
all cause SYLLR									0.286	-0.169	-0.277
									[0.193]	[0.290]	[0.363]
lone pensioner households										-0.480***	
										[0.182]	
population working in agriculture											0.132***
											[0.022]
Constant	4.973	15.081***	4.986**	7.488***	-4.212***	3.555*	-1.684	-2.675	-1.222	1.413	-5.960***
	[3.047]	[3.303]	[2.342]	[2.786]	[1.034]	[1.817]	[1.130]	[2.562]	[1.388]	[1.373]	[1.104]
Observations	151	151	151	151	151	151	151	151	151	151	151
Endogeneity test statistic	20.985	19.454	11.612	15.477		2.846	4.958			0.060	1.769
Endogeneity p-value	4.63e-06	1.03e-05	0.000655	8.35e-05		0.0916	0.0260			0.807	0.183
Hansen-Sargan test statistic	0.411	0.003	1.369	0.0201		0.510	2.748			1.091	1.121
Hansen-Sargan p-value	0.522	0.959	0.504	0.887		0.775	0.0974			0.296	0.571
Shea's partial R-squared	0.384	0.253	0.398	0.325		0.402	0.518			0.416	0.364
Kleibergen-Paap LM test statistic	40.04	28.14	39.41	33.23		40.29	31.53			16.51	27.19
Kleibergen-Paap p-value	2.02e-09	7.76e-07	1.42e-08	6.09e-08		9.26e-09	1.42e-07			0.000260	5.37e-06
Kleibergen-Paap F statistic	51.44	29.097	40.69	20.04		37.14	73.21			26.60	32.54
Pesaran-Taylor/Ramsey test statisti	2.262	0.0002	0.0236	0.0341	0.721	2.351	0.619	1.297	1.018	1.757	0.193
Pesaran-Taylor/Ramsey p-value	0.133	0.988	0.878	0.854	0.541	0.125	0.432	0.278	0.387	0.185	0.660

Note: robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table B10.3: table showing cost of life and life year estimates using spend data for 2007/8 and outcome data for 2007/2008/2009 (assumes zero health gain for 13 programmes)

A	B	C	D	E =0.01*C*D	F	G	H	I =0.01*D*G*H	J	K =E/I	L	M =I/L	N =E/M	O	P	Q =0.01*D*H*P/3	R	S =Q/R	T =E/Q	U =E/S
	PBC description	Spend (£m) 2007/8	Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2007/09	Outcome elasticity (without negative sign)	Change in annual mortality		Cost per life gained (£)	Coverage of mortality data relative to spend data	Change in annual mortality adj for coverage	Cost per life gained (£) adj for coverage		Total life years lost, <75years, 2007/09	Change in annual life years lost	Coverage of mortality data relative to spend data	Change in annual life years lost adj for coverage	Cost per life year gained (£)	Cost per life year gained adj for coverage (£)
1	Cancer	£4,573	0.890	£40.70		61,960	0.365	201.28		£202,207	0.984	204.55	£198,972		2,189,685	2,371	0.984	2,410	£17,165	£16,891
2	Circulatory problems	£6,325	0.293	£18.53		39,304	1.277	147.06		£126,018	0.992	148.25	£125,010		1,313,223	1,638	0.992	1,651	£11,315	£11,224
3	Respiratory problems	£3,431	0.536	£18.39		10,764	2.205	127.22		£144,557	0.773	164.58	£111,742		315,457	1,243	0.773	1,608	£14,798	£11,439
4	Gastro-intestinal problems	£3,805	0.622	£23.67		6,031	1.328	49.82		£475,081	0.571	87.25	£271,271		343,355	945	0.571	1,656	£25,034	£14,295
	Big four programmes summary:																			
5	Spend 2007 & mortality 2007/9	£18,134		£101.29		118,059		525.37		£192,795		604.62	£167,526		4,161,720	6,197		7,324	£16,345	£13,830
6	Spend 2006 & mortality 2006/8	£17,268		£91.24		120,801		665.10		£137,188		761.49	£119,823		4,238,786	7,399		8,604	£12,333	£10,604
7	Spend 2006 & mortality 2004/6	£17,268		£114.04		125,290		953.13		£119,650					4,335,559	10,576			£10,783	
8	Spend 2005 & mortality 2002/4	£17,625		£141.22		125,290		909.96		£155,196					4,516,953	10,986			£12,855	
9	Infectious diseases	£1,119	1.436	£16.07		1,977	0.548	15.56		£1,032,863	1.000	15.56	£1,032,863		106,092	278	1.000	278	£57,742	£57,742
10	Endocrine problems	£1,997	0.264	£5.27		1,471	0.566	2.20		£2,398,551	0.634	3.47	£1,520,681		55,492	28	0.634	44	£190,745	£120,932
11	Neurological problems	£3,165	1.035	£32.76		718	0.339	2.52		£13,003,180	0.136	18.52	£1,768,432		64,873	76	0.136	558	£431,749	£58,718
12	Genito-urinary problems	£3,439	1.004	£34.53		270	1.855	5.03		£6,866,327	0.172	29.24	£1,181,008		8,529	53	0.172	308	£652,096	£112,160
13	Trauma & injuries*	£2,918	1.686	£49.20		1,013	0.369	6.30		£7,806,376	0.175	36.01	£1,366,116		21,273	44	0.175	252	£1,115,197	£195,159
14	Maternity & neonates*	£3,662	0.514	£18.82		2,199	0.11	1.24		£15,139,113	8.213	0.15	£124,337,534		489,170	92	0.679	136	£204,168	£138,630
	Other six programmes summary:																			
15	Spend 2007 & mortality 2007/9	£16,300		£156.65		7,648		32.85		£4,768,699		102.95	£1,521,610		745,429	571		1,575	£274,309	£99,428
16	Spend 2006 & mortality 2006/8	£15,643		£93.29		7,839		16.14		£5,780,723		25.05	£3,724,129		762,991	362		639	£258,046	£146,108
17	Spend 2006 & mortality 2004/6	£15,643		£112.13		7,923		18.17		£6,172,491					757,531	249			£449,706	
18	Spend 2005 & mortality 2002/4	£12,743		£99.44		7,923		16.26		£6,115,621					751,009	337			£295,074	
	All ten programmes summary:																			
19	Spend 2007 & mortality 2007/9	£34,434		£257.94		125,707		558.22		£462,067		707.57	£364,540		4,907,149	6,768		8,900	£38,110	£28,983
20	Spend 2006 & mortality 2006/8	£32,911		£184.53		128,640		681.24		£270,881		786.54	£234,617		5,001,777	7,760		9,243	£23,780	£19,965
21	Spend 2006 & mortality 2004/6	£32,911		£226.18		133,213		971.30		£232,861					5,093,090	10,826			£20,893	
22	Spend 2005 & mortality 2002/4	£30,368		£240.67		133,213		926.22		£259,838					5,267,962	11,322			£21,256	
	Assume zero health gain in the other 13 programmes																			
	Other 13 programmes summary:																			
23	Spend 2007 & mortality 2007/9	£39,223		£478.63				0.00				0.00				0.00		0.00		
24	Spend 2006 & mortality 2006/8	£34,985		£494.43				0.00								0.00		0.00		
25	Spend 2006 & mortality 2004/6	£34,985		£452.78				0.00								0.00				
26	Spend 2005 & mortality 2002/4	£33,942		£402.43				0.00								0.00				
	All 23 programmes																			
27	Spend 2007 & mortality 2007/9	£73,657		£736.57				558.22		£1,319,496		707.57	£1,040,992			6,768		8,900	£108,829	£82,765
28	Spend 2006 & mortality 2006/8	£67,896		£678.96				681.24		£996,655		786.54	£863,228			7,760		9,243	£87,494	£73,457
29	Spend 2006 & mortality 2004/6	£67,896		£678.96				971.30		£699,024						10,826			£62,718	
30	Spend 2005 & mortality 2002/4	£64,310		£643.10				926.22		£694,330						11,322			£56,799	
	Note:	2007/8	2006/7	2005/6		Note that the YLL for maternity and neonates is estimated as [(6,456 neonate deaths*75years)+(142 maternal deaths*35years)]. This totals 489,170 life years														
31	All 23 programme spend	£73,657	£67,896	£64,310		Note that the annual mortality figures reported in cells G7 & G8 and G17 & G18 are identical because we do not have mortality data for 2002/04.														
32	% change in budget	1.00	1.00	1.00		Note that, for expenditure in 2007/8, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.														
33	proportionate change	0.01	0.01	0.01		Note that the adjustment for the coverage of the mortality & YLL data relative to the spend data uses deaths under age 75 in England in 2008.														
34	Change in budget	£736.57	£678.96	£643.10		Note that the YLL figure for trauma & injuries has been estimated assuming that each death is on average at age 67 so that, on average, 7 years of life are lost per death.														

Table B10.4: table showing cost of life and life year estimates using spend data for 2007/8 and outcome data for 2007/2008/2009 (assumes zero health gain for PBC23 and 'average' gain for other 12 programmes)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
				=0.01*C*D				=0.01*D*G*H		=E/I		=I/L	=E/M			=0.01*D*H*P/3		=Q/R	=E/Q	=E/S
	PBC description	Spend (£m) 2007/8	Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2007/09	Outcome elasticity (without negative sign)	Change in annual mortality		Cost per life gained (£)	Coverage of mortality data relative to spend data	Change in annual mortality adj for coverage	Cost per life gained (£) adj for coverage		Total life years lost, <75years, 2007/09	Change in annual life years lost	Coverage of mortality data relative to spend data	Change in annual life years lost adj for coverage	Cost per life year gained (£)	Cost per life year gained adj for coverage (£)
1	Cancer	£4,573	0.890	£40.70		61,960	0.365	201.28		£202,207	0.984	204.55	£198,972		2,189,685	2,371	0.984	2,410	£17,165	£16,891
2	Circulatory problems	£6,325	0.293	£18.53		39,304	1.277	147.06		£126,018	0.992	148.25	£125,010		1,313,223	1,638	0.992	1,651	£11,315	£11,224
3	Respiratory problems	£3,431	0.536	£18.39		10,764	2.205	127.22		£144,557	0.773	164.58	£111,742		315,457	1,243	0.773	1,608	£14,798	£11,439
4	Gastro-intestinal problems	£3,805	0.622	£23.67		6,031	1.328	49.82		£475,081	0.571	87.25	£271,271		343,355	945	0.571	1,656	£25,034	£14,295
	Big four programmes summary:																			
5	Spend 2007 & mortality 2007/9	£18,134		£101.29		118,059		525.37		£192,795		604.62	£167,526		4,161,720	6,197		7,324	£16,345	£13,830
6	Spend 2006 & mortality 2006/8	£17,268		£91.24		120,801		665.10		£137,188		761.49	£119,823		4,238,786	7,399		8,604	£12,333	£10,604
7	Spend 2006 & mortality 2004/6	£17,268		£114.04		125,290		953.13		£119,650					4,335,559	10,576			£10,783	
8	Spend 2005 & mortality 2002/4	£17,625		£141.22		125,290		909.96		£155,196					4,516,953	10,986			£12,855	
9	Infectious diseases	£1,119	1.436	£16.07		1,977	0.548	15.56		£1,032,863	1.000	15.56	£1,032,863		106,092	278	1.000	278	£57,742	£57,742
10	Endocrine problems	£1,997	0.264	£5.27		1,471	0.566	2.20		£2,398,551	0.634	3.47	£1,520,681		55,492	28	0.634	44	£190,745	£120,932
11	Neurological problems	£3,165	1.035	£32.76		718	0.339	2.52		£13,003,180	0.136	18.52	£1,768,432		64,873	76	0.136	558	£431,749	£58,718
12	Genito-urinary problems	£3,439	1.004	£34.53		270	1.855	5.03		£6,866,327	0.172	29.24	£1,181,008		8,529	53	0.172	308	£652,096	£112,160
13	Trauma & injuries*	£2,918	1.686	£49.20		1,013	0.369	6.30		£7,806,376	0.175	36.01	£1,366,116		21,273	44	0.175	252	£1,115,197	£195,159
14	Maternity & neonates*	£3,662	0.514	£18.82		2,199	0.11	1.24		£15,139,113	8.213	0.15	£124,337,534		489,170	92	0.679	136	£204,168	£138,630
	Other six programmes summary:																			
15	Spend 2007 & mortality 2007/9	£16,300		£156.65		7,648		32.85		£4,768,699		102.95	£1,521,610		745,429	571		1,575	£274,309	£99,428
16	Spend 2006 & mortality 2006/8	£15,643		£93.29		7,839		16.14		£5,780,723		25.05	£3,724,129		762,991	362		639	£258,046	£146,108
17	Spend 2006 & mortality 2004/6	£15,643		£112.13		7,923		18.17		£6,172,491					757,531	249			£449,706	
18	Spend 2005 & mortality 2002/4	£12,743		£99.44		7,923		16.26		£6,115,621					751,009	337			£295,074	
	All ten programmes:																			
19	Spend 2007 & mortality 2007/9	£34,434		£257.94		125,707		558.22		£462,067		707.57	£364,540		4,907,149	6,768		8,900	£38,110	£28,983
20	Spend 2006 & mortality 2006/8	£32,911		£184.53		128,640		681.24		£270,881		786.54	£234,617		5,001,777	7,760		9,243	£23,780	£19,965
21	Spend 2006 & mortality 2004/6	£32,911		£226.18		133,213		971.30		£232,861					5,093,090	10,826			£20,893	
22	Spend 2005 & mortality 2002/4	£30,368		£240.67		133,213		926.22		£259,838					5,267,962	11,322			£21,256	
	Other 13 PBCs? Assume zero health gain in PBC23…																			
23	PBC23: spend 2007 & mortality 2007/9	£11,763	0.563	£66.23				0.00				0.00				0.00		0.00		
24	PBC23: spend 2006 & mortality 2006/8	£10,585	0.739	£78.22				0.00								0.00		0.00		
25	PBC23: spend 2006 & mortality 2004/6	£10,585	0.759	£80.34				0.00								0.00				
26	PBC23: spend 2005 & mortality 2002/4	£8,449	0.926	£78.24				0.00								0.00				
	...and that the gain in ten PBCs (see row 19) applies to the remaining 12 PBCs																			
27	12 PBCs: spend 2007 & mortality 2007/9	£27,460		£412.41				892.53		£462,067		1,131.31	£364,540			10,821		14,229	£38,110	£28,983
28	12 PBCs: spend 2006 & mortality 2006/8	£24,400		£416.20				1,536.48		£270,881		1,773.97	£234,617			17,502		20,847	£23,780	£19,965
29	12 PBCs: spend 2006 & mortality 2004/6	£24,400		£372.44				1,599.42		£232,861						17,826			£20,893	
30	12 PBCs: spend 2005 & mortality 2002/4	£25,493		£324.20				1,247.69		£259,838						15,252			£21,256	
	All 23 programmes																			
31	23 PBCs: spend 2007 & mortality 2007/9	£73,657		£736.57				1,450.75		£507,717		1,838.88	£400,554			17,590		23,129	£41,875	£31,846
32	23 PBCs: spend 2006 & mortality 2006/8	£67,896		£678.96				2,217.72		£306,153		2,560.50	£265,167			25,262		30,090	£26,876	£22,565
33	23 PBCs: spend 2006 & mortality 2004/6	£67,896		£678.96				2,570.72		£264,113						28,652			£23,697	
34	23 PBCs: spend 2005 & mortality 2002/4	£64,310		£643.10				2,173.90		£295,827						26,575			£24,200	
	Note:	2007/8	2006/7	2005/6																
35	All 23 programme spend	£73,657	£67,896	£64,310																
36	% change in budget	1.00	1.00	1.00																
37	proportionate change	0.01	0.01	0.01																
38	Change in budget	£736.57	£678.96	£643.10																

Note that the annual mortality figures reported in cells G7 & G8 and G17 & G18 are identical because we do not have mortality data for 2002/04.

Note that the coverage of the YLL data relative to the spend data for trauma & injuries is assumed to take a value of 1.0 (that is, the ICD coverage is the same).

Note that, for expenditure in 2007/8, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

Note that the adjustment for the coverage of the mortality data relative to the spend data uses deaths under age 75 in England in 2008.

B11. Analysis of programme budgeting expenditure for 2008/9 and mortality data for 2008/2010

Outcome and expenditure models were estimated using updated data for expenditure (from 2007/8 to 2008/9) and updated mortality data (from 2007/2008/2009 to 2008/2009 /2010). Results for the outcome model are shown in Table B11.1 and results for the expenditure model are in Table B11.2. First stage regressions for these IV models can be found in Tables BA.9 and BA.10 in the annex.

B11.1 Outcome models

Most of the outcome models in Table B11.1 contain just two variables: own programme expenditure and a measure of the need for health care. The latter is usually the measure of need as employed by the Department of Health for resource allocation purposes and this incorporates the CARAN formula for acute services. For the respiratory disease programme we have added the square of the need measure to improve the model fit. In other PBCs (e.g., for the endocrine, metabolic and nutritional programme), we found that the all service measure of need performed poorly and we have replaced it with a more programme specific measure (e.g., the diabetes prevalence rate) or with a better performing proxy for need (e.g., the percentage of residents born outside the EU for maternity/neonate mortality).

The relevant statistical test suggests that expenditure is endogenous in six of the ten programmes but we have retained the IV estimates for the other four because they provide plausible results. The Hansen-Sargen test suggests that the selected instruments are valid, and the Kleibergen-Paap LM statistic suggests that they are relevant (i.e., correlated with the endogenous regressor). The Kleibergen-Paap F statistic suggests that we do not have a problem with weak instruments (although the F statistic is marginally less than the conventional target value of ten in the genitor-urinary and infectious disease programmes). Finally, the Pesaran-Taylor reset test statistics reveal no evidence of mis-specification.

Results for the big four programmes are shown in the first five columns of Table B11.1. Two results are reported for the gastro-intestinal programme. The first of these (column 4) uses two instruments and so we report the instrument validity test statistic. However, one of these instruments is insignificant in the first-stage regression and, if we drop this instrument and re-estimate the model, we obtain the result in column 5 (but of course there is no instrument validity test statistic with this re-estimation). The removal of one instrument has little impact on the coefficient on expenditure but the Kleibergen-Paap F statistic is now much greater than ten.

In all of the big four programmes the need variable has a positive and significant effect on mortality, and expenditure has the anticipated negative effect. As we have noted before, the outcome results for the other programmes (in columns 6 - 10) are similar to but more diverse than those for the big four programmes. This is to be anticipated because mortality is a much rarer outcome in these programmes than it is in the big four programmes.

Expenditure has the anticipated negative effect on mortality in the endocrine problems programme and this is statistically significant. The all service measure of need is not relevant for this PBC; instead, we find that the diabetes prevalence rate is positively associated with mortality, as is a measure of deprivation (the IMD2007).

Expenditure has a negative but statistically insignificant impact on mortality from epilepsy in the neurological programme, and the all service indicator of the need for health care is positively and significantly associated with mortality in this programme.

Expenditure also has a negative but not statistically significant effect on mortality (from renal problems) in the genitor-urinary problems programme. The prevalence of lone parent households is positively associated with mortality.

Expenditure has the anticipated negative effect on mortality in the infectious disease programme and this is statistically significant. The all service measure of need is not relevant for this PBC; instead, we find

that a measure of need associated with HIV is positively associated with mortality, as is a measure of deprivation (the IMD2007).

Expenditure has the anticipated negative effect on mortality in the maternity & neonates programme. In this PBC the coefficient on the generic all service measure of need is positive but not significant. It has been supplemented with two other indicators of deprivation – the proportion of residents born outside the EU and the proportion of those aged 16-74 without any qualifications – and both of these are positively associated with mortality.

Finally, we were unable to develop a plausible outcome model for the trauma and injuries programme.

Table B11.1: table showing uutcome models using spend data for 2008/9 (two MFFs) and mortality data for 2008/9/10

	(1) PBC 2 cancer 2008/9 outcome model instrument spend weighted second stage	(2) PBC 10 circulation 2008/9 outcome model instrument spend weighted second stage	(3) PBC 11 respiratory 2008/9 outcome model instrument spend weighted second stage	(4) PBC 13 gastro-intestinal 2008/9 outcome model instrument spend weighted second stage	(5) PBC 13 gastro-intestinal 2008/9 outcome model instrument spend weighted second stage	(6) PBC 4 endocrine 2008/9 outcome model instrument spend weighted second stage	(7) PBC 7 neurological 2008/9 outcome model instrument spend weighted second stage	(8) PBC 17 genito-urinary 2008/9 outcome model instrument spend weighted second stage	(9) PBC 1 infectious disease 2008/9 outcome model instrument spend weighted second stage	(10) PBC 1819 maternityneonate 2008/9 outcome model instrument spend weighted second stage
VARIABLES										
own programme spend per head	-0.307***	-1.319***	-1.808***	-1.287***	-1.364**	-1.170***	-0.417	-1.615	-0.504**	-0.125
	[0.084]	[0.186]	[0.488]	[0.478]	[0.549]	[0.431]	[0.473]	[1.608]	[0.223]	[0.188]
needCARAN	0.954***	2.840***	4.811***	3.907***	3.993***		1.280**			0.405
	[0.095]	[0.247]	[0.760]	[0.625]	[0.700]		[0.579]			[0.288]
needCARAN2			3.016**							
			[1.284]							
diabetes prevalence rate						0.903**				
						[0.371]				
IMD 2007						0.711***			0.528***	
						[0.108]			[0.091]	
lone parent households								1.820***		
								[0.659]		
HIV need per head									0.468***	
									[0.093]	
HIV need per head squared									0.163***	
									[0.046]	
born outside EU										0.169***
										[0.031]
population with no qualifications										0.752***
										[0.129]
Constant	6.372***	10.861***	10.818***	8.715***	9.048***	2.107**	3.233	11.065	1.844***	3.097***
	[0.381]	[0.908]	[2.111]	[2.076]	[2.386]	[1.022]	[1.987]	[8.588]	[0.500]	[0.949]
Observations	151	151	151	151	151	151	151	148	151	151
Endogeneity test statistic	11.547	25.007	30.177	14.839	11.963	6.209	2.251	0.530	2.952	0.340
Endogeneity p-value	0.000679	5.71e-07	3.94e-08	0.000117	0.000543	0.0127	0.133	0.467	0.0858	0.560
Hansen-Sargan test statistic	0.843	0.801	0.00285	0.101		0.558	4.446	3.513	4.412	0.225
Hansen-Sargan p-value	0.358	0.371	0.957	0.751		0.757	0.108	0.0609	0.220	0.635
Shea's partial R-squared	0.245	0.282	0.176	0.192	0.150	0.193	0.155	0.103	0.191	0.263
Kleibergen-Paap LM test statistic	23.51	24.85	13.79	13.60	11.64	25.23	21.85	12.51	20.29	22.02
Kleibergen-Paap p-value	7.85e-06	4.02e-06	0.00101	0.00111	0.000644	1.38e-05	7.02e-05	0.00192	0.000437	1.65e-05
Kleibergen-Paap F statistic	21.14	47.87	15.10	11.93	16.51	13.56	20.13	9.000	9.306	16.92
Pesaran-Taylor reset statistic	0.416	0.405	0.104	0.483	0.0584	1.211	0.838	1.681	0.0456	0.107
Pesaran-Taylor p-value	0.519	0.524	0.747	0.487	0.809	0.271	0.360	0.195	0.831	0.744

Note: robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

B11.2 Expenditure models

Most of the expenditure models in Table B11.2 contain just three variables: the PCT budget, a proxy for the own programme need for health care, and a proxy for the need for health care in other programmes.

The budget term is positive and statistically significant in ten of the eleven models.

The usual proxy for the own programme need for health care (i.e., the all service measure of need) is positive and significant in five of the eleven results. In a couple of programmes (respiratory disease and endocrine problems) we have added the squared value of need to improve the model fit and in both cases this term is positive and significant.

In some programmes (e.g., the endocrine PBC and the neurological PBC), we have replaced and/or supplemented the all service measure of need with a more programme specific measure (e.g., the diabetes and the epilepsy prevalence rates) and these usually have a positive and significant impact on expenditure.

In addition, in a couple of programmes we have used alternative proxies for own programme need (e.g., with the use of the Department of Health's measure of maternity need in the maternity/neonates expenditure equation and the use of HIV need in the infectious diseases programme).

For eight of the eleven programmes we have used the all cause mortality rate less the own programme mortality rate as the proxy for the need for health care in other programmes, and the coefficient on this term is negative in seven programmes and statistically significant in six of the seven. In three programmes -- maternity/neonates, GMS/PMS and trauma & injuries programmes -- we have used the all cause mortality rate as the proxy for the need for health care in other programmes due to difficulties associated with the measurement of the own programme mortality rate. The coefficient on this term is negative but not significant in these three models.

The relevant statistical test suggests that expenditure is endogenous in five of the eleven programmes but we have retained the IV estimates for two further programmes (endocrine problems and maternity/neonates) because the IV estimator provides more plausible results than the OLS estimator. In the other four programmes we report OLS results.

The Hansen-Sargen test suggests that the selected instruments are valid, and the Kleibergen-Paap LM statistic suggests that they are relevant (i.e., correlated with the endogenous regressor). The Kleibergen-Paap F statistic suggests that we do not have a problem with weak instruments. Finally, the Pesaran-Taylor reset test statistics and the Ramsey reset F statistics reveal no evidence of model mis-specification.

Table B11.2: table showing expenditure models using spend data for 2008/9 (two MFFs) and mortality data for 2008/9/10

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	PBC 2 cancer 2008/9 spend model	PBC 10 circulatory 2008/9 spend model	PBC 11 respiratory 2008/9 spend model	PBC 13 gastro-intestinal 2008/9 spend model	PBC 1 infectiousdisease 2008/9 spend model	PBC 4 endocrine 2008/9 spend model	PBC 7 neurological 2008/9 spend model	PBC 17 genito-urinary 2008/9 spend model	PBC 1819 maternity/neonates 2008/9 spend model	PBC 23a GMS/PMS 2008/9 spend model	PBC 16 trauma/injuries 2008/9 spend model
	instrument o/need weighted	instrument o/need weighted	instrument o/need weighted	instrument o/need weighted	o/need exogenous weighted	instrument o/need weighted	instrument o/need weighted	o/need exogenous weighted	instrument o/need weighted	o/need exogenous weighted	o/need exogenous weighted
VARIABLES	second stage	second stage	second stage	second stage	OLS	second stage	second stage	OLS	second stage	OLS	OLS
all cause SYLLR exc cancer	-1.216***										
	[0.186]										
PCT budget per head	0.525*	0.648	0.652*	0.456*	1.546***	0.484**	0.980***	0.697***	0.975***	0.494***	1.344***
	[0.296]	[0.552]	[0.337]	[0.254]	[0.265]	[0.240]	[0.220]	[0.209]	[0.303]	[0.140]	[0.236]
need CARAN per head	2.081***	2.606***	2.036***	2.095***		0.553		0.295		0.724**	
	[0.389]	[0.623]	[0.377]	[0.411]		[0.369]		[0.310]		[0.334]	
all SYLLR exc circulatory		-1.987***									
		[0.351]									
all SYLLR exc respiratory			-1.081***								
			[0.264]								
need CARAN per head squar			1.336***			1.602***					
			[0.501]			[0.495]					
all SYLLR exc gastro				-1.256***							
				[0.317]							
HIV need per head					0.456***						
					[0.027]						
all SYLLR exc infectious dis					-0.472**						
					[0.227]						
HIV need per head squared					0.178***						
					[0.023]						
all SYLLR excluding diabetes						-0.164					
						[0.197]					
diabetes prevalence rate						0.439***					
						[0.112]					
all SYLLR excluding epilepsy							-0.257*				
							[0.153]				
epilepsy prevalence rate							0.414***				
							[0.063]				
born outside EU								0.039***			
								[0.014]			
all SYLLR excluding renal								-0.029			
								[0.139]			
all cause SYLLR									-0.348	-0.106	-0.269
									[0.302]	[0.104]	[0.195]
maternity need per head									0.846***		
									[0.120]		
lone pensioner households										-0.166**	
										[0.079]	
permanently sick aged 16-74										-0.310***	
										[0.092]	

professional occupations										-0.124*	
										[0.064]	
working in agriculture											0.107***
											[0.022]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	PBC 2	PBC 10	PBC 11	PBC 13	PBC 1	PBC 4	PBC 7	PBC 17	PBC 1819	PBC 23a	PBC 16
	cancer	circulatory	respiratory	gastro-intestinal	infectiousdisease	endocrine	neurological	genito-urinary	maternity/neonates	GMS/PMS	trauma/injuries
	2008/9	2008/9	2008/9	2008/9	2008/9	2008/9	2008/9	2008/9	2008/9	2008/9	2008/9
	spend model	spend model	spend model	spend model	spend model	spend model	spend model	spend model	spend model	spend model	spend model
	instrument o/need	instrument o/need	instrument o/need	instrument o/need	o/need exogenous	instrument o/need	instrument o/need	o/need exogenous	instrument o/need	o/need exogenous	o/need exogenous
	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted
VARIABLES	second stage	second stage	second stage	second stage	OLS	second stage	second stage	OLS	second stage	OLS	OLS
Constant	7.556***	11.702***	6.044**	8.551***	-5.471***	0.488	-1.315	-0.521	-0.696	0.586	-3.605***
	[2.406]	[4.445]	[2.651]	[2.592]	[1.096]	[2.282]	[1.005]	[1.857]	[0.800]	[1.133]	[1.027]
Observations	151	151	151	151	151	151	151	151	151	150	151
R-squared					0.776			0.497		0.278	0.339
Endogeneity test statistic	17.101	22.697	17.212	12.023		1.803	7.163		3.243		
Endogeneity p-value	3.54e-05	1.90e-06	3.34e-05	0.000525		0.179	0.00744		0.0717		
Hansen-Sargan test statistic	0.0538	0.332	0.858	0.420		0.138	0.594		1.349		
Hansen-Sargan p-value	0.817	0.565	0.354	0.517		0.710	0.441		0.509		
Shea's partial R-squared	0.379	0.265	0.389	0.331		0.399	0.500		0.257		
Kleibergen-Paap LM statistic	39.01	29.71	37.32	33.84		38.45	35.08		22.81		
Kleibergen-Paap p-value	3.38e-09	3.54e-07	7.87e-09	4.48e-08		4.48e-09	2.41e-08		4.43e-05		
Kleibergen-Paap F statistic	39.97	26.93	44.98	20.13		47.20	75.67		16.35		
Pesaran-Taylor reset statistic	1.129	0.0810	0.000203	0.557		0.354	0.366		0.00412		
Pesaran-Taylor p-value	0.288	0.776	0.989	0.456		0.552	0.545		0.949		
Ramsey reset F statistic					1.723			1.431		0.072	1.044
Probability > F					0.165			0.236		0.975	0.375

Note: robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

B11.3 Calculation of the cost of a life and life year

Expenditure and outcome elasticities for our preferred models are shown in Table B11.3 (see columns D and H) and these are used to calculate the cost of a life and the cost of a life year, both for individual programmes and for all programmes collectively.

Again, column N reports the cost per life gained and column U reports the cost per life year gained. From the latter we can see that the cost per life year gained has increased slightly compared with that using the previous expenditure and mortality data set (i.e., for 2007 and 2007/8/9 respectively): it has increased from £13,830 to £14,650 for the big four programmes and from £28,983 to £30,883 for all ten programmes with a mortality-based outcome indicator.

If we assume that the other 13 programmes offer no health gain, then the cost per life year across all PCT expenditure has increased from £82,765 in 2007/8 to £84,974 in 2008/9.

In addition, Table B11.4 shows that if we assume that PBC23 generates a zero health gain and that the gain attributable to the remaining 12 programmes is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes in 2008/9 is £33,333. This is a 5% increase on the figure (£31,846) for the previous year.

Table B11.3: table showing cost of life and life year estimates using spend data for 2008/9 and outcome data for 2008/2009/2010 (assumes zero health gain for 13 programmes)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
				=0.01*C*D				=0.01*D*G*H		=E/I		=I/L	=E/M			=0.01*D*H*P/3		=Q/R	=E/Q	=E/S
	PBC description	Spend (£m) 2008/9	Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2008/10	Outcome elasticity (without negative sign)	Change in annual mortality		Cost per life gained (£)	Coverage of mortality data relative to spend data	Change in annual mortality adj for coverage	Cost per life gained (£) adj for coverage		Total life years lost, <75years, 2008/10	Change in annual life years lost	Coverage of mortality data relative to spend data	Change in annual life years lost adj for coverage	Cost per life year gained (£)	Cost per life year gained adj for coverage (£)
1	Cancer	£4,843	0.525	£25.43		61,899	0.307	99.77		£254,855	0.984	101.39	£250,777		2,170,660	1,166	0.984	1,185	£21,802	£21,454
2	Circulatory problems	£6,655	0.648	£43.12		38,075	1.319	325.43		£132,514	0.992	328.06	£131,454		1,285,026	3,661	0.992	3,691	£11,779	£11,685
3	Respiratory problems	£3,994	0.652	£26.04		10,660	1.808	125.66		£207,230	0.773	162.56	£160,189		311,034	1,222	0.773	1,581	£21,307	£16,470
4	Gastro-intestinal problems	£3,989	0.456	£18.19		6,015	1.364	37.41		£486,199	0.571	65.52	£277,620		341,884	709	0.571	1,241	£25,662	£14,653
	Big four programmes summary:																			
5	Spend 2008 & mortality 2008/10	£19,481		£112.78		116,649		588.27		£191,716		657.53	£171,552		4,108,604	6,758		7,698	£16,688	£14,650
6	Spend 2007 & mortality 2007/9	£18,134		£101.29		118,059		525.37		£192,795		604.62	£167,526		4,161,720	6,197		7,324	£16,345	£13,830
7	Spend 2006 & mortality 2006/8	£17,268		£91.24		120,801		665.10		£137,188		761.49	£119,823		4,238,786	7,399		8,604	£12,333	£10,604
8	Spend 2006 & mortality 2004/6	£17,268		£114.04		125,290		953.13		£119,650					4,335,559	10,576			£10,783	
9	Spend 2005 & mortality 2002/4	£17,625		£141.22		125,290		909.96		£155,196					4,516,953	10,986			£12,855	
10	Infectious diseases	£1,201	1.545	£18.56		1,828	0.504	14.23		£1,303,576	1.000	14.23	£1,303,576		100,078	260	1.000	260	£71,432	£71,432
11	Endocrine problems	£2,222	0.484	£10.75		1,398	1.17	7.92		£1,358,473	0.634	12.49	£861,272		54,779	103	0.634	163	£104,008	£65,941
12	Neurological problems	£3,466	0.98	£33.97		711	0.417	2.91		£11,690,226	0.136	21.36	£1,589,871		64,222	87	0.136	643	£388,267	£52,804
13	Genito-urinary problems	£3,779	0.697	£26.34		240	1.615	2.70		£9,749,742	0.172	15.71	£1,676,956		8,004	30	0.172	175	£877,038	£150,851
14	Trauma & injuries*	£3,255	1.344	£43.75		983	0	0.00		#DIV/0!	0.175	0.00	#DIV/0!		6,881	0	0.175	0	#DIV/0!	#DIV/0!
15	Maternity & neonates*	£3,978	0.975	£38.79		2,156	0.125	2.63		£14,760,668	8.213	0.32	£121,229,365		479,905	195	0.679	287	£198,939	£135,080
	Other six programmes summary:																			
16	Spend 2008 & mortality 2008/10	£17,901		£172.15		7,316		30.39		£5,665,475		64.11	£2,685,119		713,869	676		1,528	£254,794	£112,674
17	Spend 2007 & mortality 2007/9	£16,300		£156.65		7,648		32.85		£4,768,699		102.95	£1,521,610		745,429	571		1,575	£274,309	£99,428
18	Spend 2006 & mortality 2006/8	£15,643		£93.29		7,839		16.14		£5,780,723		25.05	£3,724,129		762,991	362		639	£258,046	£146,108
19	Spend 2006 & mortality 2004/6	£15,643		£112.13		7,923		18.17		£6,172,491					757,531	249			£449,706	
20	Spend 2005 & mortality 2002/4	£12,743		£99.44		7,923		16.26		£6,115,621					751,009	337			£295,074	
	All ten programmes summary:																			
21	Spend 2008 & mortality 2008/10	£37,382		£284.93		123,965		618.66		£460,562		721.64	£394,836		4,822,473	7,434		9,226	£38,328	£30,883
22	Spend 2007 & mortality 2007/9	£34,434		£257.94		125,707		558.22		£462,067		707.57	£364,540		4,907,149	6,768		8,900	£38,110	£28,983
23	Spend 2006 & mortality 2006/8	£32,911		£184.53		128,640		681.24		£270,881		786.54	£234,617		5,001,777	7,760		9,243	£23,780	£19,965
24	Spend 2006 & mortality 2004/6	£32,911		£226.18		133,213		971.30		£232,861					5,093,090	10,826			£20,893	
25	Spend 2005 & mortality 2002/4	£30,368		£240.67		133,213		926.22		£259,838					5,267,962	11,322			£21,256	
	Assume zero health gain in the other 13 programmes																			
	Other 13 programmes summary:																			
26	Spend 2008 & mortality 2008/10	£41,016		£499.05				0.00				0.00				0		0		
27	Spend 2007 & mortality 2007/9	£39,223		£478.63				0.00				0.00				0		0		
28	Spend 2006 & mortality 2006/8	£34,985		£494.43				0.00								0		0		
29	Spend 2006 & mortality 2004/6	£34,985		£452.78				0.00								0				
30	Spend 2005 & mortality 2002/4	£33,942		£402.43				0.00								0				
	All 23 programmes																			
31	Spend 2008 & mortality 2008/10	£78,398		£783.98				618.66		£1,267,229		721.64	£1,086,385			7,434		9,226	£105,460	£84,974
32	Spend 2007 & mortality 2007/9	£73,657		£736.57				558.22		£1,319,496		707.57	£1,040,992			6,768		8,900	£108,829	£82,765
33	Spend 2006 & mortality 2006/8	£67,896		£678.96				681.24		£996,655		786.54	£863,228			7,760		9,243	£87,494	£73,457
34	Spend 2006 & mortality 2004/6	£67,896		£678.96				971.30		£699,024						10,826			£62,718	
35	Spend 2005 & mortality 2002/4	£64,310		£643.10				926.22		£694,330						11,322			£56,799	

	Note:	2008/9	2007/8	2006/7	2005/6
31	All 23 programme spend	£78,398	£73,657	£67,896	£64,310
32	% change in budget	1.00	1.00	1.00	1.00
33	proportionate change	0.01	0.01	0.01	0.01
34	Change in budget	£783.98	£736.57	£678.96	£643.10

Note that the annual mortality figures reported in cells G7 & G8 and G17 & G18 are identical because we do not have mortality data for 2002/04.

Note that, for expenditure in 2008/9, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

Note that the adjustment for the coverage of the mortality & YLL data relative to the spend data uses deaths under age 75 in England in 2008.

Note that the YLL figure for trauma & injuries has been estimated assuming that each death is on average at age 67 so that, on average, 7 years of life are lost per death.

Note that the coverage of the YLL data relative to the spend data for trauma & injuries is assumed to take a value of 1.0 (that is, the ICD coverage is the same).

Note that the YLL for maternity and neonates is estimated as [(6,339 neonate deaths*75years)+(128 maternal deaths*35years)]. This totals 479,905 life years.

Table B11.4: table showing cost of life and life year estimates using spend data for 2008/9 and outcome data for 2008/2009/2010 (assumes zero health gain for PBC23 and average gain for other 12 programmes)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
				=0.01*C*D				=0.01*D*G*H		=E/I		=I/L	=E/M			=0.01*D*H*P/3		=Q/R	=E/Q	=E/S
	PBC description	Spend (£m) 2008/9	Spend elasticity	Change in spend (£m)		Annual mortality, <75years, 2008/10	Outcome elasticity (without negative sign)	Change in annual mortality		Cost per life gained (£)	Coverage of mortality data relative to spend data	Change in annual mortality adj for coverage	Cost per life gained (£) adj for coverage		Total life years lost, <75years, 2008/10	Change in annual life years lost	Coverage of mortality data relative to spend data	Change in annual life years lost adj for coverage	Cost per life year gained (£)	Cost per life year gained adj for coverage (£)
1	Cancer	£4,843	0.525	£25.43		61,899	0.307	99.77		£254,855	0.984	101.39	£250,777		2,170,660	1,166	0.984	1,185	£21,802	£21,454
2	Circulatory problems	£6,655	0.648	£43.12		38,075	1.319	325.43		£132,514	0.992	328.06	£131,454		1,285,026	3,661	0.992	3,691	£11,779	£11,685
3	Respiratory problems	£3,994	0.652	£26.04		10,660	1.808	125.66		£207,230	0.773	162.56	£160,189		311,034	1,222	0.773	1,581	£21,307	£16,470
4	Gastro-intestinal problems	£3,989	0.456	£18.19		6,015	1.364	37.41		£486,199	0.571	65.52	£277,620		341,884	709	0.571	1,241	£25,662	£14,653
	Big four programmes summary:																			
5	Spend 2008 & mortality 2008/10	£19,481		£112.78		116,649		588.27		£191,716		657.53	£171,552		4,108,604	6,758		7,698	£16,688	£14,650
6	Spend 2007 & mortality 2007/9	£18,134		£101.29		118,059		525.37		£192,795		604.62	£167,526		4,161,720	6,197		7,324	£16,345	£13,830
7	Spend 2006 & mortality 2006/8	£17,268		£91.24		120,801		665.10		£137,188		761.49	£119,823		4,238,786	7,399		8,604	£12,333	£10,604
8	Spend 2006 & mortality 2004/6	£17,268		£114.04		125,290		953.13		£119,650					4,335,559	10,576			£10,783	
9	Spend 2005 & mortality 2002/4	£17,625		£141.22		125,290		909.96		£155,196					4,516,953	10,986			£12,855	
10	Infectious diseases	£1,201	1.545	£18.56		1,828	0.504	14.23		£1,303,576	1.000	14.23	£1,303,576		100,078	260	1.000	260	£71,432	£71,432
11	Endocrine problems	£2,222	0.484	£10.75		1,398	1.17	7.92		£1,358,473	0.634	12.49	£861,272		54,779	103	0.634	163	£104,008	£65,941
12	Neurological problems	£3,466	0.98	£33.97		711	0.417	2.91		£11,690,226	0.136	21.36	£1,589,871		64,222	87	0.136	643	£388,267	£52,804
13	Genito-urinary problems	£3,779	0.697	£26.34		240	1.615	2.70		£9,749,742	0.172	15.71	£1,676,956		8,004	30	0.172	175	£877,038	£150,851
14	Trauma & injuries*	£3,255	1.344	£43.75		983	0	0.00		#DIV/0!	0.175	0.00	#DIV/0!		6,881	0	0.175	0	#DIV/0!	#DIV/0!
15	Maternity & neonates*	£3,978	0.975	£38.79		2,156	0.125	2.63		£14,760,668	8.213	0.32	£121,229,365		479,905	195	0.679	287	£198,939	£135,080
	Other six programmes summary:																			
16	Spend 2008 & mortality 2008/10	£17,901		£172.15		7,316		30.39		£5,665,475		64.11	£2,685,119		713,869	676		1,528	£254,794	£112,674
17	Spend 2007 & mortality 2007/9	£16,300		£156.65		7,648		32.85		£4,768,699		102.95	£1,521,610		745,429	571		1,575	£274,309	£99,428
18	Spend 2006 & mortality 2006/8	£15,643		£93.29		7,839		16.14		£5,780,723		25.05	£3,724,129		762,991	362		639	£258,046	£146,108
19	Spend 2006 & mortality 2004/6	£15,643		£112.13		7,923		18.17		£6,172,491					757,531	249			£449,706	
20	Spend 2005 & mortality 2002/4	£12,743		£99.44		7,923		16.26		£6,115,621					751,009	337			£295,074	
	All ten programmes summary:																			
21	Spend 2008 & mortality 2008/10	£37,382		£284.93		123,965		618.66		£460,562		721.64	£394,836		4,822,473	7,434		9,226	£38,328	£30,883
22	Spend 2007 & mortality 2007/9	£34,434		£257.94		125,707		558.22		£462,067		707.57	£364,540		4,907,149	6,768		8,900	£38,110	£28,983
23	Spend 2006 & mortality 2006/8	£32,911		£184.53		128,640		681.24		£270,881		786.54	£234,617		5,001,777	7,760		9,243	£23,780	£19,965
24	Spend 2006 & mortality 2004/6	£32,911		£226.18		133,213		971.30		£232,861					5,093,090	10,826			£20,893	
25	Spend 2005 & mortality 2002/4	£30,368		£240.67		133,213		926.22		£259,838					5,267,962	11,322			£21,256	
	Other 13 PBCs? Assume zero health gain in PBC23…																			
26	PBC23: spend 08, mortality 8/10	£11,663	0.494	£57.62				0.00				0.00				0.00		0.00		
27	PBC23: spend 07, mortality 7/9	£11,763	0.563	£66.23				0.00				0.00				0.00		0.00		
28	PBC23: spend 06, mortality 6/8	£10,585	0.739	£78.22				0.00								0.00		0.00		
29	PBC23: spend 06, mortality 4/6	£10,585	0.759	£80.34				0.00								0.00				
30	PBC23: spend 05, mortality 2/4	£8,449	0.926	£78.24				0.00								0.00				
	...and that the gain in 10 PBCs (see row 21) applies to the remaining 12 PBCs																			
31	12 PBCs:spend 08,mortality 8/10	£29,353		£441.43				958.47		£460,562		1,118.02	£394,836			11,517		14,294	£38,328	£30,883
32	12 PBCs: spend 07, mortality 7/9	£27,460		£412.41				892.53		£462,067		1,131.31	£364,540			10,821		14,229	£38,110	£28,983
33	12 PBCs: spend 06, mortality 6/8	£24,400		£416.20				1,536.48		£270,881		1,773.97	£234,617			17,502		20,847	£23,780	£19,965
34	12 PBCs: spend 06, mortality 4/6	£24,400		£372.44				1,599.42		£232,861						17,826			£20,893	
35	12 PBCs: spend 05, mortality 2/4	£25,493		£324.20				1,247.69		£259,838						15,252			£21,256	
	All 23 programmes																			
36	23 PBCs:spend 08,mortality 8/10	£78,398		£783.98				1,577.13		£497,094		1,839.66	£426,155			18,951		23,520	£41,369	£33,333
37	23 PBCs: spend 07, mortality 7/9	£73,657		£736.57				1,450.75		£507,717		1,838.88	£400,554			17,590		23,129	£41,875	£31,846
38	23 PBCs: spend 06, mortality 6/8	£67,896		£678.96				2,217.72		£306,153		2,560.50	£265,167			25,262		30,090	£26,876	£22,565
39	23 PBCs: spend 06, mortality 4/6	£67,896		£678.96				2,570.72		£264,113						28,652			£23,697	

40	23 PBCs: spend 05, mortality 2/4	£64,310		£643.10			2,173.90	£295,827		26,575	£24,200

	Note:	2008/9	2007/8	2006/7	2005/6
35	All 23 programme spend	£78,398	£73,657	£67,896	£64,310
36	% change in budget	1.00	1.00	1.00	1
37	proportionate change	0.01	0.01	0.01	0.01
38	Change in budget	£783.98	£736.57	£678.96	£643.10

Note that the annual mortality figures reported in cells G7 & G8 and G17 & G18 are identical because we do not have mortality data for 2002/04.
Note that the coverage of the YLL data relative to the spend data for trauma & injuries is assumed to take a value of 1.0 (that is, the ICD coverage is the same).
Note that, for expenditure in 2007/8, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.
Note that the adjustment for the coverage of the mortality data relative to the spend data uses deaths under age 75 in England in 2008.

B11.4 Comparing the cost of life year estimates associated with different data sets

Table B11.5 presents expenditure and outcome elasticities for the five combinations of expenditure and outcome data that have been used to estimate our model. It also reports the corresponding unadjusted cost of life year estimates (i.e., estimates that are unadjusted for the mismatch in the ICD10 coverage of the expenditure and mortality data). It is clear from this Table (see row 13) that the (unadjusted) cost of a life year for the ten programmes with a mortality based outcome indicator fluctuated around £22,000 for the first three sets of estimations (see columns M-O). However, using the two most recent sets of expenditure data (i.e., for 2007/8 and then for 2008/9), the figures in the table suggest that this cost has increased to about £38,000.

What are the proximate causes of this increase? Recall that the cost of a life year is calculated as

$$\frac{\text{the change in expenditure associated with a 1\% budget increase}}{\text{the change in the number of life years lost associated with this increase}}.$$

For 2006/7 (using mortality data for 2006/7/8) and for the ten programmes with a mortality based outcome indicator, the change in expenditure associated with a 1% budget increase is £184.53m and the change in the number of life years lost associated with this increase is 7,760 (see Table B8.21 in the appendix for the calculation of these figures). Thus the cost of a life year is £23,780 (=£184.53m/7,760).

For 2007/8 (using mortality data for 2007/8/9) and for the ten programmes with a mortality based outcome indicator, the change in expenditure associated with a 1% budget increase is £257.94m and the change in the number of life years lost associated with this increase is 6,768 (see Table B10.3 in the appendix for the calculation of these figures). Thus the cost of a life year is £38,110 (=£257.94m/6,768).

It is clear that the 60% increase in the cost of a life year between 2006/7 and 2007/8 is largely attributable (a) to the 40% increase in the additional expenditure (up from £184.53m to £257.94m) directed towards these ten programmes following a 1% budget increase and (b) to the 12% decline in the number of life years gained associated with this increase in expenditure (down from 7,760 to 6,768 life years).

The rise in the share of the budget increase directed towards these programmes can be attributed to the increase in the implied expenditure elasticity associated with these ten programmes (up from 0.561 to 0.749). The decrease in the number of years of life gained appears to be due (a) to an overall reduction in the (absolute) size of the outcome elasticities and (b) to a shift in the additional expenditure towards those programmes with a relatively high cost of a life year. For example, the cost of a life year for the 'small six' programmes is much larger than for the 'big four' programmes. However, in 2007/8 the spend elasticity for the small six increases from 0.561 to 0.961 (71%) while the expenditure elasticity for the big four rises from 0.528 to 0.559 (6%). A similar pattern – of additional expenditure shifting away from the low cost PBCs – can be seen within the big four programmes. However, it is not clear why such rather dramatic changes should have taken place.

Table B11.6 presents cost of life year estimates (adjusted for the mismatch in the ICD10 coverage of the expenditure and mortality data) for various combinations of programmes. These reveal similar increases in the cost of a life year between 2006/7 on the one hand and 2007/8 and 2008/9 on the other. The cost of a life year increased from £19,965 in 2006/7 to £28,983 in 2007/8 for the ten programmes with mortality rate, an increase of 45%; and it increased from £22,565 to £31,846 for all programmes if we assume a zero health gain in PBC23 and the same gain in the other 12 programmes as in the ten with a mortality rate (an increase of 41%).

One reason for this apparent step change in the cost of a life year might be the adjustment that was made to the methodology for the collection of the 2007/8 programme budgeting data. In previous years expenditure that was not directly attributable to a particular programme category was apportioned using

admitted patient care percentages.[36] In other words, if *x*% of total admitted patient care expenditure was allocated to PBC 1, then *x*% of all expenditure that was not directly attributable to a particular programme category was also allocated to PBC 1. With effect from 2007/8, however, NHS organisations were asked to select an appropriate basis for the apportionment of this non-programme specific expenditure and that, where no reasonable basis existed, such expenditure was to be allocated to the 'Other – Miscellaneous' (PBC 23X) category.

The Department of Health estimates that this allocation rule change increased the amount of expenditure attributed to PBC 23X by £700 million. It will also, of course, have reduced expenditure across other programmes by the same amount in total. However, not all programmes will have been equally affected; PBCs that are more heavily inpatient based would have 'lost' expenditure while others, such as learning disabilities, social care, and mental health, will have 'lost' considerably less. In addition, not all PCTs will have been equally affected because each will have employed different apportionment rules for the non-programme specific expenditure. [38]

Although this allocation rule change has considerably increased the estimated cost of a life year, we believe that this rule change has led to a more accurate allocation of expenditure across PBCs, and that the more recent estimates of the cost of a life year (for 2007/8 and 2008/9) are more accurate than those for the earlier years (for 2005/6 and 2006/7).

[36]Expenditure on, for example, community care, A&E, ambulance services, and outpatients can be difficult to attribute a particular PBC. Critical care, rehabilitation, and specialised commissioning across care settings will also be difficult to attribute to a particular programme.

B11.5 Adjusting the cost of a life year estimates to constant prices

The cost of a life year estimates presented above are all at current prices. To put them on a constant price basis, we need an index of pay and price inflation for the labour and goods/services purchased by the NHS. Curtis [41] reports a pay and prices index for Hospital and Community Health Services and this implies an inflation rate of 3.7% in 2006/7, 2.9% in 2007/8, and 3.9% in 2008/9.[37] If we assume that similar inflation rates also apply to the purchase of pharmaceuticals and the provision of primary care (items that are excluded from the HCHS index), then we can use these figures to put the cost of a life year estimates on a constant price basis.

For example, if we assume that PBC23 generates a zero health gain and that the gain attributable to the 12 programmes without a mortality indicator is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes in 2008/9 is £33,333 at current (2008/9) prices. The cost for 2007/8 is £31,846 at current (2007/8) prices or £33,088 at constant (2008/9) prices, and the figure for 2006/7 is £22,565 at current (2006/7) prices or £24,125 at constant (2008/9) prices. The conversion of the costs from a current to constant price basis has relatively little impact because the inflation rate over the relevant period is quite small.

B11.6 Summary and conclusion

In this section we have estimated outcome and expenditure models using PB data for 2008/9 and mortality data for 2008/9/10. The cost of an additional life year for all ten programmes with a mortality based outcome is £30,883. This is similar to the comparable figure (£28,983) for the previous year (i.e., using expenditure data for 2007/8 and mortality data for 2007/2008/2009). If we assume that PBC23 generates a zero health gain and that the gain attributable to the 12 programmes without a mortality indicator is, on average, the same as that attributable to those with a mortality outcome measure, then the cost of a life year across all programmes in 2008/9 is £33,333 and this too is similar to the figure for the previous year (£31,846).

We have also identified a pay and prices index that can be used to put the estimated costs on a constant price basis. This index has recorded an annual inflation rate of about 3.5% since 2005/6.

There appears to have been a step change in the cost of an additional life year. The cost of a life year estimates are very similar up to and including 2006/7, and they are very similar for 2007/8 and 2008/9. However, there is a substantial difference between the figures for 2004/5, 2005/6 and 2006/7 on the one hand, and for 2007/8 and 2008/9 on the other. The reason for this step change is not obvious but it might be due to changes in the algorithm used by the Department of Health to allocate non-admitted patient care activity to budget categories. Although this allocation rule change has considerably increased the estimated cost of a life year, we believe that this rule change has led to a more accurate allocation of expenditure across PBCs, and that the more recent estimates of the cost of a life year (for 2007/8 and 2008/9) are more accurate than those for the earlier years (for 2005/6 and 2006/7).

[37] With the index for 1987/8 set equal to 100, then 2005/6=240.9, 2006/7=249.8, 2007/8=257.0, and 2008/9=267.0 (Curtis, 2011, p209).

Table B11.5: table showing expenditure and outcome elasticities for five combinations of expenditure and outcome data, and corresponding (unadjusted) cost of life year estimates

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
		spend elasticities					outcome elasticities					cost of an additional life year (unadjusted for YLL coverage)				
	PBC description	(a) using spend for 2005 and mortality for 2002/4	(b) using spend for 2006 and mortality for 2004/6	(c) using spend for 2006 and mortality for 2006/8	(d) using spend for 2007 and mortality for 2007/9	(e) using spend for 2008 and mortality for 2008/10	(a) using spend for 2005 and mortality for 2002/4	(b) using spend for 2006 and mortality for 2004/6	(c) using spend for 2006 and mortality for 2006/8	(d) using spend for 2007 and mortality for 2007/9	(e) using spend for 2008 and mortality for 2008/10	(a) using spend for 2005 and mortality for 2002/4	(b) using spend for 2006 and mortality for 2004/6	(c) using spend for 2006 and mortality for 2006/8	(d) using spend for 2007 and mortality for 2007/9	(e) using spend for 2008 and mortality for 2008/10
1	Cancer	0.968	0.548	0.465	0.890	0.525	-0.394	-0.337	-0.342	-0.365	-0.307	£13,741	£16,518	£16,383	£17,165	£21,802
2	Circulatory problems	0.682	0.701	0.540	0.293	0.648	-1.370	-1.447	-1.434	-1.277	-1.319	£8,328	£8,725	£9,466	£11,315	£11,779
3	Respiratory problems	0.849	0.718	0.679	0.536	0.652	-1.574	-3.507	-2.622	-2.205	-1.808	£20,601	£8,747	£11,593	£14,798	£21,307
4	Gastro-intestinal problems	0.772	0.667	0.446	0.622	0.456	-2.018	-2.137	-1.536	-1.328	-1.364	£18,303	£15,795	£20,892	£25,034	£25,662
5	All big four PBCs	0.801	0.660	0.528	0.559	0.579	-0.941	-1.083	-0.965	-0.872	-0.825	£12,855	£10,783	£12,333	£16,345	£16,688
6	Infectious diseases	0.742	0.731	0.792	1.436	1.545	-0.152	-0.030	-0.047	-0.548	-0.504	£215,054	£1,036,377	£630,798	£57,742	£71,432
7	Endocrine problems	0.425	0.966	0.953	0.264	0.484	-0.244	-0.812	-0.842	-0.566	-1.170	£371,601	£112,882	£114,416	£190,745	£104,008
8	Neurological problems	1.111	0.648	0.616	1.035	0.98	-0.182	-0.098	-0.112	-0.339	-0.417	£503,201	£1,241,253	£1,129,960	£431,749	£388,267
9	Genito-urinary problems	1.041	0.837	0.912	1.004	0.697	-0.034	-0.073	-0.051	-1.855	-1.615	£29,144,918	£12,384,965	£20,421,090	£652,096	£877,038
10	Trauma & injuries*	0.627	0.617	0.358	1.686	1.344	-1.332	-0.527	0	-0.369	0	£282,132	£548,767	n/a	£1,115,197	n/a
11	Maternity & neonates*	0.388	0.601	0.224	0.514	0.975	-0.237	-0.035	-0.482	-0.110	-0.125	£17,490	£631,700	£45,158	£204,168	£198,939
12	All small six PBCs	0.780	0.717	0.596	0.961	0.962	-0.262	-0.122	-0.392	-0.254	-0.300	£295,074	£449,706	£258,046	£274,309	£254,794
13	All 10 PBCs with mortality	0.792	0.687	0.561	0.749	0.762	-0.844	-0.940	-0.877	-0.778	-0.747	£21,256	£20,893	£23,780	£38,110	£38,328
14	All 23 PBCs assuming zero gain in PBCs without mortality indicator											£56,799	£62,718	£87,494	£108,829	£105,460
15	GMS/PMS	0.926	0.759	0.739	0.563	0.494	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
16	All 23 PBCs assuming zero gain in PBC 23 but average gain in other PBCs without a mortality indicator											£24,200	£23,697	£26,876	£41,875	£41,369

Notes:

(i) that the spend and outcome elasticities reported for groups of programmes are the implied elasticites calculated from the totals for the relevant individual programmes (i.e., group spend elasticity=∑(PBC spend*PBC spend elasticity)/∑ PBC spend, and group outcome elasticity=∑(PBC mortality*PBC outcome elasticity)/∑ PBC mortality). For the purpose of the calculation of the implied group outcome elasticity, we have used the years of life lost as the mortality indicator. The implied group elasticities are directly comparable with the individual programme elasticities as both exclude the impact of the relevant budget elasticities. The implied group elasticities cannot be used to calculate directly the cost of a life (year) for a group of PBCs. Instead, the latter should be calculated by summing across the change in spend and the change in mortality for the individual PBCs within the group.

(ii) for each individual programme: the cost of an additional life year = expenditure elasticity*annual spend/(expenditure elasticity*outcome elasticity*annual life years lost)

(iii) for a group of programmes: the overall cost of an additional life year = ∑ (annual spend*spend elasticity) / ∑ (spend elasticity*outcome elasticity*annual life years lost)

(iv) that the results using expenditure for 2006/7 and mortality for 2004/6 incorporate MFFs for HCHS and prescribing (see Tables B8.17 and B8.18).

Table B11.6: table showing adjusted cost of life year estimates for various combinations of programmes

A	B	C	D	E
	Programme budgeting category	Cost per life year (adjusted for ICD10 coverage of spend and mortality data)		
		2006/7	2007/8	2008/9
1	Cancer	£16,121	£16,891	£21,454
2	Circulatory disease	£9,390	£11,224	£11,685
3	Respiratory problems	£8,961	£11,439	£16,470
4	Gastro-intestinal problems	£11,929	£14,295	£14,653
5	All big four programmes	£10,604	£13,830	£14,650
6	Other six programme with a mortality rate	£146,108	£99,428	£112,674
7	All ten PBCs with a mortality rate	£19,965	£28,983	£30,883
	(a) If we assume a zero health gain in those PBCs without a mortality rate…			
8	All 23 programmes	£73,457	£82,765	£84,974
	…or (b) if we assume a zero gain in PBC23 and that the average gain from the the 10 PBCs with a mortality rate is applied to the remaining programmes			
9	All 23 programmes	£22,565	£31,846	£33,333

Note that the figures for 2006/7 relate to the use of mortality for 2006/2007/2008 combined.

B12. Summary and concluding remarks

The findings presented in this report build on four previous studies[2-5]. These studies and the results presented here draw on the availability of two new data sets to obtain empirical estimates of the relationship between mortality and expenditure across all English local health authorities.

In this research we have extended the previous studies in several ways. First, we have derived plausible outcome and expenditure models for a larger number of programmes (ten) than previous studies. The cost of a life year across all ten programmes with a mortality based outcome indicator using expenditure data for 2006/7 and mortality data for 2004/5/6 is £20,893.

Second, we relate expenditure in time period *t* to mortality in that period (*t*) and in the next two periods (*t+1* and *t+2*). In other words, we assume that the health benefits associated with expenditure occur either in the same period as the expenditure or in the next two periods. This is an improvement on past practice where data constraints forced researchers to relate expenditure to the current and two previous periods. When we re-estimated our models having replaced mortality data for 2004/5/6 with that the 2006/7/8, we found that the cost of a life year across the ten programmes with a mortality based outcome indicator using expenditure data for 2006/7 is £23,780 (up from £20,893, an increase of 14%).

Third, we have noted the mismatch in the ICD10 coverage of the expenditure and mortality data. If we adjust the calculation of the cost of a life year for 2006/7 for this mismatch then the cost of a life year across the ten programmes with a mortality based outcome indicator declines from £23,780 to £19,965 (a decrease of 16%).

Fourth, previous estimates of the cost of a life year have been for individual programmes of care. In this report we have presented estimates of the cost of a life year for an enlarged number of programmes and, with the aid of assumptions about the productivity (health gain) of programmes without a meaningful mortality-based outcome indicator, we have extended our individual programme estimates to incorporate expenditure across all programmes of care. Thus for 2006/7, the cost of a life year for those PBCs with a mortality based outcome indicator is £19,965. If we assume that (a) that the health gain associated with PBC23, which includes primary care and workforce training expenditure, are reflected in the mortality rates for disease specific programmes and (b) that the average health gain across the other programmes without a mortality based outcome indicator is the same as that for those PBCs with a mortality based outcome indicator, then the cost of life year across all programmes is £22,565.

Fifth, we have extended our cost of life year estimates beyond 2006/7. Re-estimation of our model using budgeting expenditure for 2007/8 generates an all programme cost of a life year estimate of £31,846, and re-estimation of our model using budgeting expenditure for 2008/9 generates a similar cost of a life year estimate (£33,333). Together, the last two estimates suggest that there has been step change in the cost of a life year, and that this appears to have occurred between 2006/7 and 2007/8. The cost of a life year estimates are very similar up to and including 2006/7, and they are very similar for 2007/8 and 2008/9. However, there is a substantial difference between the figures for 2004/5, 2005/6 and 2006/7 on the one hand (at about £22k), and for 2007/8 and 2008/9 on the other (at about £33k). The reason for this step change is not obvious but it might be due to changes in the algorithm used by the Department of Health to allocate non-admitted patient care activity to budget categories. Although this allocation rule change has considerably increased the estimated cost of a life year, we believe that this rule change has led to a more accurate allocation of expenditure across PBCs, and that the more recent estimates of the cost of a life year (for 2007/8 and 2008/9) are more accurate than those for the earlier years (for 2005/6 and 2006/7).

Virtually all of the cost of a life year estimates presented in this report are at current prices. However, it is possible to put them on a constant price basis using the Hospital and Community Health Services pay and prices index [41]. For 2006/7, 2007/8 and 2008/9 this index recorded an annual rate of inflation of about 3.5% and so the impact of this constant price adjustment is fairly minimal. For example, if we assume that PBC23 generates a zero health gain and that the gain attributable to the 12 programmes without a mortality indicator is, on average, the same as that attributable to those with a mortality

outcome measure, then the cost of a life year across all programmes at constant 2008/9 prices is £33,333 for 2008/9, £33,088 for 2007/8, and £24,125 for 2006/7.

Finally, although previous results and our current models 'pass' the appropriate statistical tests and, in particular, the Hansen-Sargen test for valid instruments, we are aware that this test might be unable to detect the presence of invalid instruments in some (albeit rather restrictive) circumstances. Responding to this, several studies have suggested that researchers using IV techniques should subject the estimated coefficient on the endogenous variable to a sensitivity analysis. We do precisely this for the outcome equation for each of the big four models. This sensitivity analysis reveals that uncertainty associated with instrument validity has little effect on our estimate of the cost of a life year but it does increase the degree of uncertainty associated with this estimate.

We recognize that this study has a number of limitations. The cost of an additional life year estimates for those programmes with a mortality-based outcome indicator are unadjusted for the quality of life during the additional year; the quoted costs will be an under-estimate of the QALY-adjusted cost of a life year to the extent that additional life years are not in perfect health. In previous studies we have noted that a rudimentary adjustment for this issue using HODaR data increased the cost of a life year by about 50% to 60%.[2, 5]

At the same time, however, the estimated costs will exaggerate the cost of an additional QALY-adjusted year for those programmes with a mortality-based outcome indicator because they ignore any health benefits that are not associated with a reduction in mortality. In other words, expenditure that improves the quality of life (e.g., cancer palliative care) but which does not extend the length of life is implicitly given a zero health gain value.

In addition, the expenditure data relates to expenditure on all patients whereas the mortality data is based on a life expectancy of 75 years. Thus implicitly our calculations attribute a zero health gain to all expenditure on those aged over 75. To illustrate the magnitude of the potential health gain ignored by this restriction, note that in a recent study of costs associated with all inpatient and outpatient activity (excluding mental health), those aged over 75 years accounted for 25% of all costs in 2007/8[34] for details of this study).

Moreover, our cost of a life year estimates are based on the assumption that any Departmental budgetary change falls entirely on PCTs. Although PCTs account for most of the Department of Health's budget, non-PCTs still accounted for 15% of the budget in 2006/7. Because we have no information on how any budgetary change would be split between PCTs and non-PCTs, we have assumed that that any Departmental budgetary change falls entirely on PCTs. If, on the other hand, the non-PCT budget is responsive to changes in the Department's budget then our cost of a life year estimates will be too low. If the non-PCT budget is as responsive as the PCT budget, then our cost of a life year estimate for 2006/7 will be increased by 17.7% (that is, from £22,565 to £26,553).

The results presented in this study are all from the estimation of the relationship between expenditure and mortality using data for a single time period. With the availability of several years of data for both expenditure and mortality, we wanted to estimate a panel data model because a panel can offer advantages over a one period model (e.g., it is better able to handle any unobserved heterogeneity across PCTs). However, most of the instruments employed here are based on the 2001 Census and thus estimation of a panel model will not be possible until these instruments become time variant; this should occur later this year with release of the 2011 Census data at PCT level. This is one piece of work that we intend to pursue in the near future.

References

1. Health, D.o., *A new value-based approach to the pricing of branded medicines: A consultation.* 2010.
2. Martin, S., N. Rice, and P. Smith, *Does health care spending improve health outcomes?* Journal of Health Economics, 2008a: p. 826-842.
3. Martin, S., N. Rice, and P. Smith, *The link between health care spending and health outcomes for the new English Primary Care Trusts.* CHE Research Paper 42, 2008b.
4. Martin, S., N. Rice, and P. Smith, *Panel data estimates of the link between health care spending and health outcomes for English Primary Care Trusts.* Mimeo, 2010.
5. Martin, S., N. Rice, and P. Smith, *Comparing costs and outcomes across programmes of health care.* Health Economics, 2012: p. 316-337.
6. Gerdtham, U. and B. Jonsson, *'International comparisons of health expenditure' in A. Culyer and J. Newhouse (eds).* Handbook of Health Economics, 2000. **Elsevier Amsterdam**.
7. Fisher, E.S. and H.G. Welch, *Avoiding the unintended consequences of growth in medical care - How might more be worse?* Jama-Journal of the American Medical Association, 1999. **281**(5): p. 446-453.
8. Nolte, E. and M. McKee, *Does health care save lives?* The Nuffield Trust, London, 2004.
9. Cochrane, A.L., A.S.S. Leger, and F. Moore, *Health service "input" and mortality "output" in developed countries (Reprinted from Journal of Epidemiology and Community Health vol 32, pg 200-205, 1968).* Journal of Epidemiology and Community Health, 1997. **51**(4): p. 344-348.
10. Young, F.W., *An explanation of the persistent doctor-mortality association.* Journal of Epidemiology and Community Health, 2001. **55**(2): p. 80-84.
11. St Leger, S., *The anomaly that finally went away?* Journal of Epidemiology and Community Health, 2001. **55**(2): p. 79-79.
12. Gravelle, H.S.E. and M.E. Backhouse, *International cross-section analysis of the determination of mortality.* Social Science & Medicine, 1987. **25**(5): p. 427-441.
13. Cremieux, P.Y., P. Ouellette, and C. Pilon, *Health care spending as determinants of health outcomes.* Health Economics, 1999. **8**(7): p. 627-639.
14. Or, Z., *Exploring the effects of health care on mortality across OECD countries.* OECD Labour Market and Social Policy Occasional Paper No 46. OECD, Paris., 2001.
15. Nixon, J. and P. Ulmann, *The relationship between health care expenditure and health outcomes. Evidence and caveats for a causal link.* The European journal of health economics : HEPAC : health economics in prevention and care, 2006. **7**(1): p. 7-18.
16. Health, D.o., *NHS finance manual. December 2005 edition. See http://www.dh.gov.uk/assetRoot/04/13/18/26/04131826.pdf.* 2005a.
17. Smith, P.C., N. Rice, and R. Carr-Hill, *Capitation funding in the public sector.* Journal of the Royal Statistical Society Series a-Statistics in Society, 2001. **164**: p. 217-241.
18. Health, D.o., *Recurrent resource allocations: 2006/07, 2007/08.* Department of Health, London, 2005c.
19. Office, N.A., *Good governance report: review of programme budgeting.* London: NAO, 2008.
20. Appleby, J., Harrison, T., Foot, C., Smith, A. and Gilmour, S. , *Explaining variations in primary care trusts' spending on cancer services.* The King's Fund, London, 2011.
21. Baum, C.F., M.E. Schaffer, and S. Stillman, *ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. See http://ideas.repec.org/c/boc/bocode/s425401.html.* 2010.
22. Shea, J., *Instrumental relevance in multivariate linear models: a simple measure.* Review of Economics and Statistics, 1997. **79**: p. 348-352.
23. Bound, J., D.A. Jaeger, and R.M. Baker, *Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak.* Journal of the American Statistical Association, 1995. **90**(430): p. 443-450.
24. Staiger, D. and J.H. Stock, *Instrumental variables regression with weak instruments.* Econometrica, 1997. **65**(3): p. 557-586.
25. Stock, J.H. and M. Yogo, *Testing for weak instruments in linear IV regression.* NBER Technical Working Paper 284, 2002.
26. Cragg, J.G. and S.G. Donald, *Testing identifiability and specification in instrumental variable models.* Econometric Theory, 1993. **9**(2): p. 222-240.

27. Ramsey, J.B., *Tests for specification errors in classical linear least-squares regression analysis.* Journal of the Royal Statistical Society Series B-Statistical Methodology, 1969. **31**(2): p. 350-&.
28. Pesaran, M.H. and L.W. Taylor, *Diagnostics for IV regressions.* Oxford Bulletin of Economics and Statistics, 1999. **61**(2): p. 255-+.
29. Wooldridge, J., *Econometric analysis of cross section and panel data.* The MIT Press. Cambridge, 2002.
30. Durbin, J., *Errors in variables.* Review of the International Statistical Institute, 1954. **22**: p. 23-32.
31. Health, D.o., *Payment by results: tariff information.* Department of Health, London, 2007.
32. Health, D.o., *PCT recurrent revenue allocations exposition book: 2009/10 and 2010/11.* Department of Health, London, 2009.
33. Health, D.o., *Unified exposition book: 2003/04, 2004/05 and 2005/06 PCT revenue resource limits.* Department of Health, London, 2005b.
34. Dixon, J., et al., *A person based formula for allocating commissioning funds to general practices in England: development of a statistical model.* British Medical Journal, 2011. **343**.
35. Conley, T.G., C.B. Hansen, and P.E. Rossi, *Plausibly exogenous.* Review of Economics and Statistics, 2012. **94**(1): p. 260-272.
36. Small, D.S., *Sensitivity analysis for instrumental variables regression with overidentifying restrictions.* Journal of the American Statistical Association, 2007. **102**(479): p. 1049-1058.
37. Kovandic, T., M. Schaffer, and G. Kleck, *Estimating the causal effect of gun prevalence on homicide rates: a local average treatment effect approach.* IZA Bonn. Discussion paper 3589, 2008.
38. Health, D.o., *Personal communications.* 2012.
39. Martin, S. and P. Smith, *How good at commissioning health are English primary care trusts? A preliminary statistical analysis.* Report to the Health Foundation, 2009.
40. Schaffer, M.E., *Personal communication.* 2011.
41. Curtis, L., *Unit costs of health and social care 2011.* PSSRU, University of Kent., 2011.

Annex

Table BA.1: table showing national (all PCT) expenditure per head (£) and growth in expenditure (%) by PBC group and sub-group, 2003/4 - 2008/9

PBC #	PBC description	Spend (£) per head 2003/4	Spend (£) per head 2004/5	Spend (£) per head 2005/6	Spend (£) per head 2006/7	Spend (£) per head 2007/8	Growth % 2007/8	Spend (£) per head 2008/9	Growth % 2008/9
1	Infectious Diseases	17.95	20.22	23.61	20.88	22.08	6	23.46	6
1a	HIV and AIDS				7.39	8.54	16	10.36	21
1x	Infectious diseases (Other)				13.49	13.54	0	13.10	-3
2	Cancers and Tumours	64.95	75.54	83.24	81.67	90.21	10	94.55	5
2a	Cancer, Head and Neck				2.83	2.65	-6	2.72	3
2b	Cancer, Upper GI				4.05	4.38	8	4.73	8
2c	Cancer, Lower GI				6.46	6.71	4	7.47	11
2d	Cancer, Lung				3.89	4.28	10	4.48	5
2e	Cancer, Skin				1.88	2.05	9	2.05	0
2f	Cancer, Breast				7.39	8.35	13	9.34	12
2g	Cancer, Gynaecological				2.97	2.93	-1	3.05	4
2h	Cancer, Urological				7.76	7.84	1	8.17	4
2i	Cancer, Haematological				8.40	9.22	10	9.47	3
2x	Cancers and Tumours (Other)				36.04	41.79	16	43.07	3
3	Disorders of Blood	14.08	17.00	17.48	16.58	19.44	17	19.50	0
4	Endocrine, Nutritional and Metabolic	28.96	31.86	37.26	36.70	39.39	7	43.38	10
4a	Diabetes				17.76	19.44	9	21.73	12
4b	Endocrine, Nutritional and Metabolic				6.95	7.47	8	7.96	6
4x	Other Endocrine, Nutritional, Metabolic				11.99	12.48	4	13.69	10
5	Mental Health Disorders	133.31	146.83	158.95	166.53	180.90	9	191.21	6
5a	Substance Misuse				13.81	15.76	14	17.81	13
5b	Organic Mental Disorders				14.24	14.83	4	17.39	17
5c	Psychotic Disorders				23.84	31.19	31	33.69	8
5d	Child and Adolescent Mental Health				12.13	12.15	0	13.33	10
5x	Other Mental Health Disorders				102.51	106.97	4	108.99	2
6	Problems of Learning Disability	37.93	43.37	46.54	48.36	54.20	12	56.11	4
7	Neurological	29.83	35.09	41.06	55.27	62.43	13	67.64	8

7a	Chronic Pain				19.31	22.12	15	22.79	3
7x	Neurological (Other)				35.96	40.31	12	44.85	11
8	Problems of Vision	24.61	27.65	28.24	26.97	30.69	14	32.95	7
9	Problems of Hearing	5.73	6.32	6.27	6.21	8.07	30	8.16	1
10	Problems of Circulation	110.12	122.37	124.28	122.06	124.77	2	129.94	4
10a	Coronary Heart Disease				38.91	40.32	4	41.20	2
10b	Cerebrovascular disease				16.05	17.30	8	19.35	12
10c	Problems of Rhythm				7.22	8.21	14	8.43	3
10x	Problems of circulation (Other)				59.88	58.95	-2	60.96	3
11	Problems of the Respiratory System	54.60	62.71	69.56	65.07	67.68	4	77.97	15
11a	Obstructive Airways Disease				10.64	10.64	0	12.70	19
11b	Asthma				14.04	15.73	12	16.99	8
11x	Problems of the respiratory system, other				40.40	41.31	2	48.27	17
12	Dental Problems	10.78	13.55	24.91	51.93	59.45	14	62.44	5
13	Problems of Gastro Intestinal System	63.56	73.22	81.30	73.30	75.05	2	77.89	4
13a	Upper GI				19.88	19.51	-2	19.89	2
13b	Lower GI				20.46	21.92	7	22.63	3
13c	Hepatobiliary				11.26	12.23	9	12.90	5
13x	Problmes of the gastro intestinal system				21.69	21.39	-1	22.46	5
14	Problems of the Skin	20.98	24.90	26.84	28.31	30.41	7	32.34	6
14a	Burns				1.08	1.56	44	1.02	-34
14x	Problems of the Skin				27.23	28.86	6	31.32	9
15	Problems of Musculo Skeletal System	61.36	71.72	74.74	66.75	75.91	14	79.68	5
16	Problems due to Trauma and Injuries	62.31	72.13	76.41	57.29	57.56	0	63.54	10
17	Problems of Genito Urinary System	55.32	62.38	67.38	68.98	67.83	-2	73.78	9
17a	Genital tract problems				19.33	18.80	-3	19.36	3
17b	Renal problems				21.54	19.74	-8	22.29	13
17c	STD				4.26	4.71	10	5.43	15
17x	Problems of Genito Urinary system, other				23.85	24.58	3	26.69	9
18	Maternity and Reproductive Health	52.28	55.04	60.42	57.64	57.09	-1	60.44	6
19	Conditions of Neonates	11.72	13.93	13.42	13.17	15.15	15	17.23	14
20	Adverse effects and poisoning	9.68	12.32	14.25	14.59	15.84	9	18.31	16
20a	Unintended consequences of treatment				10.54	12.14	15	12.96	7
20b	Poisoning				2.13	2.44	15	2.91	19

20c	Violence				0.47	0.49	3	1.75	258
20x	Poisoning and adverse effects				1.45	0.77	-47	0.70	-9
21	Healthy Individuals	20.29	22.77	26.18	26.85	31.44	17	35.74	14
21a	NSF Prevention programme				2.30	3.75	63	4.82	29
21b	NSF Mental health prevention				0.17	0.47	176	0.46	-2
21x	Healthy Individuals (Other)				24.38	27.22	12	30.46	12
22	Social Care Needs	24.81	30.93	33.59	30.29	35.29	17	36.58	4
23	Other	136.94	157.75	171.82	209.70	232.02	11	227.71	-2
23a	GMS/PMS				141.42	147.53	4	145.26	-2
23b	Training (WDCs)				0.60	0.30	-49	0.24	-21
23x	Miscellaneous				67.67	84.19	24	82.20	-2
1 to 23	All PBCs	1052.12	1199.60	1307.76	1345.10	1452.91	8	1530.59	5

Notes:
(i) The population figures for 2003/4, 2004/5 and 2005/6 are identical (the total for England is 49,175,998).
(ii) The corresponding figure for 2006/7 is 50,476,231, for 2007/8 it is 50,695,989, and for 2008/9 it is 51,220,531.
(iii) The spend per head figures are calculated by summing expenditure across all PCTs and dividing by the national population.

Table BA.2: table showing set of socio-economic indicators available as potential instruments in the IV estimation

Indicator name	Short description	Long description
BORNEXEU	Residents born outside the European Union	Residents born outside the European Union divided by all residents (census cell definition: KS005008/KS005001)
WHITEEG	Population in white ethnic group	Population in white ethnic group divided by total population (KS006002+KS006003+KS006004)/KS006001
PCWALLTI	Population of working age with illness	Proportion of population of working age with limiting long term illness aged 16-74 (KS008003/KS09A001)
POPPUCAR	Unpaid care providers in population	Proportion of population providing unpaid care (KS008007/KS008001)
POPPUCA1	Unpaid care (<20 hrs week) in population	Proportion of population providing unpaid care of 1-19 hours a week (KS008008/KS008001)
POPPUCA2	Unpaid care (20-49 hrs) in population	Proportion of population providing unpaid care for 20-49 hours per week (KS008009/KS008001)
POPPUCA3	Unpaid care (>50 hrs week) in population	Proportion of population providing unpaid care for over 50 hours week (KS008007/KS008001)
NQUAL174	Proportion aged 16-74 with no qualifications	Proportion of population aged 16-74 with no qualifications (KS013002/KS013001
FTSTUDEN	Proportion aged 16-74 full-time students	Proportion of population aged 16-74 that are full-time students ((KS013008+KS013009)/KS013001)
HHNOCAR	Households without a car	Proportion of households without a car (KS017002/KS017001)
OWNOCC	Owner occupied households	Proportion of households that are owner occupied (KS018002+KS018003+KS018004)/KS018001)
LAHARENT	Rented social housing	Proportion of households that are rented from LA or HA ((KS018005+KS018006)/KS018001)
PRIVRENT	Rented private housing	Proportion of households that are rented from private landlords (KS018007/KS018001)
LONEPENH	Lone pensioner households	Proportion of households that are one pensioner households (KS020002/KS020001)
LONEPARH	Lone parent households	Proportion of households that are lone parent households with dependent children (KS020011/KS020001)
PERMSICK	Permanently sick of those aged 16-74	Proportion of population aged 16-74 that are permanently sick (KS09A010/KS09A001)
PC74LTUN	Long-term unemployed of those aged 16-74	Proportion of those aged 16-74 that are long-term unemployed (KS09A015/KS09A001)
WORKAGRI	Employed in agriculture	Proportion of those aged 16-74 in employment that are working agriculture (KS11A002/KS11A001)
PROFOCCU	People in professional occupations	Proportion of those aged 16-74 in managerial and professional occupations ((KS14A002+KS14A003+KS14A004)/KS14A001)

Table BA.3: table showing first stage regressions for outcome models associated with 2005/6 expenditure and mortality data for 2002/3/4

	(1)	(2)	(3)	(4)	(5)	(6)
	PBC 2	PBC 10	PBC 11	PBC 13	PBC 16	PBC 19
	cancer	circulation	respiratory	gastro-intestinal	trauma	neonates
	2005/6	2005/6	2005/6	2005/6	2005/6	2005/6
	outcome model	outcome model	outcome model	outcome model	outcome model	outcome model
	instrument spend	instrument spend	instrument spend	instrument spend	instrument spend	instrument spend
	unweighted	unweighted	unweighted	unweighted	unweighted	unweighted
VARIABLES	first stage	first stage	first stage	first stage	first stage	first stage
need per head	0.406***	1.173***	1.533***	0.970***	0.727**	
	[0.097]	[0.235]	[0.401]	[0.243]	[0.289]	
lone pensioner households	0.593***	0.229***	-0.118	0.045	0.561***	
	[0.109]	[0.084]	[0.112]	[0.093]	[0.108]	
provision of unpaid care	-0.013	0.374***		0.574***	-0.148	
	[0.135]	[0.115]		[0.089]	[0.132]	
IMD 2000		-0.152***	-0.247***	-0.047	-0.016	
		[0.056]	[0.069]	[0.060]	[0.074]	
white ethnic group		-0.007				
		[0.067]				
permanently sick			0.192**			
			[0.085]			
low birth weight births						0.393
						[0.308]
lone parent households						0.034
						[0.209]
no qualifications						-0.599***
						[0.148]
long-term unemployed						0.394***
						[0.122]
LA/HA rented accommodation						0.283**
						[0.127]
Constant	-1.373***	-0.311	-1.562***	-0.959***	-1.780***	-2.797***
	[0.304]	[0.244]	[0.344]	[0.192]	[0.260]	[0.433]
Observations	295	295	295	295	295	294
R-squared	0.297	0.629	0.434	0.571	0.396	0.197

Note: these are the first-stage regressions for the IV models reported in Table B7.4. Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table BA.4: table showing first stage regressions for expenditure models associated with 2005/6 expenditure and mortality data for 2002/3/4

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	PBC 2	PBC 10	PBC 11	PBC 13	PBC 7	PBC 16	PBC 23
	cancer	circulation	respiratory	gastro-intestinal	neurological	trauma	GMS/PMS
	2005/6	2005/6	2005/6	2005/6	2005/6	2005/6	2005/6
	spend model instrument o/need	spend model instrument o/need	spend model instrument o/need	spend model instrument o/need	spend model instrument o/need	spend model instrument o/need	spend model instrument o/need
	unweighted	unweighted	unweighted	unweighted	unweighted	unweighted	unweighted
VARIABLES	first stage	first stage	first stage	first stage	first stage	first stage	first stage
no qualifications							0.240***
							[0.038]
lone pensioner households	-0.686***	-0.244***	-0.234***	-0.266***	-0.234***	-0.234***	-0.129***
	[0.067]	[0.052]	[0.049]	[0.049]	[0.049]	[0.049]	[0.038]
private rented housing							0.072***
							[0.017]
work in agriculture							-0.006
							[0.008]
PCT budget per head	-0.146	-0.003	-0.077	-0.022	-0.077	-0.077	0.043
	[0.117]	[0.074]	[0.070]	[0.070]	[0.070]	[0.070]	[0.069]
no car households							0.092**
							[0.039]
lone parent households							0.171***
							[0.035]
permanently sick							0.125***
							[0.027]
need per head	1.933***	0.651***	0.875***	0.597***	0.875***	0.875***	
	[0.110]	[0.157]	[0.175]	[0.157]	[0.175]	[0.175]	
white ethnic group		0.197***					
		[0.038]					
provision of unpaid care	-0.371***	-0.153**	-0.217***		-0.217***	-0.217***	
	[0.071]	[0.065]	[0.055]		[0.055]	[0.055]	
IMD 2000		0.056*	0.128***	0.179***	0.128***	0.128***	
		[0.033]	[0.038]	[0.035]	[0.038]	[0.038]	
Constant	2.562***	4.103***	4.851***	5.114***	4.851***	4.851***	7.361***
	[0.159]	[0.140]	[0.109]	[0.082]	[0.109]	[0.109]	[0.139]
Observations	295	295	295	295	295	295	295
R-squared	0.804	0.680	0.858	0.849	0.858	0.858	0.881

Note: these are the first-stage regressions for the IV models reported in Table B7.5. Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table BA.5: table showing first stage regressions for outcome and expenditure models associated with 2006/7 expenditure and mortality data for 2004/5/6

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	PBC 2	PBC 2	PBC 10	PBC 10	PBC 11	PBC 11	PBC 13	PBC 13
	cancer	cancer	circulation	circulation	respiratory	respiratory	gastro	gastro
	2006/7	2006/7	2006/7	2006/7	2006/7	2006/7	2006/7	2006/7
	outcome model	spend model	outcome model	spend model	outcome model	spend model	outcome model	spend model
	uses SYLLR instrument spend	uses SYLLR instrument o/calls	uses SYLLR instrument spend	uses SYLLR instrument o/calls	uses SYLLR instrument spend	uses SYLLR instrument o/calls	uses SYLLR instrument spend	uses SYLLR instrument o/calls
	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted
	first stage	first stage	first stage	first stage	first stage	first stage	first stage	first stage
	CARAN need	CARAN need	CARAN need	CARAN need	CARAN need	CARAN need	CARAN need	CARAN need
VARIABLES	2 MFFs	2 MFFs	2 MFFs	2 MFFs	2 MFFs	2 MFFs	2 MFFs	2 MFFs
need CARAN per head	1.162***	1.602***	1.539***	0.606***	1.026***	0.836***	1.292***	0.938***
	[0.250]	[0.126]	[0.323]	[0.141]	[0.368]	[0.175]	[0.358]	[0.167]
need CARAN per head squared	0.912							
	[0.666]							
lone pensioner households	0.383***	-0.431***	0.321***	-0.221***				
	[0.134]	[0.073]	[0.111]	[0.067]				
IMD 2007	-0.153**		-0.247***	0.117***		0.104**	-0.115	0.107***
	[0.074]		[0.087]	[0.037]		[0.043]	[0.094]	[0.041]
PCT budget per head		0.120		0.183*		0.077		-0.020
		[0.124]		[0.094]		[0.084]		[0.093]
provision of unpaid care		-0.410***	0.097			-0.309***	0.373*	-0.325***
		[0.088]	[0.197]			[0.090]	[0.215]	[0.078]
white ethnic group			-0.060					
			[0.082]					
permanently sick					0.681**			
					[0.269]			
long-term unemployed					-0.123***			
					[0.035]			
limiting long-term illness					-0.785*			
					[0.449]			
Constant	5.586***	3.074***	6.387***	3.790***	3.906***	4.501***	5.496***	5.119***
	[0.235]	[0.887]	[0.363]	[0.703]	[0.474]	[0.595]	[0.314]	[0.657]
Observations	152	152	152	152	152	152	152	152
R-squared	0.438	0.846	0.623	0.814	0.623	0.840	0.554	0.827

Note: these are the first-stage regressions for the IV models reported in Table B8.16. Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table BA.6: table showing first stage regressions for outcome and expenditure models associated with 2006/7 expenditure and mortality data for 2006/7/8

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	PBC 2	PBC 2	PBC 10	PBC 10	PBC 11	PBC 11	PBC 13	PBC 13
	cancer	cancer	circulation	circulation	respiratory	respiratory	gastro	gastro
	2006/7	2006/7	2006/7	2006/7	2006/7	2006/7	2006/7	2006/7
	outcome model	spend model	outcome model	spend model	outcome model	spend model	outcome model	spend model
	SYLLR 2006/8 instrument spend	SYLLR 2006/8 instrument o/calls	SYLLR 2006/8 instrument spend	SYLLR 2006/8 instrument o/calls	SYLLR 2006/8 instrument spend	SYLLR 2006/8 instrument o/calls	SYLLR 2006/8 instrument spend	SYLLR 2006/8 instrument o/calls
	weighted	weighted	weighted	weighted	weighted	weighted	weighted	weighted
	first stage	first stage	first stage	first stage	first stage	first stage	first stage	first stage
VARIABLES	2 MFFs	2 MFFs	2 MFFs	2 MFFs	2 MFFs	2 MFFs	2 MFFs	2 MFFs
need CARAN per head	1.162***	1.574***	1.539***	0.791***	1.061***	0.909***	1.292***	1.059***
	[0.250]	[0.138]	[0.323]	[0.157]	[0.386]	[0.167]	[0.358]	[0.166]
need CARAN p/head squ	0.912				0.455			
	[0.666]				[0.599]			
lone pensioner household	0.383***	-0.375***	0.321***	-0.269***				-0.313***
	[0.134]	[0.079]	[0.111]	[0.067]				[0.072]
IMD 2007	-0.153**		-0.247***	0.097**		0.107***	-0.115	0.066
	[0.074]		[0.087]	[0.039]		[0.041]	[0.094]	[0.040]
PCT budget per head		0.126		0.128		0.020		0.040
		[0.136]		[0.101]		[0.090]		[0.091]
provision of unpaid care		-0.386***	0.097			-0.289***	0.373*	-0.203**
		[0.097]	[0.197]			[0.080]	[0.215]	[0.088]
white ethnic group			-0.060					
			[0.082]					
permanently sick					0.677**			
					[0.272]			
long-term unemployed					-0.121***			
					[0.035]			
limiting long-term illness					-0.798*			
					[0.454]			
Constant	5.586***	3.160***	6.387***	4.132***	3.864***	4.916***	5.496***	4.481***
	[0.235]	[0.963]	[0.363]	[0.729]	[0.493]	[0.637]	[0.314]	[0.654]
Observations	152	152	152	152	152	152	152	152
R-squared	0.438	0.821	0.623	0.823	0.624	0.831	0.554	0.857

Note: these are the first-stage regressions for the IV models reported in Table B8.19. Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table BA.7: table showing first stage regressions for outcome models associated with 2007/8 expenditure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	PBC 2 cancer 2007/8 instrument spend weighted	PBC 2 cancer 2007/8 instrument spend weighted	PBC 10 circulation 2007/8 instrument spend weighted	PBC 11 respiratory 2007/8 instrument spend weighted	PBC 11 respiratory 2007/8 instrument spend weighted	PBC 13 gastro-intestinal 2007/8 instrument spend weighted	PBC 13 gastro-intestinal 2007/8 instrument spend weighted	PBC 4 endocrine 2007/8 instrument spend weighted	PBC 17 genito-urinary 2007/8 instrument spend weighted	PBC 1 infectious disease 2007/8 instrument spend weighted	PBC 1819 Maternity &neonates 2007/8 instrument spend weighted	PBC 16 Trauma &injuries 2007/8 instrument spend weighted
VARIABLES	first stage	first stage	first stage	first stage	first stage	first stage	first stage	first stage	first stage	first stage	first stage	first stage
need CARAN per head	0.582**	0.545***	0.724***	1.251***	1.196***	0.999***	1.047***				1.111***	0.856***
	[0.284]	[0.105]	[0.168]	[0.094]	[0.113]	[0.105]	[0.099]				[0.259]	[0.282]
no car households										0.512*		0.288**
										[0.267]		[0.137]
lone pensioner households	0.632***	0.644***	0.468***	0.360***	0.269***	0.199						
	[0.148]	[0.119]	[0.100]	[0.103]	[0.100]	[0.133]						
IMD 2007	-0.012							-0.067		0.325		
	[0.088]							[0.053]		[0.224]		
need CARAN per head squ				1.332***	1.338***							
				[0.428]	[0.425]							
provision of unpaid care			0.441**		0.200							
			[0.174]		[0.160]							
born outside EU						-0.054***	-0.067***				0.004	-0.079**
						[0.018]	[0.017]				[0.041]	[0.039]
diabetes prevalence rate 2007/8								0.358***				
								[0.123]				
permanently sick								0.307***				
								[0.061]				
lone parent households									0.029			
									[0.124]			
CKD prevalence rate 2007/8									0.123**			
									[0.061]			
long-term unemployed									0.146**			
									[0.061]			
limiting long-term illness									0.207			
									[0.134]			
HIV need per head squared										0.128***		
										[0.031]		
HIV need per head										0.300***		
										[0.044]		
work in agriculture										0.152**		0.126***
										[0.064]		[0.033]
work in professional occupation										0.647***		
										[0.172]		
no qualifications											-0.214	
											[0.160]	
maternity need per head											0.647***	
											[0.159]	
full-time students												0.126
												[0.095]
LA/HA accommodation												-0.197*
												[0.104]
Constant	5.759***	5.747***	6.754***	4.886***	5.172***	4.531***	4.104***	4.224***	5.269***	4.226***	4.004***	4.766***
	[0.252]	[0.234]	[0.322]	[0.201]	[0.350]	[0.281]	[0.057]	[0.390]	[0.209]	[1.007]	[0.297]	[0.339]
Observations	151	151	151	151	151	151	151	151	147	151	151	151
R-squared	0.369	0.369	0.653	0.659	0.664	0.531	0.524	0.436	0.296	0.724	0.407	0.361

Note: these are the first-stage regressions for the IV models reported in Table B10.1. Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table BA.8: table showing first stage regressions for expenditure models associated with 2007/8 expenditure

	(1) PBC 2 cancer 2007/8 spend model instrument o/need weighted	(2) PBC 10 circulation 2007/8 spend model instrument o/need weighted	(3) PBC 11 respiratory 2007/8 spend model instrument o/need weighted	(4) PBC 13 gastro-intestinal 2007/8 spend model instrument o/need weighted	(5) PBC 4 Endocrine 2007/8 spend model instrument o/need weighted	(6) PBC 7 neurological 2007/8 spend model instrument o/need weighted	(7) PBC 23 GMS/PMS etc 2007/8 spend model instrument o/need weighted	(8) PBC 16 trauma & injuries 2007/8 spend model instrument o/need weighted
VARIABLES	first stage	first stage	first stage	first stage	first stage	first stage	first stage	first stage
PCT budget per head 7/8	0.071	0.066	0.039	-0.034	0.057	0.170	0.360**	0.366***
	[0.137]	[0.123]	[0.105]	[0.105]	[0.121]	[0.130]	[0.144]	[0.136]
need CARAN per head	1.613***	1.201***	1.050***	0.971***	1.090***	1.148***		
	[0.149]	[0.136]	[0.191]	[0.188]	[0.198]	[0.137]		
need CARAN p/head square			0.343					
			[0.266]					
lone pensioner households	-0.357***	-0.220***	-0.261***		-0.274***	-0.444***	-0.255***	-0.229***
	[0.067]	[0.060]	[0.063]		[0.063]	[0.055]	[0.061]	[0.072]
provision of unpaid care	-0.362***	-0.215**	-0.156*	-0.296***	-0.181*			0.195**
	[0.094]	[0.090]	[0.093]	[0.086]	[0.100]			[0.092]
IMD 2007			0.070	0.099**	0.067		0.309***	0.276***
			[0.042]	[0.043]	[0.045]		[0.043]	[0.039]
diabetes prevalence rate 2007/8					0.008			
					[0.069]			
epilepsy prevalence rate 2007/8						0.020		
						[0.049]		
white ethnic group							0.221***	
							[0.060]	
work in agriculture								0.009
								[0.011]
Constant	3.624***	4.454***	4.667***	5.298***	4.496***	3.974***	2.054**	2.630***
	[0.969]	[0.848]	[0.735]	[0.724]	[0.920]	[0.954]	[0.994]	[0.978]
Observations	151	151	151	151	151	151	151	151
R-squared	0.847	0.828	0.861	0.834	0.860	0.839	0.830	0.824

Note: these are the first-stage regressions for the IV models reported in Table B10.2. Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table BA.9: table showing first stage regressions for outcome models associated with 2008/9 expenditure

Regressors	(1) PBC 2 cancer 2008/9 outcome model instrument spend weighted first stage	(2) PBC 10 circulation 2008/9 outcome model instrument spend weighted first stage	(3) PBC 11 respiratory 2008/9 outcome model instrument spend weighted first stage	(4) PBC 13 gastro-intestinal 2008/9 outcome model instrument spend weighted first stage	(5) PBC 13 gastro-intestinal 2008/9 outcome model instrument spend weighted first stage	(6) PBC 4 endocrine 2008/9 outcome model instrument spend weighted first stage	(7) PBC 7 neurological 2008/9 outcome model instrument spend weighted first stage	(8) PBC 17 genito-urinary 2008/9 outcome model instrument spend weighted first stage	(9) PBC 1 infectious disease 2008/9 outcome model instrument spend weighted first stage	(10) PBC 1819 maternity/neonates 2008/9 outcome model instrument spend weighted first stage
need CARAN per head	1.122***	1.274***	1.228***	0.989***	1.056***		0.373**			0.659***
	[0.198]	[0.146]	[0.085]	[0.088]	[0.087]		[0.175]			[0.236]
lone pensioner households	0.490***	0.426***	0.252**	0.272**			0.287***			
	[0.127]	[0.090]	[0.110]	[0.109]			[0.109]			
IMD 2007	-0.145**					-0.082			0.236	
	[0.060]					[0.091]			[0.214]	
no car households		-0.188***							0.502**	
		[0.053]							[0.227]	
need CARAN per head sq			1.071**							
			[0.426]							
provision of unpaid care			0.339***			0.539**			1.393***	
			[0.117]			[0.232]			[0.340]	
born outside EU				-0.042***	-0.060***	0.080***				-0.031
				[0.016]	[0.015]	[0.026]				[0.032]
diabetes prevalence rate 2007/8						0.167				
						[0.132]				
permanently sick						0.380***				
						[0.104]				
epilepsy prevalence rate 2007/8							0.486***			
							[0.121]			
owner occupied households							-0.235**			
							[0.113]			
lone parent households								0.175**		0.013
								[0.080]		[0.106]
CKD prevalence rate 2007/8								0.089***		
								[0.033]		
long-term unemployed								0.148***		
								[0.045]		
HIV need per head									0.471***	
									[0.050]	
HIV need per head squared									0.146***	
									[0.027]	
no qualifications									-0.751***	-0.092
									[0.189]	[0.113]
work in agriculture									0.150***	
									[0.051]	
maternity need per head										0.834***
										[0.162]
Constant	5.937***	5.435***	5.610***	4.752***	4.167***	6.379***	4.808***	5.363***	6.010***	4.171***
	[0.221]	[0.230]	[0.238]	[0.236]	[0.048]	[0.768]	[0.193]	[0.121]	[1.513]	[0.375]
Observations	151	151	151	151	151	151	151	148	151	151
R-squared	0.521	0.612	0.746	0.665	0.648	0.559	0.477	0.378	0.791	0.614

Note: these are the first-stage regressions for the IV models reported in Table B11.1. Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table BA.10: table showing first stage regressions for expenditure models associated with 2008/9 expenditure

VARIABLES	(1) PBC 2 cancer 2008/9 spend model instrument o/need weighted first stage	(2) PBC 10 circulatory 2008/9 spend model instrument o/need weighted first stage	(3) PBC 11 respiratory 2008/9 spend model instrument o/need weighted first stage	(4) PBC 13 gastro-intestinal 2008/9 spend model instrument o/need weighted first stage	(5) PBC 4 endocrine 2008/9 spend model instrument o/need weighted first stage	(6) PBC 7 neurological 2008/9 spend model instrument o/need weighted first stage	(7) PBC 1819 maternity/neonates 2008/9 spend model instrument o/need weighted first stage
PCT budget per head 8/9	0.090	0.049	0.020	-0.115	0.055	0.180	0.452***
	[0.155]	[0.118]	[0.113]	[0.106]	[0.122]	[0.132]	[0.128]
need CARAN per head	1.589***	1.215***	1.305***	1.042***	1.112***	1.119***	
	[0.168]	[0.133]	[0.129]	[0.207]	[0.211]	[0.141]	
need CARAN per head sq			0.221				
			[0.270]				
lone pensioner households	-0.371***	-0.209***	-0.286***		-0.256***	-0.453***	-0.106*
	[0.071]	[0.060]	[0.057]		[0.062]	[0.053]	[0.059]
provision of unpaid care	-0.349***	-0.236**	-0.266***	-0.302***	-0.239**		
	[0.105]	[0.092]	[0.088]	[0.099]	[0.119]		
IMD 2007				0.099**	0.051		0.223***
				[0.045]	[0.048]		[0.044]
diabetes prevalence rate 2007/8					0.067		
					[0.064]		
epilepsy prevalence rate 2007/8						0.036	
						[0.044]	
maternity need per head							0.266***
							[0.078]
white ethnic group							0.292***
							[0.065]
Constant	3.457***	4.528***	4.687***	5.851***	4.340***	3.858***	1.901**
	[1.131]	[0.849]	[0.814]	[0.753]	[0.940]	[0.979]	[0.864]
Observations	151	151	151	151	151	151	151
R-squared	0.840	0.828	0.854	0.830	0.857	0.837	0.843

Note: these are the first-stage regressions for the IV models reported in Table B11.2. Robust standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1.

Appendix C

Translating mortality effects into life years and quality adjusted life years

C.1 Introduction

This Appendix describes how the results of the econometric work undertaken to estimate the link between NHS spending and mortality, which was detailed in Appendix B, can be translated in to effects on life years and quality adjusted life years (QALYs). This Appendix presents much of the detail of data and analyses that support Chapter 4 of the main report.

We present three sequential steps of analysis which lead to estimates of the overall cost per QALY threshold for the NHS:

i. In section C.2.1 we reconsider how the estimated effects on mortality from the econometrics work might better translate into life years by exploring the limitations of mortality data available at PCT level and the published years of life lost (YLL) figures presented in the previous chapter. We explore how these estimates might be improved using additional data and analysis.
ii. In section C.2.2 we consider how these estimates of life year effects might be adjusted for the quality of life in which they are lived, taking account of the gender and the age at which life years are gained or lost as well as the disutility associated with particular diseases.
iii. In section C.2.3 we explore ways to also take account of those effects on health not directly associated with mortality and life year effects (i.e., the 'pure' quality of life effects) to estimate an overall cost per QALY threshold.

This sequence of analysis is set out and explained based on the analysis of 2006/07 expenditure and mortality data from 2006 to 2008. At the end of each section, we present a summary which includes a central 'best' estimate as well as extreme lower and upper bounds for the cost per life year and cost per QALY threshold. The core assumptions which underpin these three values are common across sections C.2.1 to C.2.4. The central or 'best' estimate is based on two assumptions one conservative and the other more optimistic with respect to the health effects associated with expenditure. The first is that the health effects of changes in one year of expenditure are restricted to one year. Analyses in Appendix B uses 3 years of mortality data, but these are averaged to an annual value prior to estimating outcome elasticities. Therefore, the estimated outcome elasticities represent the proportionate effect on mortality in one year due to a proportionate change in expenditure. This is likely to underestimate effects on mortality since expenditure that reduces mortality risk for an individual in one year may well also reduce their risk over subsequent years; possibly over the whole of their remaining disease duration. Expenditure may also prevent disease in future patient populations. Therefore, total health effects will be underestimated and the cost per life year or QALY threshold will be overestimated. Although undoubtedly conservative, it may be offset to some extent by the more optimistic assumption used to translate mortality effects into life years. In common with YLL figures published by NHS IC and the WHO Global Burden of Disease study it is assumed that any death averted by expenditure in one year will return the individual to the mortality risk of the general population, i.e., the years of life gained associated with each death averted are based on what would have been their life expectancy taking account of their of age and gender (using life tables for the general population).

The extreme upper and lower bounds for cost per life year and cost per QALY thresholds are based on making both assumptions either optimistic (providing the lower bound for the threshold) or both conservative (an upper bound for the threshold). The lower bound is based on assuming that health effects are not restricted to one year but apply to the remaining disease duration for the population at risk during the expenditure year (although this still does not account for the effects of expenditure on preventing disease). The upper bound is based on the combination of assuming that health effects are restricted to one year and that any death averted is only averted for the minimum duration consistent with the mortality data used to estimate the outcome elasticities in Appendix B. It is very important to note that the lower and upper bounds represent extreme values rather than alternative but plausible views that could reasonably be taken.

The three sequential steps of analysis, which provide a cost per life year threshold, through a cost per life year adjusted for quality to a cost per QALY threshold, are explained and detailed in sections C.2.1 to C.2.3, using the analysis of 2006 expenditure and mortality data from 2006 to 2008. In Section C.2.4, further analysis using these data highlight which PBCs have the greatest influence on the overall threshold. An exploration of the impact of the uncertainty over the outcome and spend elasticities in estimates of the threshold is also presented in Section C.2.5. The sequence of analyses is then applied to 2008/09 expenditure and 2008 to 2010 mortality data; results of the cost per QALY threshold for the most recent years of analysis are presented in Section C.3. In Section C.4 we present our best estimate of the threshold cost per QALY based on the analysis of 2007/8 expenditure and mortality data from 2007 to 2009.

C.2 Analysis of 2006/07 expenditure and 2006 to 2008 mortality data

C.2.1 From mortality to life years

In this section we summarise our examination of a number of issues associated with available PCT-based mortality data and the associated published estimates of YLL. We then examine how, given the limited information available about the population at risk in each PBC, we might take proper account of the fact that some of the observed deaths would have occurred anyway (had the same population not been at risk in the particular PBC) when estimating YLL, i.e., taking account of unobserved counterfactual deaths. This allows us to estimate the YLL that better reflects the effect of expenditure on the mortality observed in each PBC, and infer the excess deaths associated with each PBC. Finally we present cost per death averted and cost per life year which accounts for the issues raised in this section.

C.2.1.1 Mortality and YLL coverage

The mortality data that is available at PCT level does not offer full coverage of all deaths across all the ICDs that make up each PBC. Table C.1 illustrates, using a few PBCs as examples, the mapping of three digit ICD-10 to PBCs (column 1) and the incomplete coverage of these ICDs in mortality data (column 2). A more detailed account of the extent of coverage is presented in Table B.5.1 in Appendix B.

Table C.1. Illustrating coverage

PBC		ICD codes covered by the spend data [1]	ICD codes covered by the mortality data (NHS IC) [2]	Coverage of mortality data relative to spend data (2008) [3]
1	Infectious diseases	large parts of A00-B99	A00-B99	1.000
2	Cancer	C00-C97, D00-D49	C00-C97	0.984
4	Endocrine	E000-E899	E10-E14	0.634
10	Circulatory	I00-I99, Q20-Q28	I00-I99	0.992
11	Respiratory	A150-A169,* A190-A199, J000-J989, Q300-Q349, R000-R099	J12-J18, J40-J44, J45-J46	0.773

National (English) data are, however, available that cover all deaths associated with all the ICDs that make up each PBC. Therefore, it is possible to adjust the incomplete reporting of mortality at PCT level (see section B5.1 in Appendix B) before applying the estimated outcome elasticities to calculate the deaths averted due to expenditure. Applying published estimates of YLL per death to all the deaths averted using coverage adjustment factors (as illustrated in column 3 of Table C.1) provides the estimate of the cost per life year reported in Appendix B. Note that the proportionate effects on mortality (due to changes in expenditure) are therefore assumed to be similar for mortality that is and is not recorded at PCT level. This seems more reasonable than assuming no effect of expenditure on mortality that happens not to be recorded at PCT level.

The published estimates of YLL (NHS IC) used in Chapter 3 only include deaths below 75 years (but exclude deaths below 1 year) and are based on the difference between age 75 and the age of each death below 75. These estimates have the same limited coverage as PCT level mortality data, so are not available for all the ICDs that make up each PBC. Therefore, applying the available estimates of YLL per death to the estimated number of deaths averted requires an assumption that the YLL per death is similar for those groups of ICDs covered and not covered by the published YLL figures.

This can be examined by using national ONS data to calculate YLL in the same way as NHS IC, but with full coverage of all the ICDs that make up each PBC. Although ONS data provides complete coverage and reports gender, age at death is only reported in 5 year ranges (these data are not available at PCT level so could not be used when estimating outcome elasticities in Chapter 3). Therefore, using ONS data to

estimate YLL requires taking the midpoint[1] of each range as the age of death, i.e., assuming reported deaths are equally likely over the range in which they are reported. For this reason it is not possible to precisely recover the published YLL figures using ONS data for those ICD groupings that can be precisely matched to the NHS IC coverage. However, the differences are small (ranging from -1% to 2% as shown in Table C.2 below), suggesting that taking the midpoint of each range as the age of death is a reasonable approximation.

Table C.2. Estimates of YLL for NHS IC and ONS for those ICD groupings that can be precisely matched to the NHS IC coverage

PBC		$YLL_{<75}$ (NHS IC)* [1]	$YLL_{<75}$ (ONS) + [2]	Difference in YLL [3]
1	Infectious diseases	35,517	35,688	0.5%
2	Cancer	735,674	744,240	1%
4	Endocrine problems	19,224	19,445	1%
10	Circulatory	453,878	461,062	2%
18+19	Maternity & neonates	164,200	163,105	-1%

* does not take into account coverage adjustment

\+ deaths age<1 included in PBC 18+19

Published estimates of YLL are available from NHS IC for PBC16 (Trauma and injuries) but ONS does not provide the information required to calculate YLL for this PBC. The estimated outcome elasticity for PBC 16 (Trauma and injuries) was zero for 2006 and could not be estimated for 2008 expenditure. Therefore, this PBC does not contribute any changes in health outcomes, although the changes in this expenditure are included in subsequent estimates of cost per life year and QALY thresholds. However, there was a very limited coverage of mortality data recorded at PCT level and the expenditure data for this PBC. In addition, the mortality data that was available (ICDs S72, S02, S06 and T90) was less likely to be associated with changes expenditure in this PBC and more likely to be associated with changes in expenditure in others. Consequently the health effects of changes in expenditure in PBC 16 may be underestimated. Therefore, this PBC does not contribute any changes in health outcomes due to changes in expenditure in subsequent estimates of cost per life year and QALY thresholds.

The differences between estimates of YLL based on ONS and NHS IC data are, however, much more significant and are reported in Table C.3. These reflect differences in the distribution of ages at death between those groups of ICDs covered and not covered in the NHS IC figures. For example, NHS IC figures available at PCT level for PBC7 (neurological problems) have low coverage of all deaths in this PBC (0.136 in column 1). The deaths that are reported in NHS IC are associated with epilepsy and the YLL (22,046 in column 2) reflects the generally younger age at death in this group. When adjusted for full coverage (22,046/0.136 = 162,100 in column 3) the estimated YLL is much greater than the YLL based directly on all deaths by age group reported for the PBC in ONS. This difference in YLL reflects the fact that the deaths in PBC7 which are not covered by NHS IC figures tend to be in older age groups so generate fewer YLL.

[1] The calculated midpoints are as follows,

Age range	< 1	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44
midpoint	0.5	3.0	7.5	12.5	17.5	22.5	27.5	32.5	37.5	42.5
Age range	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90+
midpoint	47.5	52.5	57.5	62.5	67.5	72.5	77.5	82.5	87.5	92.5

Table C.3. Estimates of YLL for NHS IC and ONS

PBC		Coverage of mortality data relative to spend data [1]	$YLL_{<75}$ (NHS IC) [2]	$YLL_{<75}$ adjusted (NHS IC) [3]	$YLL_{<75}$ no adjustment needed (ONS) [4]	Difference from adjusted NHS IC to ONS [5]
1	Infectious diseases	1.000	35,517	35,517	40,928	15%
2	Cancer	0.984	735,674	747,636	758,804	1%
4	Endocrine problems	0.634	19,224	30,322	41,548	37%
7	Neurological problems	0.136	22,046	162,100	93,755	-42%
10	Circulatory	0.992	453,878	457,538	481,246	5%
11	Respiratory	0.773	108,074	139,812	147,465	6%
13	Gastro-intestinal	0.571	115,303	201,931	177,532	-12%
17	Genito-urinary	0.172	3,343	19,438	17,380	-11%
18+19	Maternity & neonates*	0.679	164,200	241,826	15,409	-94%

Using ONS data also allows deaths under the age of 1 year to be appropriately assigned to PBCs via the ICD in which they occurred (NHS IC YLL figures exclude deaths under one year), rather than assigning them all to PBC18 & 19 as in Appendix B.[2] This explains the large reduction in YLL for PBC18 & 19 (Maternity and neonates) as much of the mortality is re-assigned to ICDs which contribute to other PBCs. Since most of the deaths that are re-assigned are allocated to PBC1 (infectious diseases) the YLL for this PBC increases despite complete reporting of deaths at PCT level and full coverage by NHS IC figures (see Table C.4).

Table C.4. Estimates of YLL for NHS IC and ONS including deaths age <1

PBC	$YLL_{<75}$ (from NHS IC) [1]	$YLL_{<75}$ (from ONS)* [2]	Difference in YLL [3]
1	35,517	40,928	15%
2	735,674	744,960	1%
4	19,224	19,445	1%
10	453,878	464,763	2%
18+19	164,200	15,409†	-91%

* deaths age<1 included in PBC of death
† does not include YLL from deaths <28 days

Using ONS data to calculate YLL in the same way as the published NHS IC figures, but overcoming some of the issues associated with the reporting of mortality at PCT level and the coverage of published estimates of YLL, generates similar estimates of a cost per life year threshold (see column 2 Table C.6) to those reported in Appendix B.

C.2.1.2 Life expectancy and YLL

As noted above, the NHS IC estimates of YLL only include deaths below 75 years and are based on the difference between age 75 and the age of each death below 75. Implicitly, this treats 75 as the appropriate normal life expectancy for males and females for the population at risk in each PBC. However, with the exception of maternity and neonates, most deaths in PBCs occur above the age of 75 and life expectancies are significantly greater than 75. Based on 2006 to 2008 data, life expectancy at birth is greater than 75 (77.74 for males and 81.88 for females).[3] Given the need to reflect the normal life expectancy for the at risk population, it is more appropriate to use the age distribution of the general population, and calculate life expectancy conditional on age averaged over the general population's age

[2] The YLL available from NHS IC represented all deaths from maternity and all deaths under 28 days across PBCs. The coverage factor (0.68) adjusts this YLL to represent maternity and all deaths < 1year across PBCs. The calculation is described in Appendix B, footnote (v) of Table B5.1.
[3] Figures for England, from http://www.ons.gov.uk/ons/rel/subnational-health4/life-expec-at-birth-age-65/2004-06-to-2008-10/statistical-bulletin.html#tab-National-life-expectancy.

distribution. General population life expectancies are estimated to be 80.7 for males and 84.4 for females, These life expectancy estimates will always be higher than life expectancies at birth.

Based on ONS data, YLLs can be re-recalculated using the above estimates of gender specific life expectancy for the general population. When increasing life expectancy (LE) two effects occur, both of which tend to increase estimates of YLL. Firstly, more deaths are included in the YLL calculation (those that occur between age 75 and LE) and secondly, each death previously counted below 75 will generate 5.7 or 9.4 more YLL for males and females respectively. The effect on the number of deaths and the YLL for each PBC of using the life expectancy of the general population is reported in Table C.5.

Table C.5. The difference in YLL by life expectancy

PBC		Deaths$_{<75}$ (ONS) [1]	Deaths$_{<LE}$ (ONS) [2]	Difference in deaths due to increased LE [3]	YLL$_{<75}$ (ONS) [4]	YLL$_{<LE}$ (ONS) [5]	Difference in YLL due to increased LE [6]
1	Infectious diseases	2,050	3,710	81%	40,928	62,051	52%
2	Cancer	62,944	95,212	51%	758,804	1,345,013	77%
4	Endocrine	2,367	4,000	69%	41,548	65,015	56%
7	Neurological	5,095	8,975	76%	93,755	145,526	55%
10	Circulatory	41,487	82,098	98%	481,246	916,170	90%
11	Respiratory	14,000	30,500	118%	147,465	310,326	110%
13	Gastro-intestinal	10,611	15,827	49%	177,532	273,303	54%
17	Genito-urinary	1,588	4,197	164%	17,380	39,098	125%
18+19	Maternity & neonates	226	226	0%	15,409	17,167	11%

Life expectancy (LE): male=80.7, female=84.4

The number of deaths counted below LE increases for every PBC except for maternity & neonates because, as expected, all deaths are below age 75 in PBC18 & 19. However, YLL increases for all PBCs reflecting the additional years otherwise expected to be lived to an older LE. Of course including more of the deaths observed in each PBC and the greater YLL associated with them will generate more deaths averted and more life years gained when applying the same proportionate effects from the outcome elasticities estimated in Appendix B. Therefore, the cost per death averted and cost per life year thresholds are expected to be lower using these figures than those reported in Appendix B.

The impact on the cost per life year and cost per death averted thresholds is summarised in Table C.6. A detailed breakdown of the changes in spend and YLL across PBCs is presented in Table C.7. A listing of the spend and outcome elasticities used in threshold calculations throughout this section is in Table C.8.

Table C.6: Summary of cost per death averted and cost per life year threshold.

	Using 75 as the cut-off (ONS)		Using LE as the cut-off (ONS)	
	Cost per death averted [1]	Cost per LY gained [2]	Cost per death averted [3]	Cost per LY gained [4]
big 4 PBC's	£122,756	£10,398	£63,426	£5,487
11 PBCs (with mortality)	£240,433	£20,031	£124,655	£10,660
All 23 PBCs (zero health effects for remaining 12 PBCs)	£884,579	£73,697	£458,620	£39,218
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£271,739	£22,639	£140,886	£12,048

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

Table C.7: Breakdown of the cost per death averted and cost per life year thresholds

			Using 75 as the cut-off (ONS)						Using LE as the cut-off (ONS)					
PBC	PBC description	Change in spend, £m	N death (<75)	Change in N deaths	Cost per death averted, £	YLL	Change in YLL	Cost per LY gained, £	N deaths (<LE)	Change in N deaths	Cost per death averted, £	YLL	Change in YLL	Cost per LY gained, £
2	Cancer	£19	62 944	100.10	£191,500	758 804	1 207	£15,885	95 212	151.42	£126,599	1 345 013	2 139	£8,962
10	Circulatory	£33	41 487	321.26	£103,560	481 246	3 727	£8,928	82 098	635.73	£52,333	916 170	7 094	£4,690
11	Respiratory	£22	14 000	249.25	£89,482	147 465	2 625	£8,495	30 500	543.	£41,074	310 326	5 525	£4,037
13	Gastro-intestinal	£17	10 611	72.69	£227,013	177 532	1 216	£13,568	15 827	108.42	£152,198	273 303	1 872	£8,814
	Big 4				£122,756			£10,398			£63,426			£5,487
1	Infectious diseases	£8	2 050	0.76	£10,936,680	40 928	15	£547,796	3 710	1.38	£6,043,179	62 051	23	£361,319
4	Endocrine	£18	2 367	18.99	£929,559	41 548	333	£52,957	4 000	32.1	£550,066	65 015	522	£33,842
7	Neurological	£17	5 095	3.52	£4,889,114	93 755	65	£265,693	8 975	6.19	£2,775,491	145 526	100	£171,172
17	Genito-urinary	£32	1 588	0.74	£42,993,075	17 380	8	£3,928,251	4 197	1.95	£16,267,096	39 098	18	£1,746,202
16	Trauma & injuries*	£10	NA	0.00	NA	NA	0	NA	NA	0	NA	NA	0	NA
18+19	Maternity & neonates*	£8	226	0.24	£32,813,038	15 409	17	£481,261	226	0.24	£32,813,038	17 167	19	£431,977
	First 11 PBC's				£240,433			£20,031			£124,655			£10,660
3	Disorders of Blood	£11		46.57	£240,433		559	£20,031		89.83	£124,655		1 050	£10,660
5	Mental Health	£204		849.17	£240,433		10 193	£20,031		1 637.87	£124,655		19 153	£10,660
6	Learning Disability	£31		128.05	£240,433		1 537	£20,031		246.98	£124,655		2 888	£10,660
8	Vision	£24		100.54	£240,433		1 207	£20,031		193.92	£124,655		2 268	£10,660
9	Hearing	£6		26.60	£240,433		319	£20,031		51.3	£124,655		600	£10,660
12	Dental	£23		97.72	£240,433		1 173	£20,031		188.48	£124,655		2 204	£10,660
14	Skin	£11		43.72	£240,433		525	£20,031		84.34	£124,655		986	£10,660
15	Musculo skeletal	£15		62.93	£240,433		755	£20,031		121.38	£124,655		1 419	£10,660
20	Poisoning and AE	£4		18.27	£240,433		219	£20,031		35.23	£124,655		412	£10,660
21	Healthy Individuals	£18		76.27	£240,433		915	£20,031		147.1	£124,655		1 720	£10,660
22	Social Care Needs	£68		281.19	£240,433		3 375	£20,031		542.35	£124,655		6 342	£10,660
23	Other	£78		0	NA			NA			NA			NA
	All (23 PBCs)				£271,739			£22,639			£140,886			£12,048

Note that we have been unable to obtain a satisfactory outcome model for trauma & injuries and have assumed a zero outcome elasticity.

Note that, for expenditure in 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

Table C.8: Outcome and spend elasticities.

PBC	PBC description	Total spend 2006/07, (£) [1]	Spend elasticities unadjusted[i] [2]	Spend elasticities adjusted [2]	Outcome elasticities[ii] [3]
2	Cancer	£4,122	0.465	0.465	0.342
10	Circulatory problems	£6,161	0.540	0.540	1.434
11	Respiratory problems	£3,285	0.679	0.679	2.622
13	Gastro-intestinal problems	£3,700	0.446	0.446	1.536
	Big 4	£17,268			
1	Infectious diseases	£1,054	0.792	0.792	0.047
4	Endocrine problems	£1,853	0.953	0.953	0.842
7	Neurological problems	£2,790	0.616	0.616	0.112
17	Genito-urinary problems	£3,482	0.912	0.912	0.051
16	Trauma & injuries*	£2,892	0.358	0.358	-
18+19	Maternity & neonates*	£3,574	0.224	0.224	0.482
	First 11 PBC's	£32,912			
3	Disorders of Blood	£837	0.700	1.338	-
5	Mental Health Disorders	£8,406	1.271	2.429	-
6	Learning Disability	£2,441	0.660	1.261	-
8	Vision	£1,362	0.929	1.775	-
9	Hearing	£314	1.067	2.039	-
12	Dental problems	£2,621	0.469	0.896	-
14	Problems of the Skin	£1,429	0.385	0.736	-
15	Musculo-skeletal system	£3,369	0.235	0.449	-
20	Poisoning and AE	£737	0.312	0.596	-
21	Healthy Individuals	£1,355	0.708	1.353	-
22	Social Care Needs	£1,529	2.314	4.422	-
23	Other	£10,585	0.739	0.739	-
	All (23 PBCs)	£67,896			

[i] The spend elasticities reflect how a 1% increase in budget is distributed across PBCs; however, in the econometrics, these were estimated separately for each PBC (unadjusted estimates in column 2) and because of this, its direct application to spend generates a change in budget bigger than the 1%. An adjustment was thus applied to the remaining 12 PBC's (except PBC23 that was left unchanged), by multiplying each by a common factor – the magnitude of the unadjusted spend elasticities is changed but proportionality to the original elasticities is maintained

[ii] without the negative sign

The cost per death averted (or life saved) threshold should not be over interpreted because this is of little direct policy interest since lives are never saved (death is only delayed) and the significance of a death averted depends critically on how long it is averted and the quality of life in which additional years are lived (see Section C.2.2). However, establishing the number of deaths averted which are associated with net YLL is useful because it enables an assessment of the number of life years gained associated with each death averted. Table C.9 presents the YLL saved for each death averted implied by the assumptions underlying calculations of the cost per life year threshold in Table C.7. For the 11 PBCs with mortality signal, each death averted is assumed to be associated with a gain of 11.7 YLL (when LE is used, column 2). This value is smaller than when using 75 years old as a cut-off (column 1) because a higher proportion of deaths closer to the cut-off age are being considered (i.e., with lower YLL associated).

Table C.9: Implied YLL per death averted for each PBC

PBC	PBC description	Implied YLL per death averted (<75) [1]	Implied YLL per death averted (<LE) [2]
2	Cancer	12.1	14.1
10	Circulatory problems	11.6	11.2
11	Respiratory problems	10.5	10.2
13	Gastro-intestinal problems	16.7	17.3
	Big 4	11.8	11.6
1	Infectious diseases	20.0	16.7
4	Endocrine problems	17.6	16.3
7	Neurological problems	18.4	16.2
17	Genito-urinary problems	10.9	9.3
16	Trauma & injuries*	NA	NA
18+19	Maternity & neonates*	68.2	76.0
	First 11 PBC's	12.0	11.7

There are good reasons why YLL figures calculated as the difference between age of death and LE are likely to be overestimated. This is dealt with in the next section (Section C.2.1.3). In Section C.2.1.4 we take account of the fact that some of the deaths observed in a PBC would have occurred anyway in a similar 'normal' population (i.e., the counterfactual population not at risk through membership of the PBC) so not all observed deaths are 'excess' and generate YLL.

C.2.1.3 YLL and accounting for counterfactual deaths

The estimates of YLL based on ONS data overcome many of the limitations of the published NHS IC figures. However, the YLLs reported in Table C.5 are calculated in the same way as the NHS IC figures, by taking the difference between a fixed LE and the age at death for deaths observed below that LE. Simply taking the difference between a fixed LE and the age at death of deaths that occur below LE and ignoring those deaths that occur above LE, is only an accurate representation of the YLL if it is reasonable to assume that no deaths would have otherwise occurred prior to LE (so all 'normal' deaths must occur at LE) and that there are no deaths (survivors) beyond LE in the at risk population, i.e. all deaths below LE are excess deaths and there are no excess deaths above LE. The estimate of YLL in the previous section may thus be biased for two reasons: i) it does not account for the fact that not all deaths observed below LE are 'excess' deaths in the sense that some deaths would have occurred (at the same age) in a similar population not at risk in the PBC and ii) some of the deaths observed above LE may be 'excess' deaths that would not otherwise have occurred at that age (see breakdown of deaths below and above LE in Table C.10).

The overall effect on YLL, and on the cost per life year, will depend on the number of deaths above and below LE that are excess. However, it is more likely that deaths below LE are 'excess'. Estimates of YLL are required which take account of the 'counterfactual' deaths that would have occurred even if the population in the PBC was not at risk through membership of the ICD codes that make it up, but faced the same mortality risks as the general population, accounting for the age and gender distribution of the PBC population.

Table C.10. Number of deaths below and above LE in 2006/07/08, by PBC

PBC		<LE 2006 [1]	>LE 2006 [2]	<LE 2007 [3]	>LE 2007 [4]	<LE 2008 [5]	>LE 2008 [6]	Annual N deaths <LE [7]	Annual N deaths > LE [8]
1	Infectious diseases	3,824	3,420	3,902	3,735	3,403	2,589	3,710	3,248
2	Cancer	95,549	34,192	95,331	35,455	94,758	37,144	95,213	35,597
4	Endocrine	4,006	2,661	3,967	2,750	4,028	2,882	4,000	2,764
7	Neurological	8,454	5,762	8,845	6,501	9,626	6,871	8,975	6,378
10	Circulatory	84,909	78,369	80,610	78,481	80,779	76,407	82,099	77,752
11	Respiratory	29,925	34,549	29,540	35,060	32,036	35,227	30,500	34,945
13	Gastro-intestinal	15,893	8,311	15,658	8,376	15,930	8,274	15,827	8,320
17	Genito-urinary	4,056	6,049	4,072	6,558	4,465	6,673	4,198	6,427
18+19	Maternity & neonates	195	0	216	0	267	0	226	0

Life expectancy (LE): male=80.7, female=84.4

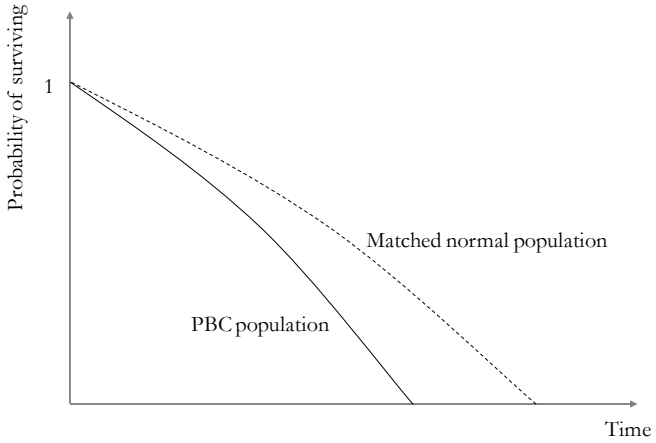
Ideally, with reliable information about the size of the population at risk in each PBC and its age and gender distribution it would be possible to estimate the number of deaths that would be expected to occur had this population not been at risk, based on mortality data for the general population. The difference between deaths observed across all ages in the PBC and the deaths expected to have occurred in this matched 'normal' population would provide the number of 'excess' deaths by age and gender. These 'counterfactual' deaths will occur in the other PBCs insofar as all deaths are recorded in an ICD code, taking account of the unavoidable fact that everyone must die of something at some time. For

example, even if all observed cancer mortality was avoidable and could in principle be eliminated with sufficient expenditure, lives would not be 'saved' but deaths delayed and reallocated to other causes.[4]

The YLL associated with each of these excess deaths is the life expectancy conditional on gender and on surviving to the age at which the excess death occurred. The total YLL for the at risk population is simply the sum of these YLLs over all excess deaths, which could occur at any age. We do not (and will never) know the counterfactual expected age of death for each individual patient. However, two perfectly matched populations of individuals, one at risk and another not at risk in the PBC can be compared in terms of their survival curves (Figure C.1). The area below each survival curve reflects the life expectancy and the area between the two survival curves returns the YLL. This is equivalent to comparing the average age of death across patients in the population at risk in the PBC (*N* patients), with the average age of death in the matched, not at risk, population (for simplicity assumed to be equally sized). Equation (1) describes the YLL per patient as the difference in the average age of death, age_{death} , observed for each individual, *i* (out of *N* individuals), in each population. The YLL for the population is simply the per patient YLL multiplied by the size of the population N.

$$\begin{aligned} YLL_{per\ patient} &= LE^{\text{norm}} - LE^{\text{PBC}} = \frac{1}{N}\sum_{i=1}^{\text{N}} \text{age}_{death,i}^{\text{norm}} - \frac{1}{N}\sum_{i=1}^{\text{N}} \text{age}_{death,i}^{\text{PBC}} \\ YLL &= N \cdot YLL_{per\ patient} \end{aligned} \quad (1)$$

Figure C.1: Survival curve of a population at risk in a PBC and of a matched 'normal' population



The difficultly is that routinely available data do not provide any information about the size of the population at risk or its age and gender distribution (matching criteria). Thus a matched population cannot be generated, and the area between the two curves cannot be evaluated. Therefore, it is not possible to directly estimate excess deaths or compare survival curves. Even if the size of the at risk population is unknown we can still use information that might be available about its age and gender distribution (or make reasonable assumptions) to estimate a matched 'normal' LE using life tables for the general population – such a LE summarises the area under the counterfactual survival curve ($LE^{norm} = \frac{1}{N}\sum_{i=1}^{\text{N}} \text{age}_{death,i}^{\text{norm}}$ in equation 1). Unfortunately, it is not possible to also calculate the LE for the population at risk in the PBC (or represent the survival curve) without information about the size of the at risk population – if it was possible, the difference between these life expectancies would approximate the YLL per patient at risk in a PBC.

Fortunately, we can still recover a consistent estimate of YLL using the normal LE of a matched population that is not at risk (a summary of the counterfactual average age of death), alongside the death

[4] Note that the outcome elasticities are based on PBC mortality that is sensitive to changes in expenditure (i.e., is avoidable) at the margin so no assumptions about how much of the PBC mortality is avoidable is required.

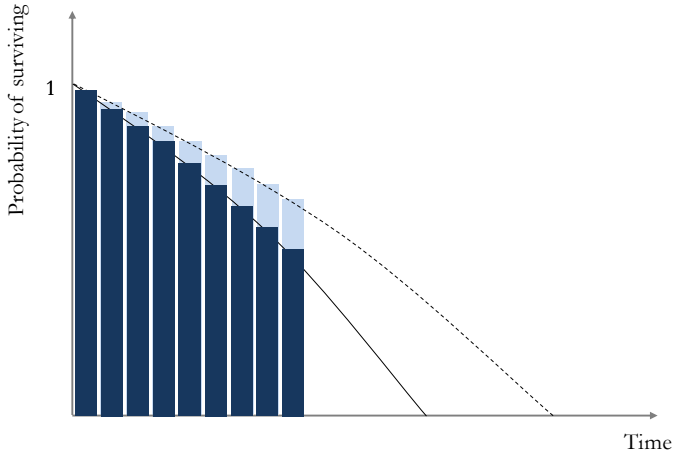
data available for the PBC population. Equation (2) shows that population YLL can be approximated by subtracting the age at which each observed death in a PBC has occurred to the normal LE.

$$\mathrm{YLL} = \mathrm{N} \cdot \left(\mathrm{LE}^{\mathrm{norm}} - \frac{1}{N}\sum_{i=1}^{\mathrm{N}} \mathrm{age}_{\mathrm{death,i}}^{\mathrm{PBC}}\right) = \sum_{i=1}^{\mathrm{N}} \left(\mathrm{LE}^{\mathrm{norm}} - \mathrm{age}_{\mathrm{death,i}}^{\mathrm{PBC}}\right) \quad (2)$$

The data on the PBC observed deaths available expresses the ages at which deaths occurred in age groups (*k* out of *K* groups). Following from Equation (2), the population YLL can be evaluated considering the number of patients dying in each of the age groups, $\mathrm{N}_{\mathrm{die},k}$, as depicted in Equation (3). This is equivalent to comparing survival curves where age is discretized into intervals and the midpoint of the intervals used as age of death – this is illustrated in Figure C.2.

$$\mathrm{YLL} = \sum_{k=1}^{K} \left(\mathrm{LE}^{\mathrm{norm}} - \mathrm{age}_{death,k}\right) \cdot \mathrm{N}_{\mathrm{die},k} \quad (3)$$

Figure C.2: Area between the survival curves, discretized.



The calculations (in Equations 2 and 3) require all observed deaths — both those that occur below and those that occur above this LE — to be taken into account. Those deaths occurring below LE generate YLL - compared to the average of a matched population not at risk. However, we must also account for those deaths that occur at ages above LE. These deaths generate life years 'gained' (YLG) compared to the average of a matched population not at risk. Therefore, the appropriate estimate is a net YLL (i.e., YLL – YLG). In effect, by subtracting YLG from YLL we take account of the fact that not all deaths below LE are excess deaths but some deaths above LE are. Insofar as deaths above LE have been observed in a specific PBC, the net YLL estimate will always be lower than the estimate of YLL. Consequently, the estimates in Section C.2.1.3 overestimate YLL and hence underestimate the cost per life year threshold.

Using the life expectancy of the general population

Routinely available data provides the age and gender of observed deaths but no information about the age and gender distribution of the at risk population itself. Using observed age and gender at death (see columns 5 and 6 of Table C.11) as an indication of the distribution of the at risk population will significantly overestimate the LE of a normal matched population insofar as a disease may be chronic (not all PBC mortality occurs on entry into the at risk population). If mortality risk increases over the disease duration more deaths would be observed in groups that have been prevalent for some time (i.e., are older) than those that are incident. Older age groups will thus be overrepresented in observed deaths compared to a matched normal population. For these reasons LE and YLL would be overestimated using age at death as a proxy for the age distribution of the at risk population, and the cost per life year would be underestimated.

Table C.11: Average age and life expectancy for PBCs based on age of the general population

PBC		Sex	Average age of general population [1]	LE of general population [2]	Age at death [3]	LE at age of death [4]
1	Infectious diseases	m	38.5	80.7	72.8	87.5
		f	40.8	84.4	79.3	91.1
2	Cancer	m	38.5	80.7	73.3	86.5
		f	40.8	84.4	73.8	88.8
4	Endocrine	m	38.5	80.7	72.5	87.1
		f	40.8	84.4	77.9	90.6
7	Neurological	m	38.5	80.7	72.8	87.2
		f	40.8	84.4	77.7	90.5
10	Circulatory	m	38.5	80.7	76.4	87.9
		f	40.8	84.4	82.7	91.7
11	Respiratory	m	38.5	80.7	79.4	89.0
		f	40.8	84.4	82.9	91.8
13	Gastro-intestinal	m	38.5	80.7	68.9	85.7
		f	40.8	84.4	77.1	90.1
17	Genito-urinary	m	38.5	80.7	81.6	90.1
		f	40.8	84.4	84.0	92.3
18+19	Maternity & neonates	m	38.5	80.7	1.1	78.3
		f	40.8	84.4	11.4	82.7

In the absence of additional external information the net YLL could be based on the life expectancy of the general population, reflecting its current age and gender distribution. Such net YLL estimates are reported in Table C.12, and illustrate the impact of accounting for counterfactual deaths in the way described above. The YLL reported in column 5 of Table C.12 are calculated the same way and are the same as the figures previously reported (column 5 of Table C.5). That is, they do not account for deaths that would have otherwise occurred below LE or the very many deaths that occur above LE. With the exception of PBC18 & 19 many death occur above the LE of the general population (see column 4 in Table C.12) in all PBCs. As a consequence, there are LYG associated with all other PBCs (see column 6) so the net YLL in column 7 are lower than YLL based on the same life expectancy. Therefore, failure to account for counterfactual deaths would lead to an overestimate of the YLL associated with a PBC and the effects of expenditure on YLL. The cost per life year threshold would be underestimated (see Table C.15).

Table C.12. Net YLL using life expectancy of the general population

PBC		Average2006-2008						
		LE of Males [1]	LE of Females [2]	Deaths <LE [3]	Deaths >LE [4]	YLL [5]	YLG [6]	Net YLL [7]
1	Infectious diseases	80.7	84.4	3,710	3,248	62,052	18,796	43,256
2	Cancer	80.7	84.4	95,213	35,597	1,345,038	175,350	1,169,689
4	Endocrine	80.7	84.4	4,000	2,764	65,016	15,864	49,152
7	Neurological	80.7	84.4	8,975	6,378	145,529	34,621	110,908
10	Circulatory	80.7	84.4	82,099	77,752	916,192	444,694	471,498
11	Respiratory	80.7	84.4	30,500	34,945	310,334	215,829	94,505
13	Gastro-intestinal	80.7	84.4	15,827	8,320	273,308	45,295	228,012
17	Genito-urinary	80.7	84.4	4,198	6,427	39,099	40,530	-1,431
18+19	Maternity & neonates	80.7	84.4	226	0	17,167	0	17,167

However, these figures are only correct insofar as the distribution of age and gender in each PBC is similar to the general population. For example, if the at risk population tends to be younger the correct LE for the PBC will be lower. A lower LE will mean that there are less deaths below LE each generating fewer YLL, and more deaths above LE each generating more LYG. The net YLL will thus tend to be

lower. Similarly, if the at risk population tends to be older than the general population, the correct LE will be higher and net YLL will also tend to be higher.

This explains the apparent net gain in YLL (negative net YLL) for PBC17 (Genito-urinary) where most deaths occur at ages greater than the LE of the general population so that LYG exceeds YLL. As we are able to show later (see Table C.14) this is because the age distribution in this PBC tends to be older than the general population, i.e., the LE for a matched normal population should be higher with fewer deaths above and more below this LE.

Using additional information about age and gender distribution

It is evident that estimates of YLL require some account to be taken of counterfactual deaths. In the absence of routinely available information this requires examination of alternative sources of information which might provide a basis for more credible assumptions about the age and gender distribution of the at risk population in each PBC than either, the distribution of observed deaths or of the general population.[5] The WHO Global Burden of Disease (GBD) study, updated in 2008 using 2004 data (see Addendum 1 for more details[6]) provides a range of summary health indicators for the UK, which are, in part, based on estimates of the incidence and duration of sequelae associated with different types of disease by age and gender. Therefore, the type of information used by WHO in the GBD Study to generate summary estimates for the UK can also be used to improve the assumptions required about the age and gender distribution of the PBC populations. Importantly, at this stage, we do not need to rely on estimates of the absolute size of the at risk population, but only the relative 'share' by age and gender.

Specifically, the information reported by GBD (estimates specific to the UK provided in the National Burden of Disease toolkit) reported the incidence and duration of sequelae associated with different types of disease by age and gender. Since it is possible that a patient may experience more than one of the types of sequelae reported in GBD we use the gender and age distribution of the sequelae with the highest prevalence, i.e., the minimum estimate of prevalence consistent with these figures (Addendum 1 to this Appendix), to evaluate the age and gender distribution within each disease.

GBD classifies diseases by U-codes, which are groups of three digit ICD-10 codes (see Addendum 1 to this Appendix for details of how U-codes map to ICD-10 codes).[7] Since we know which ICD codes contribute to each PBC we can map information from U-codes to PBCs via the ICD codes that contribute to each. The resulting average age and life expectancy for each PBC is reported in columns 3 and 4 of Table C.13 using the information available from GBD in combination with life tables for the general population.

---

[5] Although this research was not funded to purchase access to GPRD data we were able to examine a sample of it which comprised of 22,313,086 rows/patient–ICD10 events (3 digit) representing 4,229,910 patients with data on new diagnosis of diseases observed between 1 Jan 2006 and 24 June 2011 (see Addendum 1). Although GPRD data could, in principle, provide this type of information the difficulties of reliability, face validity and interpretation of the sample data in the form available to us meant that it was not directly useful.

[6] We are aware that the 2000-2002 WHO GBD study and the update which was published in 2008 using 2004 data has itself recently been updated. However, the report and tools where not publically available at the time this research was conducted.

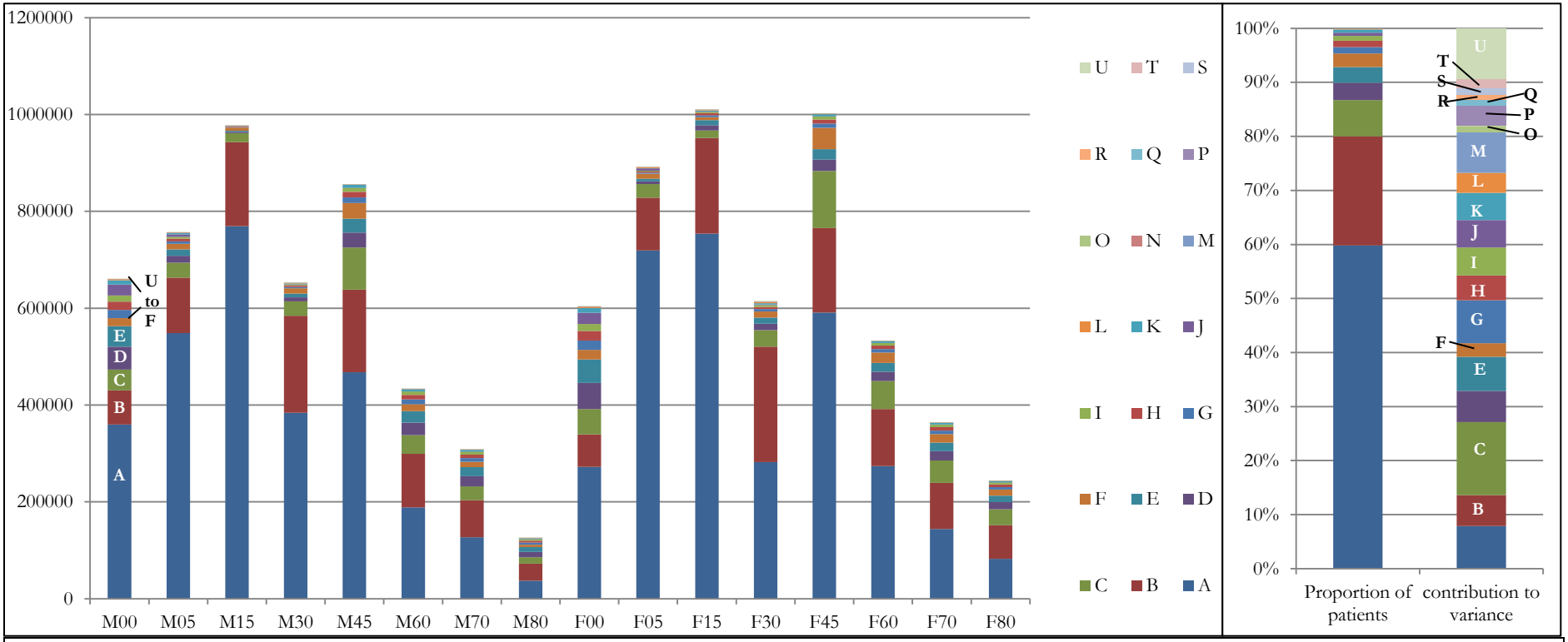
[7] Throughout the analyses in this Appendix, mortality, life years and QALY were not assigned to procedural ICD codes (i.e. those in ICD chapter Z, Factors influencing health status and contact with health services). Health effects from increased spending on these ICD codes would either be non-existant or would be evident in other ICD codes related to the procedure. . This means that no health effects are associated with PBC 22 Social Care (which only includes procedural ICD codes), although changes in expenditure on PBC 22 are included. This is likely to overestimate the threshold because any health effects associated with PBC 22 will not be reflected in the estimated outcome elasticities of other PBCs unless the effects happen to be correlated with changes in expenditure in those other PBCs.

Table C.13. Average age and life expectancy for PBCs based on GBD

PBC		Sex	Average age of general population [1]	LE of general population [2]	Proportion males in PBC (GBD) [3]	Average age in PBC (GBD) [4]	Normal LE of PBC population (GBD) [5]
1	Infectious diseases	m	38.5	80.7	54.1%	28.6	79.6
		f	40.8	84.4	45.9%	30.2	83.6
2	Cancer	m	38.5	80.7	28.0%	61.3	83.0
		f	40.8	84.4	72.0%	52.3	84.7
4	Endocrine	m	38.5	80.7	38.4%	44.2	81.0
		f	40.8	84.4	61.6%	50.8	84.7
7	Neurological	m	38.5	80.7	28.1%	24.8	79.6
		f	40.8	84.4	71.9%	23.5	83.3
10	Circulatory	m	38.5	80.7	51.6%	55.4	83.0
		f	40.8	84.4	48.4%	57.9	86.5
11	Respiratory	m	38.5	80.7	48.0%	32.1	80.3
		f	40.8	84.4	52.0%	33.7	84.0
13	Gastro-intestinal	m	38.5	80.7	42.9%	35.8	80.6
		f	40.8	84.4	57.1%	41.9	84.5
17	Genito-urinary	m	38.5	80.7	85.9%	63.2	83.5
		f	40.8	84.4	14.1%	47.3	85.6
18+19	Maternity & neonates	m	38.5	80.7	16.3%	3.0	78.7
		f	40.8	84.4	83.7%	24.1	83.1

These summary estimates suggest that some of the PBC populations may be on average older than the general population (e.g., Cancer, Circulatory and Genito-urinary PBCs) or younger (e.g., Maternity & neonates, Infectious diseases and Neurological). However, when trying to interpret these summaries it should be noted that the average age reported in Table C.13 is the average over the ages at which sequelae occur within the ICDs contributing to the PBC. Therefore, a similar average age can reflect very different age distributions. Some reflect a markedly bimodal distribution, e.g., PBC11, Respiratory, where there is high incidence at very young and older ages (see Figure C.3), or very different age distributions across the type of diseases that contribute to the PBC. For example PBC7 (Neurological) includes dementia which accounts for the vast majority of the PBC population older than 70. However, a greater proportion of the population is in much younger age groups with other conditions, especially migraine (see Addendum 1 to this Appendix for a detailed description of age and gender distributions in all PBCs). When interpreting these summary estimates it should also be noted that the reported life expectancies are not the life expectancies at the average ages reported in column 3 (Table C.13), but the average over the life expectancies for each age group within the contributing ICDs weighted by the age distribution of the sequelae with maximum prevalence from GBD U-codes.

Figure C.3: Distribution of PBC11 prevalence by age, gender and contributing ICDs, alongside proportion of prevalent patients in the PBC and contribution to variance of each ICD



A- Chronic lower respiratory diseases (J40-J47)
B- Acute upper respiratory infections (J00-J06)
C- Influenza and Pneumonia (J09-J18)
D- Lung diseases due to external agents (J60-J70)
E- Other diseases of upper respiratory tract (J30-J39)
F- Other acute lower respiratory infections (J20-J22)
G- Other respiratory diseases principally affecting the interstitium (J80-J84)
H- Other diseases of pleura (J90-J94)
I- Other diseases of the respiratory system (J95-J99)
J- Other bacterial diseases (A30-A49)
K- Suppurative and necrotic conditions of lower respiratory tract (J85-J86)
L- Mycoses (B35-B49)
M- Tuberculosis (A15-A19)
N- Other diseases casused by chlamydiae (A70-A74)
O- Other viral diseases (B25-B34)
P- Symptoms and signs circulatory and respiratory systems (R00-R09)
Q- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)
R- Persons encountering health services for examination and investigation (Z00-Z13)
S- Persons encountering health services for specific procedures and health care (Z40-Z54)
T- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)
U- Congenital malformations and deformations respiratory system (Q30-Q34)

The implications for net YLL of using these PBC specific estimates of 'normal' life expectancy are reported in Table C.14. As expected, the net YLL for those PBCs with a LE greater than the general population are higher than those reported in column 7 in Table C.12 (e.g., PBC10 Circulatory and PBC17 Genito-urinary, which now has positive net YLL). Similarly those PBCs with a LE less than the general population have lower net YLL than reported in column 7 in Table C.12 (e.g., PBC1 Infectious diseases and PBC18 & 19 Maternity & neonates, where the effect of a lower LE is more modest as there are no deaths above either of the estimates of LE).

Table C.14. Net YLL using life expectancy for each PBC

PBC		LE of Males [1]	LE of Females [2]	Average2006-2008 Deaths <LE [3]	Deaths >LE [4]	YLL [5]	YLG [6]	Net YLL [7]
1	Infectious diseases	79.6	83.6	3,498	3,460	58,686	21,724	36,962
2	Cancer	83.0	84.7	101,203	29,607	1,473,733	126,549	1,347,184
4	Endocrine	81.0	84.7	4,068	2,696	66,283	15,058	51,225
7	Neurological	79.6	83.3	8,370	6,983	135,686	41,770	93,917
10	Circulatory	83.0	86.5	96,694	63,157	1,102,020	278,251	823,768
11	Respiratory	80.3	84.0	29,549	35,897	298,343	230,313	68,030
13	Gastro-intestinal	80.6	84.5	15,824	8,323	273,117	45,414	227,703
17	Genito-urinary	83.5	85.6	4,969	5,655	47,229	29,101	18,127
18+19	Maternity & neonates	78.7	83.1	226	0	16,801	0	16,801

The impact on the cost per life year threshold of the issues discussed in this Section is summarised in columns 3 and 4 of Table C.15.

Table C.15. Summary of cost per life year threshold

	Using cut-off in estimating YLL (ONS)		Using net YLL estimates	
	cut-off of 75 [1]	cut-off of LE of the GP [2]	Using LE of the GP [3]	Using LE of the PBC population (GBD) [4]
big 4 PBC's	£10,398	£5,487	£10,421	£8,080
11 PBCs (with mortality)	£20,031	£10,660	£19,928	£15,628
All 23 PBCs (zero health effects for remaining 12 PBCs)	£73,697	£39,218	£73,317	£57,497
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£22,639	£12,048	£22,523	£17,663

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

Taking account of counterfactual deaths by calculating net YLL based on the life expectancy of the general population (see column 3) provides similar estimates to those reported in Appendix B. Assuming that PBC populations have the same age and gender distribution as the general population when the, albeit limited, information that is available suggests otherwise, seems inappropriate. Therefore, our preferred central estimate of the cost per life year threshold is reported in column 4 (Table C.15). These are lower than those based on the general population, reflecting the impact on net YLL of evidence that the population at risk in some key PBCs (especially PBC2 and 10) tend to be older than the general population. A detailed breakdown of the changes in spend and YLLs across PBCs that originate this central estimate are presented in columns 5 to 7 of Table C.16. In Section C.2.1.5 we consider extreme upper and lower bounds that might be placed on this central estimate.

Table C.16: LY threshold using net YLL estimates (non-zero health effects for remaining PBCs except GMS).

PBC	PBC description	Change in spend, £m [1]	Using LE of the GP: Net YLL [2]	Change in Net YLL [3]	Cost per LY gained, £ [4]	Using LE of the PBC population: Net YLL [5]	Change in net YLL [6]	Cost per LY gained, £ [7]
2	Cancer	£19	1 169 689	1 860	£10,305	1 347 184	2 142	£8,947
10	Circulatory problems	£33	471 498	3 651	£9,112	823 768	6 379	£5,216
11	Respiratory problems	£22	94 505	1 683	£13,256	68 030	1 211	£18,415
13	Gastro-intestinal problems	£17	228 012	1 562	£10,564	227 703	1 560	£10,579
	Big 4				£10,421			£8,080
1	Infectious diseases	£8	43 256	16	£518,314	36 962	14	£606,574
4	Endocrine problems	£18	49 152	394	£44,765	51 225	411	£42,953
7	Neurological problems	£17	110 908	77	£224,601	93 917	65	£265,235
17	Genito-urinary problems	£32	- 1 431	- 1	-£47,709,995	18 127	8	£3,766,371
16	Trauma & injuries*	£10	NA	0	NA	NA	0	NA
18+19	Maternity & neonates*	£8	17 167	19	£431,977	16 801	18	£441,387
	First 11 PBC's				£19,928			£15,628
3	Disorders of Blood	£11		562	£19,928		716	£15,628
5	Mental Health Disorders	£204		10 245	£19,928		13 064	£15,628
6	Learning Disability	£31		1 545	£19,928		1 970	£15,628
8	Problems of Vision	£24		1 213	£19,928		1 547	£15,628
9	Problems of Hearing	£6		321	£19,928		409	£15,628
12	Dental problems	£23		1 179	£19,928		1 503	£15,628
14	Skin	£11		528	£19,928		673	£15,628
15	Musculo skeletal system	£15		759	£19,928		968	£15,628
20	Poisoning and AE	£4		220	£19,928		281	£15,628
21	Healthy Individuals	£18		920	£19,928		1 173	£15,628
22	Social Care Needs	£68		3 393	£19,928		4 326	£15,628
23	Other	£78		0	NA		0	NA
	All (23 PBCs)				£22,523			£17,663

Note that we have been unable to obtain a satisfactory outcome model for trauma & injuries and have assumed a zero outcome elasticity.
Note that, for expenditure in 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

C.2.1.4 Inferring excess deaths

We have been able to establish a measure of net YLL which takes account of deaths that would have occurred anyway below a normal LE for the PBC population (i.e., not all deaths observed in a PBC are excess) and that some deaths observed above this LE would not otherwise have occurred at that age (i.e., some of these deaths are excess). As explained in Section C.2.1.3, net YLL calculated in this way is equivalent to first establishing the number of excess deaths at each age, then calculating YLL for each excess death (based on the LE conditional on the age at which each excess death occurred) and then summing these YLL across all excess deaths (i.e., across all ages). In other words, the estimates of net YLL imply a number of excess deaths required to generate them in each PBC. Therefore, it is possible to solve for the total number of excess deaths based on the net YLL and the average YLL per observed death (the average of the sum of the YLLs for every observed death where the YLL for each observed death is the difference between age at death and LE conditional on age of death). The net YLL divided by the average YLL per death provides the number of excess deaths required, which on average will generate the estimated net YLL.

In the absence of information about the age distribution of excess deaths, calculations assume that the average YLL associated with observed and excess deaths are similar. Insofar as excess deaths are thought likely to generate more YLL than observed deaths the number of excess deaths will tend to be overestimated. This would tend to underestimate the cost per excess death averted. However, the cost per life year estimates remain unchanged and do not require such an assumption.

The implied excess deaths associated with net YLL based on the LE of the PBCs (see column 7 Table C.14) are reported in Table C.17. With the exception of PBC18&19, excess deaths are some proportion

of total observed deaths in each PBC. The proportion of excess deaths differs by PBC reflecting the distribution of deaths relative to the LE of the PBC. For example, in those PBCs where a large proportion of deaths occur below LE (see column 3 and 4) excess deaths tend to be greater proportion of total deaths (e.g., PBC2, 13 and 10). Where most deaths occur above LE excess deaths as a proportion of total deaths tend to be lower (e.g., PBC11, 17 and 1). Nevertheless, the impact of the age distribution of deaths and the age distribution of the at risk population (summarised as LE) on the calculation of excess deaths is not always obvious as both will affect the numerator (net YLL) as well the denominator (average YLL per death) in this calculation.

Table C.17: Excess deaths implied by net YLL.

PBC		Net YLL [1]	YLL per observed death [2]	Excess deaths [3]	Total deaths [4]	% excess deaths [5]
1	Infectious diseases	36,962	13.4	2,797	6 958	40%
2	Cancer	1,347,184	14.1	95,715	130 810	73%
4	Endocrine	51,225	13.7	3,769	6 764	56%
7	Neurological	93,917	13.7	6,909	15 353	45%
10	Circulatory	823,768	10.5	79,218	159 851	50%
11	Respiratory	68,030	9.2	7,386	65 445	11%
13	Gastro-intestinal	227,703	15.2	15,199	24 147	63%
17	Genito-urinary	18,127	8.3	2,172	10 625	20%
18+19	Maternity & neonates*	16,801	73.9	226	226	100%

Excess deaths are calculated for each gender by dividing net YLLs by the YLL per death (column [3] = column [1] / column [2])
* The number of excess deaths estimated in PBC18&19 was initially estimated to be 230, higher than the number of total deaths. This is due to the use of approximations (i.e. in the life expectancy, or in using the net YLL) thus, for consistency, we assumed this to be 100% of the total deaths.

Estimates of net YLL and changes in life years due to expenditure (see Table C.14 and C.15) have already accounted for the fact that not all deaths are excess and do not generate YLL. Nevertheless, solving for the number of implied excess deaths associated with these net YLL estimates allows a comparison of the cost per excess and observed PBC death avoided and an examination of the interpretation that can be placed of the life years expected to be gained from an excess or observed death averted.

Since only deaths observed in the PBC can be used to estimate the effects of expenditure (excess deaths are not directly observed since they rely on an unobserved counterfactual population and would occur outside the PBC), the outcome elasticities can be interpreted as the proportionate change in observed PBC mortality due to a proportionate change in PBC expenditure. Equally, however, they can also be interpreted as the proportionate effect on excess death due to a proportionate change in expenditure so can be applied to either total observed or total excess deaths. Observed PBC mortality that is sensitive to changes in expenditure can be regarded as 'avoidable' and it is only this mortality that contributes to the estimates of outcome elasticities (not all observed mortality is necessarily avoidable and sensitive to expenditure - such mortality will not contribute to the estimates). Not all observed mortality is excess when compared to the counterfactual population but this is unrelated to the question of how sensitive it is to expenditure, i.e., observed mortality will be just as sensitive to expenditure whether or not it is regarded as excess. Therefore, the estimated outcome elasticities can be applied to either observed PBC deaths or excess PBC deaths

The cost per excess death and the cost per PBC death averted are reported in Table C.18, and a detailed breakdown of changes in spend and excess or total deaths across PBCs is shown in Table C.19. The cost per PBC death averted is, of course; significantly lower than the cost per excess death as excess deaths are only a proportion of total deaths (see Table C.17). Also the cost per PBC death averted are substantially lower than those reported in Appendix B (see Tables B8.22 and B8.23), since these estimates do not restrict the effects of expenditure to PBC deaths under 75.

Table C.18. Summary of the cost per death averted threshold

	Cost per excess death averted, £ [1]	Cost per PBC death averted, £ [2]
big 4 PBC's	£91,129	£32,864
11 PBCs (with mortality)	£177,692	£64,774
All 23 PBCs (zero health effects for remaining 12 PBCs)	£653,748	£238,310
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£200,829	£73,208

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

Table C.19: Breakdown of the cost per death averted threshold.

			PBC deaths			Excess deaths		
PBC	PBC description	Change in spend, £m [1]	Total PBC deaths [2]	Change in PBC deaths [3]	Cost per PBC death averted, £ [4]	Excess deaths [5]	Change in excess deaths [6]	Cost per excess death averted, £ [7]
2	Cancer	£19	130 809	208.03	£92,147	95 715	152.22	£125,934
10	Circulatory problems	£33	159 851	1237.82	£26,878	79 218	613.43	£54,235
11	Respiratory problems	£22	65 446	1165.14	£19,142	7 386	131.49	£169,616
13	Gastro-intestinal problems	£17	24 148	165.42	£99,757	15 199	104.12	£158,488
	Big 4				£32,864	0		£91,129
1	Infectious diseases	£8	6 958	2.59	£3,222,218	2 797	1.04	£8,014,595
4	Endocrine problems	£18	6 765	54.28	£325,291	3 769	30.24	£583,830
7	Neurological problems	£17	15 353	10.59	£1,622,486	6 909	4.77	£3,605,579
17	Genito-urinary problems	£32	10 625	4.94	£6,425,694	2 172	1.01	£31,430,287
16	Trauma & injuries*	£10	NA	0	NA	NA	0	NA
18+19	Maternity & neonates*	£8	226	0.24	£32,813,038	226	0.24	£32,813,038
	First 11 PBC's				£64,774			£177,691
3	Disorders of Blood	£11		172.87	£64,774		63.01	£177,692
5	Mental Health Disorders	£204		3152.02	£64,774		1149.00	£177,692
6	Learning Disability	£31		475.30	£64,774		173.26	£177,692
8	Problems of Vision	£24		373.19	£64,774		136.04	£177,692
9	Problems of Hearing	£6		98.72	£64,774		35.99	£177,692
12	Dental problems	£23		362.72	£64,774		132.22	£177,692
14	Skin	£11		162.30	£64,774		59.16	£177,692
15	Musculo skeletal system	£15		233.59	£64,774		85.15	£177,692
20	Poisoning and AE	£4		67.80	£64,774		24.71	£177,692
21	Healthy Individuals	£18		283.09	£64,774		103.19	£177,692
22	Social Care Needs	£68		1043.74	£64,774		380.47	£177,692
23	Other	£78		0	£0		0	NA
	All (23 PBCs)				£73,208			£200,828

Recall from Appendix B that the measure of mortality that is available at PCT level and used to estimate the outcome elasticities is restricted to deaths under 75, as are the published estimates of YLL associated with them (see Section C.2.1.1). However, to restrict effects only to those under 75 would imply that there is no excess mortality above 75 or equivalently that there are no health effects of PBC expenditure above 75. Rather than assume no effects of NHS activity in older populations we apply the effects that can be observed to the whole PBC but account for deaths that would have otherwise occurred in our estimate of net YLL in Section C.2.1.3. Table C.20 illustrates the number deaths averted for a 1% change in budget implicit in the alternative calculations of the cost per death averted threshold.

Table C.20: Illustration of the number of deaths averted for a 1% change in budget

	Using deaths < 75 (Appendix B, Table B8.21)		Using excess deaths (Table C.18)		Using PBC deaths (Table C.18)	
	Cost per death averted (<75), £ [1]	Number of deaths averted (<75) for a 1% change in budget [2]	Cost per excess death averted, £ [3]	Number of excess deaths averted for a 1% change in budget [4]	Cost per PBC death averted, £ [5]	Number of PBC deaths averted for a 1% change in budget [6]
big 4 PBC's	£137,188	665	£91,129	1,001	£32,864	2,776
11 PBCs (with mortality)	£270,881	681	£177,692	1,039	£64,774	2,849
All 23 PBCs (zero health effects for remaining 12 PBCs)	£996,655	681	£653,748	1,039	£238,310	2,849
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£306,153	2,218	£200, 828	3,381	£73,208	5,191

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

In many respects, whether or not PBC deaths at older ages are as sensitive to changes in expenditure is not critical since any observed deaths that might be averted at older ages are less likely to generate life years gained because they are more likely to have occurred anyway in that year (i.e., are excess so generate zero life years gained anyway). Therefore, they will have very limited impact on cost per life year or subsequently on cost per QALY estimates (in Sections C.2.2 and C.2.3). For this reason, it is the cost per life year rather than cost per death averted, whether excess or observed, that is of primary interest. The cost per PBC or excess death averted (or life saved) should thus not be over interpreted since lives are never saved (death is only delayed). However, establishing the number of excess and PBC deaths averted which are associated with net YLL is useful because it enables an assessment of the number of life years gained associated with each death averted. These are reported for each PBC in Table C.21 and range from 74.3 years per excess death for PBC 18 & 19 Maternity & neonates to 8.3 per excess death for PBC17 Genito-urinary. On average, across all 11 PBCs each excess death averted is associated with 11.4 life years gained.

Table C.21: Implied YLL per death averted for each PBC

PBC	PBC description	Implied YLL per excess death averted [1]	Implied YLL per PBC death averted [2]
2	Cancer	14.07	10.30
10	Circulatory problems	10.40	5.15
11	Respiratory problems	9.21	1.04
13	Gastro-intestinal problems	14.98	9.43
	Big 4	11.28	4.07
1	Infectious diseases	13.21	5.31
4	Endocrine problems	13.59	7.57
7	Neurological problems	13.59	6.12
17	Genito-urinary problems	8.34	1.71
16	Trauma & injuries	NA	NA
18+19	Maternity & neonates	74.34	74.34
	First 11 PBC's	11.37	4.14

However, clinicians or the evaluative literature cannot distinguish whether an observed death is excess or not. What can be observed is whether groups of similar patients with and without access to a treatment survive and for how long. Therefore, it is the life years associated with each observed death that provides a context that can be interpreted based on experience and evidence of how effective those interventions that could be invested or disinvested tend to be. The average life years expected to be gained associated with each observed PBC deaths averted takes account of that fact that some deaths that are avoided in the PBC are not delayed for very long but quickly occur elsewhere and do not generate LY gained (i.e., they were not excess deaths). The portion of observed deaths that are regarded as excess depend on how time is discretised. The data available reports deaths in annual intervals so in this context 'quickly' means within one year. If deaths were reported in narrower time intervals then a greater proportion of observed

deaths would be regarded as excess and, in the limit, with continuous time all observed deaths would be excess. Of course, the average YLL associated with them would be smaller and is approximated by the net YLLs per observed death reported (the effects of approximation is likely to be small but unavoidable as it is due to deaths being reported in annual intervals).

However, establishing the number of excess and PBC deaths averted which are associated with net YLL is useful because it enables an assessment of the number of life years gained associated with each death averted. On average across all 11 PBCs each excess death averted is associated with 11.4 life years gained. These are reported for each PBC in Table C21 in Appendix C and range from 74.3 years per excess death for PBC 18 & 19 Maternity & neonates to 8.3 for PBC17 Genito-urinary. However, clinicians or the evaluative literature cannot distinguish whether an observed death is excess or not. What can be observed is whether groups of similar patients with and without access to a treatment survive and for how long. Therefore, it is the life years associated with each observed death that provides a context that can be interpreted based on experience and evidence of how effective those interventions that could be invested or disinvested tend to be. The average life years expected to be gained associated with each observed PBC deaths averted takes account of that fact that some deaths that are avoided in the PBC are not delayed for very long but quickly occur elsewhere and do not generate LY gained (i.e., they were not excess deaths). What portion of observed deaths are regarded as excess depend on how time is discretised.

The data available reports deaths in annual intervals so in this context 'quickly' means within one year. If deaths were reported in narrower time intervals then a greater proportion of observed deaths would be regarded as excess and in the limit with continuous time all observed deaths would be excess. Of course, the average YLL associated with them would be smaller and is approximated by the net YLLs reported in Table 4.5 per observed death (the effects of approximation is likely to be small but unavoidable as it is due to deaths being reported in annual intervals). These are also reported for each PBC in Table C21 in Appendix C and range from 74.3 years per observed death for PBC 18 & 19 Maternity & neonates[8] to 1.0 for PBC11 Respiratory problems, i.e., the YLL per PBC death are much lower for those PBCs where a small proportion of observed deaths are excess. On average across all 11 PBCs each PBC death averted is associated with 4.1 life years gained.

C.2.1.5 Summary of cost per life year estimates

The sequence of analysis set out above has enabled an examination of the impact of the limitations associated with the incomplete reporting mortality data at PCT level and incomplete coverage of published YLL estimates. We have also been able to consider effects above 75 while taking account of that fact that many deaths would have occurred anyway, despite the limited information available about the population at risk within a PBC. The GBD Study does provide some information about the age and gender distribution of the population at risk in a PBC so offers some improvement over the other assumptions that would otherwise be required (i.e., that the distribution of age and gender is the same as the general population or follows the distribution of observed deaths). For this reason the cost per life year threshold in column 4 of Table C.15 and repeated in lines 1 to 4 in Table C.22 are regarded as the central or best estimates given the evidence available and the credibility of alternative assumption that could be made. As explained in Section C.1, these are based on the conservative assumption that any health effects of changes in expenditure are restricted to one year, which, to some extent, may be offset by the more optimistic assumption any death averted returns the individual to the mortality risk faced by the general population, matched for age and gender.

[8] This is the same as life years associated with excess deaths since all observed deaths in this PBC are excess.

Table C.22: Summary of the cost per life year threshold with upper and lower bounds

	Best estimate	
Effect of expenditure on mortality:	*1 year*	
YLL per PBC death averted:	*~ 4.1 YLL*	
big 4 PBC's	£8,080	[1]
11 PBCs (with mortality)	£15,628	[2]
All 23 PBCs (zero health effects for remaining 12 PBCs)	£57,497	[3]
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£17,663	[4]
	Lower bound	
Effect of expenditure on mortality:	*Remainder of disease*	
YLL per PBC death averted:	*~ 4.1 YLL*	
big 4 PBC's	£3,846	[5]
11 PBCs (with mortality)	£6,106	[6]
All 23 PBCs (zero health effects for remaining 12 PBCs)	£22,463	[7]
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£6,901	[8]
	Upper bound	
Effect of expenditure on mortality:	*1 year*	
YLL per PBC death averted:	*2 YLL*	
big 4 PBC's	£16,432	[9]
11 PBCs (with mortality)	£32,387	[10]
All 23 PBCs (zero health effects for remaining 12 PBCs)	£119,155	[11]
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£36,604	[12]

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

It does not seem credible to imagine that NHS expenditure has no health effects in the 12 PBCs which do not have sufficient mortality reported at PCT level to estimate outcome elasticities - what is implied by the estimate reported in line 3. Therefore, it is the estimates reported in lines 2 and 4 that are of policy interest. The estimate of £15,628 per life year (line 2) is restricted to the effects of changes in expenditure in the 11PBCs where outcome elasticities can be estimated. The threshold of £17,663 per life year uses the estimated health effects of expenditure in these PBC as a surrogate for health effects in the others, i.e., assuming that the effects that can be observed will be similar to those that cannot. However, no health effects are assigned to PBC23 (General Medical Services) on the basis that any health effects of this expenditure would be recorded in the other PBCs.

It would be inappropriate to assign all the change in GMS expenditure to the estimate of cost per life year based only on the 11 PBCs with outcome elasticities because it would imply that GMS only contributes to these PBCs. Restricting attention to the 11 PBCs with outcome elasticities but allocating part of the change in GMS expenditure to them based on their proportional share of changes in overall expenditure would yield the same cost per life year as reported in line 4. [9]

The extreme upper and lower bounds for the cost per life year thresholds in Table C.22 are based on making the necessary assumptions about duration of health effects and how long a death might be averted optimistic (providing the lower bound for the threshold) or conservative (an upper bound for the threshold). The lower bound (lines 5 to 8) is based on assuming that health effects are not restricted to one year but apply to the whole of the remaining disease duration of the population at risk in PBCs during the expenditure year. Estimates of the average disease durations across the PBCs used in this calculation are depicted in Table C.23 (column 2).[10] These were obtained from the GBD Study (see Addendum 1 to this Appendix). Although this lower bound for the threshold combines optimistic assumptions, it is possible, indeed likely, that at least some expenditure may have effects on the health outcomes of future patients that are not currently part of the population at risk in a PBC, e.g., investments or disinvestment in prevention will have an impact on populations that are incident to PBCs in the future. Such effects are not captured in any of the estimates presented in this chapter so all are conservative with respect to this type of health effect of changes in expenditure.

[9] It should be noted that including changes in GMS expenditure but not assigning health effects to this PBC is likely to overestimate the threshold because any health effects associated with GMS will not be reflected in the estimated outcome elasticities of other PBCs unless the effects happen to be correlated with changes in expenditure in those PBCs.

[10] This information is also used in Section C.2.3.

Table C.23: Disease duration by PBC (GBD).

PBC	Duration of disease for an incident patient (years), GBD [1]	Remaining duration of disease for at risk population (years), GBD [2]
1	6.21	3.11
2	1.19	0.59
3	1.07	0.53
4	24.83	12.42
5	7.41	3.70
6	3.46	1.73
7	30.91	15.45
8	13.96	6.98
9	16.40	8.20
10	3.21	1.61
11	11.24	5.62
12	0.33	0.17
13	0.27	0.13
14	1.01	0.50
15	9.56	4.78
16	3.74	1.87
17	1.11	0.56
18	0.58	0.29
19	9.71	4.86
20	0.93	0.47
21	1.07	0.53
22	3.74	1.87
23	3.74	1.87

The upper bound (lines 9 to 12 in Table C.22) is based on the combination of assuming that health effects are restricted to one year for the population currently at risk and that any death averted is only averted for the minimum duration consistent with the mortality data. The econometrics work used the average of 3 years of mortality (2006 to 2008), so the estimated outcome elasticities are based on differences in mortality that remain after averaging over three years. Therefore, the estimated effects are based on differences in observed PBC deaths that must have been sustained, on average, for more than a minimum of 2 years. This is because whilst variation in mortality the first year of data will only contribute to estimates if differences are sustained for a minimum of 3 years, variation in mortality in the second year will only contribute if it is sustained for a minimum of 2 years, and in the third year only if sustained for 1 year. If differences in mortality are similar each year (the three years contribute equally to the estimates) then estimated effects must have been sustained, on average, for a minimum of 2 years.[11] These estimates can be interpreted as an upper bound given the data available and therefore the analysis that has been feasible.

C.2.2 Adjusting life years for quality of life

The central or best estimates of the cost per life year threshold, which were presented in Table C.22 (lines 2 and 4) take no account of the health related quality of life in which years of life, expected to be gained or lost through changes in expenditure, are likely to be lived. Even if attention is restricted to the direct health consequences of changes in mortality, estimates of the cost per life year will tend to overestimate the effects of changes in expenditure (underestimate the threshold) compared to a more complete measure of health that accounts for the quality in which the years of life are expected to be lived. In this Section we examine the ways in which the life years reported in Section C.2.1 can be adjusted for quality, taking account of information that is available about: i) how quality of life differs by age and gender (see Section C.2.1), and ii) how the quality of life years associated with mortality changes might be affected by the types of diseases that make up each PBC (see Section C.2.2). Throughout we continue to take

[11]Since some of the variation in mortality in 1st year that is not sustained to the 3rd year will nevertheless be sustained for 1 or 2 years, 2 life years per death averted represents somewhat less than the minimum, consistent with restricting life years gained to the observed mortality data.

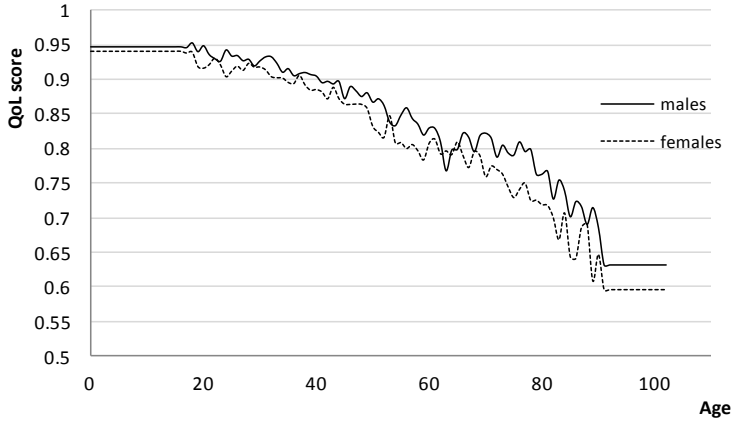
account for counterfactual deaths in the way described in Section C.2.1.3 by making the adjustment for quality to the life years associated with every observed death before calculating a quality adjusted net YLL. The implications for a cost per quality adjusted life year (QALY) threshold that only accounts for the health effects of mortality changes are presented in Section C.2.3.2. In Section C.2.3 we explore the ways in which the likely direct effects of expenditure on quality of life (other than through mortality) might also be taken into account.

C.2.2.1 Quality of life based on the general population

The most commonly used metric of health related quality of life in the UK is EQ5D,[1] which is specified in the NICE reference case for methods of technology appraisal.[2] This metric has 5 dimensions of quality each with three possible levels. Each of these 243 possible health states is valued relative to a score of one, which represents full or best imaginable health (the best score across all 5 dimensions), and a score of zero, which represents death, based on a representative sample of the UK population.[3] Therefore, insofar as the years of life expected gained or lost through changes in expenditure would be lived in this state of full health, the cost per life year thresholds reported in Table C.22 would also be the cost per QALY thresholds, albeit ones that only account for the health effects of mortality changes.

However, unsurprisingly, there is good evidence that, on average, the general population is not in this state of full health. Therefore, the quality of life score associated with the health states experienced by the general population are less than 1, and are expected to decline with age and to differ by gender. These quality of life 'norms' for the general population by age and gender are illustrated in Figure C.4 based on an analysis of data from the Health Survey for England (HSE, see Addendum 1 to this Appendix for a description on HSE data and the analysis of quality of life norms illustrated in Figure C.4).

Figure C.4: Quality of life for the general population by age and gender



These quality of life norms can be applied to the YLL associated with all observed deaths in each PBC, taking account of gender and age at death. The results are reported in column 4 to 6 of Table C.24.

Table C.24: Net YLL adjusted for quality of life 'norms'

PBC		Unadjusted life years			Quality adjusted life years		
		YLL [1]	YLG [2]	Net YLL [3]	YLL [4]	YLG [5]	Net YLL [6]
1	Infectious diseases	58,686	21,724	36,962	47,481	14,618	32,864
2	Cancer	1,473,733	126,549	1,347,184	1,143,445	84,036	1,059,409
4	Endocrine	66,283	15,058	51,225	52,856	9,973	42,883
7	Neurological	135,686	41,770	93,917	109,349	28,262	81,087
10	Circulatory	1,102,020	278,251	823,768	848,046	183,330	664,717
11	Respiratory	298,343	230,313	68,030	231,578	154,743	76,835
13	Gastro-intestinal	273,117	45,414	227,703	216,256	30,277	185,979
17	Genito-urinary	47,229	29,101	18,127	35,929	18,947	16,982
18+19	Maternity & neonates	16,801	0	16,801	14,568	0	14,568

There are two effects of adjusting life years for quality: i) since quality of life norms are always less than 1 the adjusted YLL and YLG are always lower than the unadjusted values in columns 1 and 2 (previously reported in Table C.5); and ii) deaths above LE are necessarily at older ages with poorer quality of life norms than those below, so the difference between adjusted and unadjusted values is greater for YLG than YLL (Table C.25 illustrates these effects by showing the implied QoL scores applied to YLL and YLG). The overall effect of quality adjustment on net YLL is the balance of these two effects, and tends to reduce the net YLL (compare column 6 and 3 in Table C.24). The only exception is PBC11 (Respiratory) which has a large proportion of deaths occurring above the life expectancy of the PBC population (see Table C.14).

Table C.25: Implied quality of life score in the net YLL adjustment for quality of life 'norms'

PBC		QoL score for YLL [1]	QoL score for YLG [2]
1	Infectious diseases	0.81	0.67
2	Cancer	0.78	0.66
4	Endocrine	0.80	0.66
7	Neurological	0.81	0.68
10	Circulatory	0.77	0.66
11	Respiratory	0.78	0.67
13	Gastro-intestinal	0.79	0.67
17	Genito-urinary	0.76	0.65
18+19	Maternity & neonates	0.87	NA

The quality adjusted net YLL figure in Column 6 suggest that the health effects of mortality are lower than when relying only on unadjusted life years in Section C.2.1.3. Therefore, the health effects of changes in expenditure on this more complete measure of health will also be lower. The implications of these adjustments to a cost per QALY threshold that only accounts for the direct health effects of mortality are summarised in Table C.26, and detailed in Table C.27. As expected, the cost per QALY threshold based on adjusting the life years gained or lost (column 2, Table C.26) is higher than a threshold based on unadjusted life years (column 1 in Tables C.26, these results were previously reported in Tables C.15 and C.22).

Table C.26: Summary of cost per QALY threshold based on population norms and mortality effects

	Cost per life year threshold [1]	Cost per QALY threshold Population norms [2]
big 4 PBC's	£8,080	£9,631
11 PBCs (with mortality)	£15,628	£18,622
All 23 PBCs (zero health effects for remaining 12 PBCs)	£57,497	£68,513
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£17,663	£21,047

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal except GMS.

Table C.27: A breakdown of the cost per QALY threshold based on population norms

PBC	PBC description	Change in spend, £m [1]	Change in QALY [3]	Cost per QALY gained, £ [4]
2	Cancer	£19	1685	£11,378
10	Circulatory problems	£33	5147	£6,464
11	Respiratory problems	£22	1368	£16,304
13	Gastro-intestinal problems	£17	1274	£12,952
	Big 4			£9,631
1	Infectious diseases	£8	12	£682,211
4	Endocrine problems	£18	344	£51,309
7	Neurological problems	£17	56	£307,201
17	Genito-urinary problems	£32	8	£4,020,316
16	Trauma & injuries*	£10	0	NA
18+19	Maternity & neonates*	£8	16	£509,044
	First 11 PBC's			£18,622
3	Disorders of Blood	£11	601	£18,622
5	Mental Health Disorders	£204	10964	£18,622
6	Learning Disability	£31	1653	£18,622
8	Problems of Vision	£24	1298	£18,622
9	Problems of Hearing	£6	343	£18,622
12	Dental problems	£23	1262	£18,622
14	Skin	£11	565	£18,622
15	Musculo skeletal system	£15	812	£18,622
20	Poisoning and adverse effects	£4	236	£18,622
21	Healthy Individuals	£18	985	£18,622
22	Social Care Needs	£68	3630	£18,622
23	Other	£78	0	NA
	All (23 PBCs)			£21,047

Table C.28 depicts the judgements over life years, quality of life weights and total QALYs implicit in calculations of the threshold cost per QALY in Table C.26. Specifically, columns 1 and 2 of Table C.28 report the number of life years associated with each death averted for each PBC; as expected, the values are equal to those in Table C.21 as estimates rely on the net YLLs evaluated in Section C.2.1.3. In columns 3 and 4, the number of QALYs gained associated with each death averted are presented. These ranged from 64.46 QALYs gained per PBC death averted for PBCs18&19 (Maternity and Neonates) to 1.17 QALYs per PBC death averted for PBC11 (Respiratory) – column 4. In general, these values are expected to be smaller than the unadjusted YLL per PBC death averted in column 2. The exception is PBC 11 (respiratory) – in this PBC, the number of YLL and YLG are more similar than in other PBCs (respectively, columns 1 and 2 of Table C.24), and given that YLGs are weighted more heavily (with lower QoL scores) than YLL, the netting of adjusted estimates returns a higher number than the netting of unadjusted estimates. On average, across all 11 PBCs each PBC death averted is associated with 3.5 QALYs gained.

Table C.28: Implied YLL per excess death averted and implied QoL score per YLL gained, for each PBC

PBC	PBC description	Implied YLL per excess death averted [1]	Implied YLL per PBC death averted **[2]**	Implied QALYs gained per excess death averted [3]	Implied QALYs gained per PBC death averted [4]
2	Cancer	14.07	10.30	11.07	8.10
10	Circulatory	10.40	5.15	8.39	4.16
11	Respiratory	9.21	1.04	10.40	1.17
13	Gastro-intestinal	14.98	9.43	12.24	7.70
	Big 4	11.28	4.07	9.46	3.41
1	Infectious diseases	13.21	5.31	11.75	4.72
4	Endocrine	13.59	7.57	11.38	6.34
7	Neurological	13.59	6.12	11.74	5.28
17	Genito-urinary	8.34	1.71	7.82	1.60
16	Trauma & injuries*	NA	NA	NA	NA
18+19	Maternity & neonates*	74.34	74.34	64.46	64.46
	First 11 PBC's	11.37	4.14	9.54	3.48

C.2.2.2 Adjusting age related quality of life for disease decrements

Adjusting life years for age and gender related quality of life norms assumes that any life year gained through a change in expenditure would be lived in a similar quality of life to the general population. It is possible however, that patients benefiting from reduced mortality may, nevertheless, continue to be affected by the type of diseases that make up each PBC and experience the quality of life associated with the original disease.

The Health Outcome Data Repository (HODaR)[4] provides over 30,000 observations of EQ-5D measures of quality of life by ICD code and the age and gender of the patients in the sample (see Addendum 1 to this Appendix). Although this is a rich UK data set, there were a limited number of observations for some of the less common ICD codes. For this reason HODaR was supplemented with information from the Medical Expenditure Panel Survey (MEPS)[5] which also provides EQ-5D by ICD and reports the average age of respondents (see Addendum 1 to this Appendix). These data provided a means of estimating the quality of life associated with each ICD code at the average age of respondents in the pooled sample (ICD estimates of the quality of life score and age were pooled across datasets by considering the number of patients from each dataset contributing to estimates, i.e. a weighted average). The quality of life associated with each PBC was then expressed as the average of the quality of life associated with its component ICDs. The average quality of life scores across ICDs which contribute to each PBC and the average age and gender of respondents were used to calculate a PBC disease related decrement (disutility) based on quality of life norms from the general population – it is important to note that by expressing the quality of life effects of different diseases as age related decrements we do not require the HODaR and MEPS samples to necessarily be representative of the age distribution of the population at risk in the PBCs.

Table C.29 summarises the data from HoDAR and MEPS and the quality of life decrements used further in calculations of the threshold, namely: the number of patients for which quality of life scores were available (column 1), the average age of these patients by gender (columns 2 and 3), the average quality of life scores across PBCs (column 4), the quality of life scores for the population norms by gender (columns 5 and 6), and the calculated disease related decrements (columns 7 and 8).

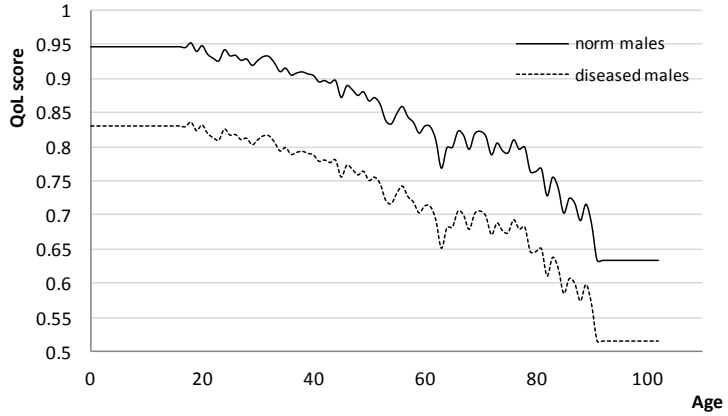
Table C.29: QoL scores per PBC from different sources

PBC	HoDAR/MEPS N	HoDAR/MEPS average age male	HoDAR/MEPS average age female	HoDAR/MEPS QoL score for diseased	Population norms male	Population norms female	Disease related decrement compared to population norms male	Disease related decrement compared to population norms female
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
1	263	54.0	47.1	0.667	0.859	0.830	0.192	0.163
2	13 324	64.3	59.8	0.692	0.809	0.830	0.117	0.138
3	2 464	58.6	58.1	0.656	0.859	0.830	0.203	0.174
4	7 128	57.3	56.5	0.701	0.859	0.830	0.157	0.128
5	12 733	47.8	47.9	0.557	0.859	0.830	0.301	0.272
6	301	25.8	25.3	0.671	0.937	0.924	0.266	0.253
7	10 296	55.8	53.8	0.546	0.859	0.830	0.312	0.283
8	11 536	63.8	64.5	0.719	0.809	0.796	0.089	0.077
9	1 023	61.7	59.8	0.778	0.809	0.830	0.031	0.051
10	33 854	64.4	64.1	0.629	0.809	0.796	0.179	0.167
11	19 646	48.4	47.2	0.634	0.859	0.830	0.224	0.195
12	1 811	40.9	40.0	0.781	0.910	0.894	0.129	0.113
13	23 138	57.3	55.5	0.653	0.859	0.830	0.206	0.177
14	5 659	54.8	54.0	0.695	0.859	0.830	0.164	0.134
15	34 590	56.4	56.6	0.578	0.859	0.830	0.280	0.251
16	2 652	46.0	58.3	0.652	0.859	0.830	0.207	0.178
17	13 651	57.5	53.0	0.711	0.859	0.830	0.147	0.118
18+19	1 566	37.8	31.7	0.848	0.910	0.894	0.063	0.047
20	1 569	59.1	52.3	0.584	0.859	0.830	0.275	0.246
21	7 488	60.6	60.3	0.661	0.809	0.796	0.147	0.135
22	25	78.4	81.4	0.156	0.798	0.636	0.642	0.480
23	1 002	62.6	60.8	0.639	0.809	0.796	0.170	0.158

no gender details were available from MEPS so assumed 50:50 split of frequency
only primary diagnosis is used from HoDAR data
a lower bound of 0 is assumed for disutility for each PBC

Figure C.5 illustrates the use of the decrement to quality of life norms for PBC1 (Infectious disease) across a range of ages. For PBC1, the quality of life score was evaluated across the component ICD codes was evaluated to be 0.667 in HoDAR and MEPS, at an average age of 54 for male respondents. Since the quality of life norms for males age 54 is 0.859 this suggests a decrement associated with membership of PBC1 of 0.192, which can then be applied to quality of life norms by age as illustrated in Figure C.5.

Figure C.5: Quality of life for males in PBC1 (Infectious disease) and the general population by age



In principle, it would be possible to estimate disease related disutility by age rather than assume a fixed additive decrement. HODaR does provide age for each reported quality of life score but MEPs only provides average age of respondents in published summaries. However, even with access to 'raw' scores and the age and gender of each, it is very unlikely that there would be sufficient data to estimate age related decrements in each of the component ICDs. It would, however, be possible to assume a proportionate rather than fixed decrement by age. However, the average age of respondents in the pooled HODaR and MEPs sample (columns 2 and 3 of Table C.29) tends to be older than the age distribution of the PBC populations (columns 3 and 4 of Table C.13). Given that older individuals are expected to have a lower quality of life (norm), relative decrements can overestimate the decrements observed in younger patients. By applying overestimated decrements, the quality adjusted net YLL would be underestimated and the cost per QALY threshold increased compared to the fixed decrement applied here.

Quality of life norms adjusted for disease related decrements can be applied to the YLL associated with observed deaths in each PBC, taking account of gender and age at death in the same way as Section C.3.1. To do so, the 'PBC decrements' calculated from HoDAR and MEPS were applied to each observed death and the age at which each life year was gained or lost (from ONS). The results are reported in columns 4 to 6 of Table C.30. The overall effect of quality adjustment that also applies a disease related decrement is to reduce the net YLL to a greater extent than adjustment with population norms alone (compare column 6 in Table C.30 to column 6 in Table C.24).

Table C.30: Net YLL adjusted for disease and age related quality of life

PBC		Unadjusted life years			Quality adjusted life years		
		YLL	YLG	YLL	YLG	YLL	YLG
		[1]	[2]	[3]	[4]	[5]	[6]
1	Infectious diseases	58,686	21,724	36,962	37,055	10,793	26,262
2	Cancer	1,473,733	126,549	1,347,184	955,690	67,930	887,760
4	Endocrine	66,283	15,058	51,225	43,394	7,844	35,550
7	Neurological	135,686	41,770	93,917	68,893	15,842	53,050
10	Circulatory	1,102,020	278,251	823,768	656,145	135,241	520,905
11	Respiratory	298,343	230,313	68,030	169,269	106,505	62,764
13	Gastro-intestinal	273,117	45,414	227,703	163,593	21,677	141,916
17	Genito-urinary	47,229	29,101	18,127	29,749	15,152	14,598
18+19	Maternity & neonates	16,801	0	16,801	13,662	0	13,662

The implied quality of life weights (considering the disease related decrements) for YLL and YLG are shown in Table C.31. Note that, as expected, the weights assume a lower value than in Table C.25.

Table C.31: Implied QoL weights in the net YLL adjusted for disease and age related quality of life

PBC		QoL weights for YLL [1]	QoL weights for YLG [2]
1	Infectious diseases	0.63	0.50
2	Cancer	0.65	0.54
4	Endocrine	0.65	0.52
7	Neurological	0.51	0.38
10	Circulatory	0.60	0.49
11	Respiratory	0.57	0.46
13	Gastro-intestinal	0.60	0.48
17	Genito-urinary	0.63	0.52
18+19	Maternity & neonates	0.81	NA

Combining quality of life adjustments for both population norms and disease related decrements assumes that any life years gained due to a reduction in mortality will be lived in the diseased state until life expectancy, i.e., that all diseases are not just chronic but disease duration is lifelong. Inevitably this assumption means that the health effects of changes in mortality will be reduced. Consequently the cost per QALY threshold reported in Table C.32 (column 2) will be higher than adjusting life years gained for population norms in Table C.26 (column 2). A detailed breakdown of the cost per QALY threshold based on disease related disability and mortality effects is shown is Table C.33.

Table C.32: Summary of cost per QALY threshold based on disease related disutility

	Cost per life year threshold [1]	Cost per QALY threshold Disease related disutility [2]
big 4 PBC's	£8,080	£12,109
11 PBCs (with mortality)	£15,628	£23,395
All 23 PBCs (zero health effects for remaining 12 PBCs)	£57,497	£86,072
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£17,663	£26,441

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal except GMS.

Table C.33: Breakdown of the cost per QALY threshold based on disease related disutility

PBC	PBC description	Change in spend, £m [1]	Change in QALY [4]	Cost per QALY gained, £ [5]
2	Cancer	£19	1412	£13,578
10	Circulatory problems	£33	4034	£8,248
11	Respiratory problems	£22	1117	£19,960
13	Gastro-intestinal problems	£17	972	£16,974
	Big 4		0	£12,109
1	Infectious diseases	£8	10	£853,712
4	Endocrine problems	£18	285	£61,892
7	Neurological problems	£17	37	£469,558
17	Genito-urinary problems	£32	7	£4,676,874
16	Trauma & injuries*	£10	0	NA
18+19	Maternity & neonates*	£8	15	£542,801
	First 11 PBC's		0	£23,395
3	Disorders of Blood	£11	479	£23,395
5	Mental Health Disorders	£204	8727	£23,395
6	Learning Disability	£31	1316	£23,395
8	Problems of Vision	£24	1033	£23,395
9	Problems of Hearing	£6	273	£23,395
12	Dental problems	£23	1004	£23,395
14	Skin	£11	449	£23,395
15	Musculo skeletal system	£15	647	£23,395
20	Poisoning and adverse effects	£4	188	£23,395
21	Healthy Individuals	£18	784	£23,395
22	Social Care Needs	£68	2890	£23,395
23	Other	£78	0	NA
	All (23 PBCs)			£26,441

The number of life years gained associated with each death averted (columns 1 and 2 in Table C.34) is, again, consistent with previous estimates (Tables C.28 and C.21). The average number of QALYs gained across all 11 PBCs is 2.8 QALY per death averted (column 4 in Table C.34). As expected this value is lower than in the previous section (column 4 in Table C.28).

Table C.34: Implied YLL per death averted and implied QoL score per YLL gained, for each PBC

PBC	PBC description	Implied YLL per excess death averted [1]	Implied YLL per PBC death averted [2]	Implied QALYs gained per excess death averted [3]	Implied QALYs gained per PBC death averted [4]
2	Cancer	14.07	10.30	9.28	6.79
10	Circulatory	10.40	5.15	6.58	3.26
11	Respiratory	9.21	1.04	8.50	0.96
13	Gastro-intestinal	14.98	9.43	9.34	5.88
	Big 4	11.28	4.07	7.53	2.71
1	Infectious diseases	13.21	5.31	9.39	3.77
4	Endocrine	13.59	7.57	9.43	5.26
7	Neurological	13.59	6.12	7.68	3.46
17	Genito-urinary	8.34	1.71	6.72	1.37
16	Trauma & injuries*	NA	NA	NA	NA
18+19	Maternity & neonates*	74.34	74.34	60.45	60.45
	First 11 PBC's	11.37	4.14	7.60	2.77

C.2.2.3 Summary of the cost per QALY threshold based only on mortality effects

The analysis to this point is summarised in Table C.35. The three estimates of a cost per QALY threshold are based on assuming that each life year gained is either: lived in full health (see column 1, equal to the cost per life year estimates in Table C.22), lived in a quality of life that reflects age and gender norms of the general population (column 2); or lived in a quality of life that reflects the original disease state (column 3).

Table C.35: Summary of QALY threshold estimates based only on mortality effects

	[1] *(QoL score =1)*	[2] *(population norms)*	[3] *(Disease related disutility)*	
		Best estimate		
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
*YLL per death averted**:*	*~ 4.1 YLL*	*~ 4.1 YLL*	*~ 4.1 YLL*	
*QALYs per death averted**:*	*~4.1 QALYs*	*~3.5 QALYs*	*~2.8 QALYs*	
big 4 PBC's	£8,080	£9,631	£12,109	[1]
11 PBCs (with mortality)	£15,628	£18,622	£23,395	[2]
All 23 PBCs*	£17,663	£21,047	£26,441	[3]
		Lower bound		
Effect of expenditure on mortality:	*Remainder of disease*	*Remainder of disease*	*Remainder of disease*	
*YLL per death averted**:*	*~ 4.1 YLL*	*~ 4.1 YLL*	*~ 4.1 YLL*	
*QALYs per death averted**:*	*~4.1 QALYs*	*~3.5 QALYs*	*~2.8 QALYs*	
big 4 PBC's	£3,846	£4,252	£5,319	[4]
11 PBCs (with mortality)	£6,106	£6,852	£8,568	[5]
All 23 PBCs*	£6,901	£7,744	£9,683	[6]
		Upper bound		
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
*YLL per death averted**:*	*2 YLL*	*2 YLL*	*2 YLL*	
*QALYs per death averted**:*	*2 QALY*	*~1.9 QALY*	*~1.5 QALY*	
big 4 PBC's	£16,432	£17,456	£21,747	[7]
11 PBCs (with mortality)	£32,387	£34,492	£42,967	[8]
All 23 PBCs*	£36,604	£38,983	£48,561	[9]

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal except GMS.
** see Tables C.20, C.27 and C.33

The weights reflecting the quality in which each of the years of life saved is lived implied in each of these three estimates is shown in Table C.36.

Table C.36: Implied QoL weight per YLL gained

PBC	PBC description	Full health [1]	Population norms [2]	Disease related disutility [3]
2	Cancer	1	0.79	0.66
10	Circulatory	1	0.81	0.63
11	Respiratory	1	1.13	0.92
13	Gastro-intestinal	1	0.82	0.62
	Big 4	1	0.84	0.67
1	Infectious diseases	1	0.89	0.71
4	Endocrine	1	0.84	0.69
7	Neurological	1	0.86	0.56
17	Genito-urinary	1	0.94	0.81
16	Trauma & injuries	NA	NA	NA
18+19	Maternity & neonates	1	0.87	0.81
	First 11 PBC's	1	0.84	0.67

Assuming that life years gained are lived in full health is not credible and should be regarded as an underestimate of the threshold given what is known about quality of life norms for the general population (see Figure C.4). Equally, assuming that all life years gained are lived in the quality of life of the original disease state does not seem credible either and is likely to overestimate the threshold since it assumes that all disease is not only chronic but lifelong and all life years would be lived in the diseased state until death. The information that is available about disease duration suggests that many types of disease that comprise the PBCs are not chronic and certainty not lifelong (see Table C.23). Although adjusting life years gained for the quality of life of the general population taking account of age and gender (in column 2, Table C35) is likely to underestimate the cost per QALY threshold based only on mortality effects, it probably represents the 'best' of the three alternative estimates available at this stage of the analysis . The lower and upper bounds are based on combining optimistic and pessimistic assumptions about the duration of health effects and how long a death might be averted as described in Section C.2.1.5.

However, it should be noted that these cost per QALY thresholds only account for the direct health effects of changes in mortality due to changes in expenditure. Insofar as much, or at least some part, of NHS activity and expenditure is intended to improve quality of life, not just mortality, then these estimates will underestimate total health effects and overestimate a cost per QALY threshold based on a more complete measure of possible health effects. In Section C.2.3 we explore the ways in which the likely effects of expenditure on quality of life (other than through mortality) might also be taken into account.

C.2.3. Including quality of life effects during disease

The cost per QALY thresholds presented in Section C.2.2 only account for the health (QALY) effects of changes in mortality due to changes in expenditure. It does not seem credible to suppose that all NHS activity and expenditure only influences mortality with no effect on the quality of life while alive and experiencing a disease. Insofar as changes in NHS expenditure will also affect quality of life as well as mortality then total health effects will be underestimated and the thresholds presented in Table C.35 will overestimate the cost per QALY threshold. In this section we explore ways to also take account of those effects on health not directly associated with mortality and life year affects (i.e., the 'pure' quality of life effects) to estimate an overall cost per QALY threshold.

The routine reporting of quality of life outcomes are increasingly available at PCT level (see Addendum 1 for a description of these data). In principle, the variation in such measures of outcome across PCTs could be used to estimate outcome elasticities for quality of life rather than mortality effects using similar econometric methods to those described in Appendix B (see Section B.8.8 for the results of an exploratory econometric analysis of these data). However, the currently limited coverage of routine reporting of these outcomes means that it is not feasible to estimate quality of life effects across all the PBCs using these data. Here we explore how estimates of effects of expenditure that can be observed (i.e., on mortality) can be used to infer the likely effects on what cannot be directly observed (quality of life), rather than making extreme assumptions that are not credible (e.g., assuming that changes in expenditure will have no effects on quality of life outcomes).

In Section C.2.3.1 we use three alternative estimates of the ratio of QALYs to life years lost due to different types of disease as a means of inferring the change in QALYs that is likely to be associated with the estimated change in YLL, i.e., applying the total QALYs lost associated with each YLL with disease. This is consistent with regarding the estimates of the mortality and life year effects as a surrogate for a more complete measure of the health effects of a change in expenditure.

However, these ratios of QALYs lost to life years lost due to disease in those PBC where outcome elasticities could not be estimated cannot inform estimates of the threshold (there are no estimated life year effects with which to apply the ratios). Nonetheless, the sources of information on which ratios are based also provide much of the information required to calculate the QALY burden of disease in these areas, which can be used to inform estimates of the threshold. Therefore, in section C.2.3.2 we use an estimate of the QALY burden of disease, infer a proportionate effect on burden from the observed effects on life years, and then apply this proportionate effect to the measures of QALY burden for all the other PBCs. In this way we can use all the information available about the mortality and quality of life effects of the different types of disease that make up each PBC, including those where mortality based outcome elasticities are not available.

C.2.3.1 Using ratios of QALYs to YLL

The ratio of the total QALYs to years of life lost (YLL) due to a disease indicates the number of QALYs associated with each YLL. Therefore, any change in YLL is expected to generate a number of QALYs indicated by the ratio - in this way, the estimated effects on mortality and life years are interpreted as a surrogate for a more complete measure of total health effects, which is reasonable. For example, a disease with a ratio greater than 1 suggests that each YLL across the at risk population is associated with

more than one QALY, i.e., there are significantly greater quality of life effects while experiencing the disease. Therefore, a change in expenditure that leads to 1 life year gained in this type of disease may generate a greater QALY effect than the same life year effects in a disease where this ratio is less than 1, i.e., where most of the effect of disease is on mortality rather than quality of life. Therefore, using these ratios provides a means of accounting for the likely effect on quality of life other than through effects on mortality.

To understand the differences between the three ratios presented below it is useful to regard the total QALY lost to YLL ratio (R) for a particular disease as the sum of two ratios: i) the QALYs lost due to premature death to YLL ratio (R_{death}) and ii) the QALYs lost during disease (while alive) to YLL ratio (R_{alive}), as depicted in Equation 4.

$$R = \frac{QALY\ lost}{YLL} = \underbrace{\frac{QALY\ lost_{\ premature\ death}}{YLL}}_{R_{death}} + \underbrace{\frac{QALY\ lost_{\ while\ alive}}{YLL}}_{R_{alive}}, \quad (4)$$

Insofar as YLL would not have been lived in full health, the quality of life effects captured in R_{death} are estimated to be lower than 1. Note that the analyses in Section C.2.2 already imply a R_{death} ratio at PBC level. The second component of the ratio, R_{alive}, represents QALYs lost during disease for the at risk population as a proportion of the YLL observed in the same population – in diseases for which quality of life during disease is compromised but life expectancy is not changed significantly R_{alive} may thus assume high values. The ratios do not represent the balance of QALY gains due to mortality and morbidity in a single patient, but rather in the population. Where R_{death} if lower than 1, only when the pure QALY effects offset the less than full quality of life of the YLL is the ratio greater than one. Therefore, ratios less than one are possible even when disease has measurable quality of life effects for those experiencing it.

DALY to YLL ratios

The WHO GBD study provides UK specific estimates of the years of life lived with disability and the years of life lost due to different types of disease. Diseases in GBD are classified using U-codes that can then be mapped to ICD-10, as illustrated in Table C.37 using a few examples (Addendum 1 provides more details on the mapping procedure). GBD uses Disability Adjusted Life Years (DALYs) as a measure of the burden of disease. This DALY measure has two components: i) the years of life lived with disability (YLD), which evaluates the number of years lived with disability over the durations of disease, and incorporates weights (between zero and one) to reflect the scale of disability experienced in each year; and ii) the years of life lost (YLL).

Table C.37: Illustration of the mapping between U-code and ICD

U-code	ICDs
U037 (Other infectious diseases)	A02,A05,A20-A28,A31,A32,A38,A40-A49,A65-A70,A74-A79, A81,A82, A83.1-A83.9, A84-A89,A92-A99, B00-B04,B06-B15,B25-B49,B58-B60, B64, B66-B72, B74.3-B74.9,B75,B82-B89,B92-B99, G04
U016 (Tetanus)	A33-A35
U061 (Mouth and oropharynx cancers)	C00-C14
U057 (Iron-deficiency anaemia)	D50, D64.9

The total DALY associated with a disease is simply YLL+YLD. Therefore, the DALY to YLL ratio is (YLL+YLD)/YLL or equivalently YLL/YLL + YLD/YLL. Since the first term (YLL/YLL = R_{death}) must equal one and the second (R_{alive} = YLD/YLL) must be ≥ 0, a ratio based on DALYs must necessarily be bounded by one.

$$R_{DALY} = \frac{DALY}{YLL} = \frac{(YLL+YLD)}{YLL} = \underbrace{1}_{R_{death}} + \underbrace{\frac{YLD}{YLL}}_{R_{alive}} \quad (5)$$

This is illustrated in Table C.38a for the four different diseases (classified by U-codes) introduced in Table C.36 which reflect diseases where mortality is the major component (e.g., U016) and where the impact of disease on the quality of life while alive is the major component (e.g., U141).

Table C.38a: Examples of DALY to YLL ratios

Ucode	DALY ratios	(R_{death} + R_{alive})
U037 (Other infectious diseases)	1.23	(1+0.23)
U016 (Tetanus)	1.00	(1+0)*
U061 (Mouth and oropharynx cancers)	1.05	(1+0.05)
U141 (Spina bifida)	2.34	(1+1.34)**

* Given the short disease duration, it is only mortality effects that contribute to the ratio
** Quality of life effects during disease contribute significantly to estimates of the ratio

Note that the estimates of GBD YLL used here are derived using UK data on mortality (relating to the year 2004) by age and gender groups – we assume these data to be from ONS and thus consistent with the data used in this work. However, the calculation of YLLs in GBD differs from both the approach adopted by NHS IC and the approach adopted here of using net YLL. For each death observed in the data, GBD evaluates YLL by considering the life expectancy at the age at which the death occurred (and gender).[6] This is expected to overestimate net YLL (which accounts for counterfactual deaths, as detailed in Section C.2.2.3). This will make no difference to the first term in the QALY ratio (R_{death}) since an overestimate of YLL affects both denominator and numerator of the ratio. However, the second term (R_{alive}) is likely to be underestimated. Therefore the ratios will tend to underestimate the QALY effects of expenditure and overestimate the cost per QALY threshold. This will be adjusted for in Section C.2.3.2, where our preferred analysis based on burden of disease is presented.

Adjusting DALYs for quality of life norms

The use of DALY ratios bounded below by one essentially assumes that YLL would have otherwise been lived in a state of full health. As was discussed in section C.2.3.1 this is not credible given information available about the quality of life in the general population (see Figure C.4). It would lead to overestimating the QALYs associated with mortality and life year effects and underestimating the cost per QALY threshold. Therefore, it is important to adjust these DALY ratios for the quality of life norms by age and gender in the same way as described in Section C.2.3.1. Equation 6 shows how the adjusted ratio is formulated when YLLs are adjusted by the quality of life in the general population, u_n. This is a simplified representation of the adjustment as despite gender and age having been considered in calculations these are not shown in the notation below.

$$R_{DALY\ adj} = \frac{u_n \cdot YLL}{YLL} + \frac{YLD}{YLL} \mathrel{+}= u_n + \underbrace{\frac{YLD}{YLL}}_{R_{alive}} \qquad (6)$$

$$\downarrow$$

$$R_{death}$$

The effect of this adjustment (within each U-code, see Addendum 1) is illustrated in Table C.38b. Now those types of disease where mortality rather than quality of life with the disease is the major component can have ratios less than one. Indeed the first term of these ratios (R_{death}) is consistent with (but not equivalent to) the analysis in Section C.2.3.1, where the ratio of quality adjusted net YLLs to unadjusted net YLLs represents this ratio on average for each PBC.

Table C.38b: Examples of adjusted DALY to YLL ratios

Ucode	Adjusted DALY ratios	(R_{death} + R_{alive})
U037 (Other infectious diseases)	1.01	(0.78+0.23)
U016 (Tetanus)	0.78	(0.78+0)
U061 (Mouth and oropharynx cancers)	0.83	(0.78+0.05)
U141 (Spina bifida)	2.18	(0.85+1.34)

Using quality of life estimates (based on HODAR and MEPS)

The disability weights used in the DALY measure (in R_{alive}) are not based on the same description of health states as the EQ5D measure, nor are the weights based on a representative sample of the UK population responding to choice based elicitation questions. EQ5D based quality of life decrements (in relation to age adjusted quality of life norms) associated with different types of disease can be estimated from HODaR and MEPS data for the groups of ICD codes that make up each U-code. The calculations of the quality of life decrements from HODaR were conducted as previously described in Section C.2.2.2. In summary, the average quality of life scores across the ICDs which contribute to each U-code (based on the contributing ICD codes, see Table C.36 and Addendum 1 for how ICD codes map to U-codes) and the average age and gender of respondents from HODaR and MEPS were used to calculate a disease decrement for each U-code, based on quality of life norms from the general population. Note that, by expressing the quality of life effects of different diseases as age related decrements (see Figure C.5), we do not require the HODaR and MEPS samples to necessarily be representative of the age distribution of the population at risk.

The disease related quality of life decrements can then be used to replace the DALY disability weights in R_{alive} reported in Tables C.38a and C.38b. This final adjustment is illustrated in Table C.38c: for example, the evidence about quality of life from HODaR and MEPS suggests that the impact of U037 on quality of life is greater than indicated by DALY disability weights. The quality of life effects of U141, although still very significant, are lower than indicated by DALY disability weights.

Table C.38c: Examples of QALY to YLL ratios (HODaR and MEPS)

Ucode	QALY ratios (HoDAR and MEPs)	(R_{death} + R_{alive})
U037 (Other infectious diseases)	1.37	(0.78+0.60)
U016 (Tetanus)	0.78	(0.78+0
U061 (Mouth and oropharynx cancers)	0.80	(0.78+0.02)
U141 (Spina bifida)	1.88	(0.85+1.03)

By turning what were originally DALY ratios into EQ5D QALY ratios, we regard the QALY to YLL ratios rather than DALY or modified DALY ratios as the preferred basis of estimating a cost per QALY threshold. We consider these estimates to provide a more complete picture of the likely health effects of changes in expenditure.

U-code QALY ratios to ICD QALY ratios

Information about the size and age and gender distribution is only available at U-code level. Therefore U-code ratios are applied to all the ICD codes that contribute to a particular U-code. Note that, unlike ICD codes, U-codes do not map directly to PBCs so some ICDs in different PBCs may belong to the same U-code and therefore have the same U-code ratio. Some ICDs are not included in the U-code classification of disease. Some of these are procedural codes (84 out of 1562)where mortality and QALY effects were not assigned mortality of QALY effects anyway (any health effects would be evident in other ICD codes) so it was not necessary to impute ratios for them. Of the others, some were associated with PBC16 (186 out of 1562) with a zero outcome elasticity so did not require imputation either. Imputation based on the median ratio across the ICDs within the PBC was required for the remaining (482 out of 1562). Eighty eight of these are not mapped into U-codes – these include three big categories of ICDs: Symptoms and signs (R00–R69), Abnormal clinical and laboratory findings, not elsewhere classified (R70–R99) and Ill-defined and unknown causes of mortality (R95–R99). The remaining 394 were associated with U-codes where the ratio was undefined because the denominator (YLL) was zero. In both these cases, values were also imputed based on the median ratio across the ICDs within the PBC. Since the distribution of ratios within a PBC tend to be highly positively skewed, imputation based on the

median is likely to be conservative with respect to health effects and especially in the latter case where mortality effects appear to be a much less important aspect of the disease.
Table C.39 illustrates the variation observed in the ratios (imputed) across ICDs within the same PBC.

Table C.39: Percentiles of the ratio across ICDs, by PBC

		Percentiles of the adjusted DALY ratios						
PBC	*PBC description*	*5%*	*15%*	*25%*	*50%*	*75%*	*85%*	*95%*
2	Cancer	0.76	0.76	0.76	0.81	0.85	0.85	0.91
10	Circulatory problems	0.86	0.86	0.86	0.86	0.96	1.00	2.65
11	Respiratory problems	0.22	0.73	1.00	1.67	1.67	1.96	2.67
13	Gastro-intestinal problems	0.86	0.96	1.01	1.63	1.63	1.78	2.73
1	Infectious diseases	0.00	0.83	1.01	1.01	1.01	1.01	2.64
4	Endocrine problems	0.77	1.37	1.43	2.97	2.97	2.97	2.97
7	Neurological problems	0.86	1.01	1.01	2.01	2.01	2.01	2.30
17	Genito-urinary problems	0.74	0.77	0.77	1.10	1.10	1.10	12.41
18	Maternity	0.00	0.79	0.81	20.39	20.39	20.39	20.39
19	Neonates	1.17	1.17	1.17	1.17	2.29	2.29	2.29
		Percentiles of the QALY ratios (HoDAR and MEPs)						
PBC	*PBC description*	*5%*	*15%*	*25%*	*50%*	*75%*	*85%*	*95%*
2	Cancer	0.76	0.76	0.76	0.79	0.80	0.80	0.83
10	Circulatory problems	0.83	0.83	0.83	0.83	0.94	1.01	1.83
11	Respiratory problems	0.73	0.86	1.37	2.09	2.09	2.24	2.80
13	Gastro-intestinal problems	0.84	1.01	1.37	1.70	1.70	2.17	7.10
1	Infectious diseases	0.83	1.37	1.37	1.37	1.37	1.37	3.26
4	Endocrine problems	0.77	2.37	2.55	5.12	5.12	5.12	10.15
7	Neurological problems	0.84	0.90	1.37	5.90	5.90	5.90	5.90
17	Genito-urinary problems	0.74	0.78	0.78	0.99	0.99	0.99	9.80
18	Maternity	0.81	0.81	0.83	49.30	49.30	49.30	49.30
19	Neonates	0.87	0.87	0.87	0.88	0.88	0.88	0.88

Allocating effects at PBC level to ICD codes

Tables C.38a,b and c illustrate how QALY ratios can be calculated for and differ by U-code. Unsurprisingly, these ratios differ across the type of diseases that make up each PBC (Table C.39).When using this information to estimate a cost per QALY threshold the mortality and life year effects observed at PBC level must be allocated in some way to the component ICD codes before ratios are applied to LY effects and the resulting QALY effects are summed across all the contributing ICD codes.

Alternatively, one could calculate an average of the ratios within a PBC and then apply this 'average ratio' to life year effects at PBC level, rather than calculate QALY effects at ICD level by applying the relevant ratio. This would be inappropriate for two reasons. Firstly, ratios should not be averaged; instead, the total QALYs lost and YLL should be summed across ICDs and the ratio of these sums used to represent a PBC level estimate (i.e., a ratio of averages). Secondly, and even if the appropriate estimate of the QALY to YLL ratio is calculated at the PBC level, this estimate would assume ICDs to be equally representative of the PBC – i.e., that expenditure would be equally likely to affect any of the ICDs that compose a particular PBC. This is unlikely to be true not only due to the inherent differences in the disease described by the ICD coding, but also as ICDs are likely to differ significantly in what concerns the size of the at risk population they represent.

It is important to consider explicitly how other information might inform the different ways in which the effects observed at PBC level might be generated by the distribution of impacts at ICD level, i.e., where investment or disinvestment is likely to occur within the PBC and therefore which ICDs are likely to contribute most to overall health effects. An important and complementary element to the econometric analysis of routinely reported information at PBC level, was to investigate this by looking at local level information available within the NHS. The details of this investigation are reported in Addendum 2. The review of local data sources suggested that there is very little routinely collected data on investment and disinvestment by local NHS organisations beyond the high-level aggregate data on spending by PB which

are used in the econometric analysis. Although more disaggregated data on spending decisions about specific services relevant to particular ICD codes could in principle be acquired through additional primary research (surveys or Freedom of Information requests) this would be costly and with a risk that information acquired in this way may not be complete, consistent or representative.

In the absence of useful information at a local level, it is possible to assume that a change in PBC expenditure will be allocated equally (on a per patient basis) across the component ICD codes, i.e., any investment or disinvestment is equally likely across the population at risk within the PBC. Hospital Episode Statistics (HES) (see Addendum 1) provides information about the costs associated with each ICD by PCT. The variation in per patient costs between PCTs (where total costs allocated to individual ICDs were divided by the number of patients using services in the PCT) was analysed to establish which ICDs contribute most to the variability in HES costs within a PBC, across PCTs. The ICDs that contribute most to this variance might be expected to be more likely to have been subject to differential investment or disinvestment across PCTs. Unfortunately total PBC costs are not available at ICD level across PCTs so could not be used for this purpose. Costs from HES data are only a component of total PBC costs (41% of total PBC costs for the 11 PBCs where mortality effect can be estimated) and contribute less to the variability in PBC costs across PCTs (HES contribute only 23% of the variability for the 11 PBCs where mortality effect can be estimated).

There are differences in relative weight assigned to ICDs based solely on the size of the population or its contribution to variance in HES costs. If investment or disinvestment within a PBC tends to focus on ICD codes representing areas of marginal value the health effects of a change in PBC expenditure may be overestimated and a cost per QALY threshold underestimated when allocating effects equally across the population at risk within each PBC. However, weighting ICDs based on HES data is likely to favour those ICDs which represent more severe disease requiring more hospital care. This may over represent ICDs with lower QALY to YLL ratios if mortality effects tend to be a major component of these types of disease and maybe conservative with respect to the health effects of changes in expenditure.

The implications for a cost per QALY threshold that uses the estimated mortality and life year effects as a surrogate for a more complete measure of the likely heath effects (i.e., that includes quality of life as well as quality adjusted life year effects) is summarised in Table C.40 and detailed in Table C.41. These results use the contribution to variance in HES costs to 'weight' the different ICD codes within a PBC (allocate the life year effects), before applying the QALY ratios associated with each ICD (see footnote [12] reporting results using weights based on size of the population).

Table C.40: Summary of the QALY threshold using ratios

	Cost per QALY threshold		
	DALY ratios [1]	Adjusted DALY ratios [2]	QALY ratios (HoDAR and MEPs) [3]
big 4 PBC's	£5,402	£6,419	£5,990
11 PBCs (with mortality)	£9,958	£11,718	£10,297
All 23 PBCs	£11,254	£13,244	£11,638 *

* Preferred analysis

[12] The table below reports the cost per QALY threshold using a relative weight based on the size of the ICD population to allocate health effects.

	Cost per QALY threshold		
	DALY ratios [1]	Adjusted DALY ratios [2]	QALY ratios (HoDAR and MEPs) [3]
big 4 PBC's	£4,400	£5,100	£2,340
11 PBCs (with mortality)	£8,066	£9,267	£4,212
All 23 PBCs	£9,117	£10,474	£4,760

Table C.41: Breakdown of the QALY threshold using ratios by PBC

PBC	PBC description	Change in spend, £m [1]	Adjusted DALY ratios: Change in QALY [2]	Adjusted DALY ratios: Cost per QALY gained, £ [3]	QALY ratios (HoDAR and MEPs): Change in QALY [4]	QALY ratios (HoDAR and MEPs): Cost per QALY gained, £ [5]
2	Cancer	£19	1 763	£10,871	1 699	£11,283
10	Circulatory problems	£33	7 677	£4,334	6 713	£4,956
11	Respiratory problems	£22	2 379	£9,375	3 215	£6,937
13	Gastro-intestinal problems	£17	2 396	£6,886	3 605	£4,577
	Big 4			£6,419		£5,990
1	Infectious diseases	£8	21	£388,430	27	£305,724
4	Endocrine problems	£18	1 077	£16,396	2 036	£8,673
7	Neurological problems	£17	296	£58,158	342	£50,295
17	Genito-urinary problems	£32	15	£2,158,296	12	£2,623,379
16	Trauma & injuries*	£10	0	NA	0	NA
18+19	Maternity & neonates*	£8	125	£64,173	273	£29,327
	First 11 PBC's			£11,718		£10,297
3	Disorders of Blood	£11	956	£11,718	1 087	£10,297
5	Mental Health Disorders	£204	17 423	£11,718	19 828	£10,297
6	Learning Disability	£31	2 627	£11,718	2 990	£10,297
8	Problems of Vision	£24	2 063	£11,718	2 348	£10,297
9	Problems of Hearing	£6	546	£11,718	621	£10,297
12	Dental problems	£23	2 005	£11,718	2 282	£10,297
14	Skin	£11	897	£11,718	1 021	£10,297
15	Musculo skeletal system	£15	1 291	£11,718	1 469	£10,297
20	Poisoning and adverse effects	£4	375	£11,718	426	£10,297
21	Healthy Individuals	£18	1 565	£11,718	1 781	£10,297
22	Social Care Needs	£68	5 769	£11,718	6 566	£10,297
23	Other	£78	0	NA	0	NA
	All (23 PBCs)			£13,244		£11,638

Since all the analysis in this Section seeks to use the estimated mortality and life year effects as a surrogate for a more complete measure of likely health effects, it is the cost per QALY threshold for all 23 PBCs that is most relevant. As expected, this threshold (£11,638), is lower than a cost per QALY threshold based only the quality adjusted life year effects (£21,047 and £26,441 in Table C.34 that assumes no effects of NHS expenditure on quality of life itself). This difference gives some indication of the relative importance of QALY effects due to avoidance of premature death and the QALY effects of avoiding disability during disease. Table C.42 reports how the estimated QALY effects for each PBC can be decomposed into that part associated with quality adjusted life year effects and that part associated with 'pure' quality of life effects. These results appear credible for the first 11PBCs, where those for which mortality is the major concern have a much greater share of total QALY effects associated with avoidance of premature death (e.g., PBC2 and PBC10) compared to those where quality of life is the major concern (e.g., PBC 7).

Table C.42: Decomposing estimated QALY effects by PBC

PBC		QALY change (total)	QALY change (death)	% QALY gained due to avoidance of premature death	% QALY gained due to avoidance of disability while alive
2	Cancer	1,699	1,641	97%	3%
10	Circulatory	6,713	4,856	72%	28%
11	Respiratory	3,215	923	29%	71%
13	Gastro-intestinal	3,605	1,193	33%	67%
1	Infectious diseases	27	11	40%	60%
4	Endocrine	2,036	323	16%	84%
7	Neurological	342	52	15%	85%
17	Genito-urinary	12	6	52%	48%
16	Trauma & injuries*	0	0	NA	NA
18+19	Maternity & neonates*	273	15	6%	94%
3	Disorders of Blood	1,087	547	50%	50%
5	Mental Health	19,828	9,979	50%	50%
6	Learning Disability	2,990	1,505	50%	50%
8	Problems of Vision	2,348	1,181	50%	50%
9	Problems of Hearing	621	313	50%	50%
12	Dental problems	2,282	1,148	50%	50%
14	Skin	1,021	514	50%	50%
15	Musculo skeletal	1,469	739	50%	50%
20	Poisoning and AE	426	215	50%	50%
21	Healthy Individuals	1,781	896	50%	50%
22	Social Care Needs	6,566	3,304	50%	50%
23	Other	0	0	NA	NA

The ratios of QALYs to YLL due to disease in those PBC where outcome elasticities could not be estimated cannot be used to inform estimates of the threshold because there are no estimated life year effects with which to apply the ratios. Therefore, as in previous sections, the estimated effect of expenditure on health for the 11 PBCs with outcome elasticities is applied to the estimated changes in PBC expenditure for the other 12 PBCs (excluding GMS for the reasons given in Section C.2.1.5), i.e., assuming that the health effects that can be observed of a change in expenditure will be similar to those that cannot. However, the use of QALY ratios also implies that the share of total health effects between quality adjusted life year effects and that part associated with 'pure' quality of life effects are also similar to those PBC with estimated outcome elasticities. Summing the different types of health effects across these 11PBCs suggests that 50% is due to avoidance of premature death and 50% due to avoidance of disability. This is clearly not credible when applied to the other PBCs, e.g., mental health, vision and hearing are likely have a much greater share of total health effects associated with quality of life effects and very little associated with premature mortality.

By comparing the change in QALY in each PBC (that originates cost per QALY threshold estimates, column 2 in Table C.43), with the corresponding change in YLL (column 6, Table C.16), we can infer the implied QALY to YLL ratio in each of the PBCs with a mortality signal. These are shown in Table C.40. The QALY to YLL ratio implied by the analysis using QALY ratios for all 11 PBC with outcome elasticities is 1.52, which suggests that every life year is associated with 1.52 QALYs on average across these PBCs. However, this implied QALY ratio differs across these PBCs, ranging from 0.79 in PBC2 to 15.05 in PBC18+19 (see column 4 of Table C.43). It should be noted that the implied QALY ratio of 1.35 for the 11 PBC with outcome elasticities is a ratio of QALYs to unadjusted YLL. The proportion of total QALY effects due to premature deaths for the same PBCs (50% in Table C.41) also implies a ratio - equal to two. However, this is a ratio of total QALY effects to quality adjusted YLL. The difference between these two ratios is the denominator, i.e., quality adjusted YLL are lower than unadjusted YLL.

Table C.43: Implied QALY to YLL ratios.

PBC	PBC description	Adjusted DALY ratios: Implied QALY per LY gained [1]	Adjusted DALY ratios: Implied QALY per excess death averted [2]	Adjusted DALY ratios: Implied QALY per PBC death averted [3]	QALY ratios (HoDAR and MEPs): Implied QALY per LY gained [4]	QALY ratios (HoDAR and MEPs): Implied QALY per excess death averted [5]	QALY ratios (HoDAR and MEPs): Implied QALY per PBC death averted [6]
2	Cancer	0.82	11.58	8.48	0.79	11.16	8.17
10	Circulatory problems	1.20	12.51	6.20	1.05	10.94	5.42
11	Respiratory problems	1.96	18.09	2.04	2.65	24.45	2.76
13	Gastro-intestinal problems	1.54	23.02	14.49	2.31	34.63	21.80
	Big 4	1.26	14.20	5.12	1.35	15.21	5.49
1	Infectious diseases	1.56	20.64	8.30	1.98	26.22	10.54
4	Endocrine problems	2.62	35.61	19.84	4.95	67.31	37.51
7	Neurological problems	4.56	61.99	27.90	5.27	71.69	32.26
17	Genito-urinary problems	1.75	14.56	2.98	1.44	11.98	2.45
16	Trauma & injuries*	NA	NA	NA	NA	NA	NA
18+19	Maternity & neonates*	6.88	511.33	511.33	15.05	1118.85	1118.85
	First 11 PBC's	1.33	15.16	5.53	1.52	17.26	6.29

The problem is that using QALY to YLL ratios means that much of the information that is available about the other 12 PBCs cannot be used to inform the estimates of the cost per QALY threshold. Fortunately, the sources of information on which ratios are based also provide much of the information required to calculate the QALY burden of disease in these areas. Section C.2.3.2 explores how measures of burden can be used to estimate a cost per QALY threshold that captures the likely effects of a change in expenditure on all aspects of health while using all the information that is available about all the PBCs.

C.2.3.2 Using estimates of the QALY burden of disease

In this Section we use estimates of the QALY burden of disease, infer a proportionate effect on burden from the observed effects on life years, and then apply this proportionate effect to the measures of QALY burden for all PBCs. In this way we can use all the information available about the mortality and quality of life effects of the different types of disease that make up each PBC, particularly for those where mortality based outcome elasticities are not available. Applying a proportionate effect to measures of QALY burden of disease is equivalent to assuming that any effects on life years are lived at quality of life that reflects a proportionate improvement to the quality of life with disease. It also allows quality of life effects of changes in expenditure to be included; also based on proportionate improvement in the quality of life with disease.

In Section C.2.2, each life year gained could be assumed to be lived in full health, lived in a quality of life that reflects age and gender norms of the general population or lived in a quality of life that reflects the original disease state. Applying an estimated proportionate effect on the life year burden of disease to measures of QALY burden of disease implies a proportionate improvement in the quality of life with disease applied to any life year effects. Therefore, basing estimates on measures of QALY burden provides are more conservative estimate of the QALY effects of changes in mortality than the best estimate reported in Section C.2.2, which was based on quality of life norms.

The total QALY burden of disease for the population with disease in a particular year includes: i) the quality adjusted years of life lost due to all the disease related mortality that could occur in this population over their remaining duration of disease and ii) the reduction in quality of life while alive also for their remaining disease duration. These components of burden represent, respectively, the QALY lost due to premature death ($QALYl_{death}$) and the QALY lost while alive ($QALYl_{alive}$) as a consequence of disease.

$$Burden = QALYl_{death} + QALYl_{alive} \quad (7)$$

However, applying the estimated proportionate effects on mortality and life years to such a measure of total burden would provide an estimate of the effects of a change in expenditure, not just in one year, but

in all the remaining years of disease for the population at risk in that year. Recall from Section C.2.1 that we have adopted the conservative assumption that changes in expenditure will only have health effects in one year for the population with disease in that year. Therefore, it is not a measure of total burden that is required, but a measure of the QALY burden of disease during one year for the population with disease (prevalent and incident) in that year. The estimated outcome elasticities can then be appropriately (and directly) applied to this measure of burden. Of course, it would be possible to solve for a lower outcome elasticity that could be applied to total burden which would return the required estimate of total QALY effects restricted to one year.

The information from GBD used to derive QALY ratios in Section C.2.3.2 includes information about the YLL and duration of disease for those incident to a U-code, i.e., the measure of QALY burden from the information included in the ratios is a measure of the total burden of the disease but only for the population that is incident (rather than total population with disease) in one year. Assuming that incidence is stable over the disease duration this is also equivalent to the QALY burden of disease during one year for the population with disease (i.e., those that are incident and prevalent) in that year. This is valid as long as estimates of the quality of life decrement of disease from HODaR and MEPS are assumed representative of average effects across those earlier (incident) and later (prevalent) in their disease duration.

However, in moving from ratios to absolute measures of burden it becomes more important to examine and then adjust for any inconsistency between information about YLL and size of the incident population from GBD (which is available by U-codes and can be mapped to ICDs), and the information about net YLL and observed deaths for each PBC based on ONS data as described in Section C.2.2.3 – see Table C.44.

Table C.44: Comparing deaths and YLL from ONS and GBD.

			deaths			YLL		
		Excess deaths ONS	All deaths ONS	All deaths GBD*	adjustment factor (deaths)	Net estimates ONS	Total YLL GBD*	adjustment factor (YLL)
		[1]	[2]	[3]	[4]	[5]	[6]	[7]
1	Infectious diseases	2,797	6,958	1,408	4.94	36,962	25,142	1.47
2	Cancer	95,715	130,810	140,124	0.93	1,347,184	1,932,637	0.70
4	Endocrine	3,769	6,765	7,509	0.90	51,225	95,401	0.54
7	Neurological	6,909	15,353	12,854	1.19	93,917	164,796	0.57
10	Circulatory	79,218	159,852	178,454	0.90	823,768	1,750,608	0.47
11	Respiratory	7,386	65,446	67,441	0.97	68,030	594,529	0.11
13	Gastro-intestinal	15,199	24,147	28,329	0.85	227,703	396,829	0.57
17	Genito-urinary	2,172	10,625	8,606	1.23	18,127	77,338	0.23
18+19	Maternity & neonates	226	226	2,211	0.10	16,801	149,868	0.11
	Total	*213,391*	*420,182*	*446,936*	*0.94*	*2,683,717*	*5,187,148*	*0.52*

There are a number of reasons for the potential inconsistencies: i) GBD is based on earlier years of mortality data; ii) the imprecision of mapping from U-codes to PBC via ICD codes; and iii) the YLL reported in GBD are based on life expectancy at the age of death (see Section C.2.2 and C.2.3) and will overestimate the net YLL. The YLL by U-code, reported in GBD, that are mapped to ICDs are adjusted by these proportionate differences (column 7 of Table C.44) to ensure that the YLLs associated with all contributing ICD codes are consistent with (do not overestimate) the net YLL for the PBC as a whole. The variation across ICDs in the adjusted QALY burden associated with mortality gains (for the population with disease in a particular year) is depicted in column 2 of Table C.45.

Table C.45: Variation across ICDs of the QALY burden of disease for a patient with disease in a particular year

PBC		Burden while alive	Burden due to premature death	Burden
		Median [5th to 95th percentile]		
		[1]	[2]	[3]
1	Infectious diseases	0.47 [0.00 to 0.82]	0.25 [0.11 to 42.87]	0.72 [0.72 to 42.87]
2	Cancers and Tumours	0.09 [0.00 to 0.10]	2.82 [0.51 to 5.11]	2.92 [0.58 to 5.11]
3	Disorders of Blood	0.05 [0.05 to 0.07]	0.01 [0.01 to 0.03]	0.06 [0.06 to 0.09]
4	Endocrine	0.10 [0.00 to 0.17]	0.01 [0.01 to 4.82]	0.11 [0.11 to 4.82]
5	Mental Health	0.10 [0.07 to 0.22]	0.02 [0.00 to 0.04]	0.12 [0.07 to 0.26]
6	Learning Disability	0.10 [0.06 to 0.10]	0.02 [0.00 to 5.34]	0.12 [0.10 to 5.41]
7	Neurological	0.27 [0.00 to 0.37]	0.02 [0.02 to 22.79]	0.29 [0.25 to 22.79]
8	Vision	0.05 [0.00 to 0.06]	0.00 [0.00 to 20.99]	0.05 [0.03 to 20.99]
9	Hearing	0.05 [0.00 to 0.05]	0.00 [0.00 to 20.99]	0.05 [0.00 to 20.99]
10	Circulation	0.09 [0.09 to 0.19]	0.37 [0.06 to 0.39]	0.48 [0.18 to 0.56]
11	Respiratory system	0.14 [0.00 to 0.21]	0.01 [0.00 to 5.17]	0.15 [0.00 to 5.18]
12	Dental	0.03 [0.01 to 0.03]	0.01 [0.00 to 0.01]	0.04 [0.01 to 0.04]
13	Gastro intestinal system	0.10 [0.00 to 0.18]	0.05 [0.00 to 23.67]	0.15 [0.00 to 23.67]
14	Skin	0.06 [0.00 to 0.06]	0.02 [0.02 to 20.99]	0.08 [0.08 to 20.99]
15	Musculo skeletal system	0.10 [0.00 to 0.10]	0.02 [0.00 to 20.99]	0.12 [0.06 to 20.99]
16	Trauma and injury	NA	NA	NA
17	Genito Urinary system	0.11 [0.00 to 0.13]	0.04 [0.00 to 8.90]	0.15 [0.05 to 8.90]
18	Maternity	0.01 [0.00 to 0.01]	0.00 [0.00 to 4.68]	0.01 [0.00 to 4.68]
19	Conditions of neonates	0.00 [0.00 to 0.00]	0.03 [0.02 to 0.03]	0.03 [0.02 to 0.03]
20	Poisoning and AE	0.03 [0.00 to 0.06]	0.00 [0.00 to 18.63]	0.03 [0.02 to 18.63]
21	Healthy Individuals	0.05 [0.05 to 0.05]	0.01 [0.01 to 0.01]	0.06 [0.06 to 0.06]

Note: QALY burden of disease reflects burden while alive in one year and mortality burden in one year. Any mortality effects of disease in one year can lead to the loss of more than one life year, and for this reason burden due to premature death may assume values bigger than 1.

It is QALY burden per patient with disease in a particular year that is reported in this Table, including the median and range across the ICD codes contributing to each PBC. Such measure of burden considers quality of life burden while alive in one year and mortality burden for the same time period. Note that mortality effects of disease in one year can lead to the loss of more than one life year, and for this reason burden due to premature death (and consequently overall burden) may assume values bigger than 1. Burden values in this Table reflect variation across ICDs and should not be misinterpreted as the 'average' QALY burden for the PBC, as this depends on how PBC effects are allocated to ICDs and the 'average' burden for groups of PBCs depends on how a change in overall expenditure is shared between them, i.e., the expenditure elasticities estimated for each PBC in Appendix B.

Due to the earlier years of data and imprecision in mapping from U-codes to ICDs there might also be some inconsistency in estimates of the total incidence of disease for a PBC. Insofar as disease related mortality risk is stable, the same number of deaths should be observed in GBD and ONS data for the same at risk population. The PBC deaths recorded in GBD and those observed in ONS data (columns 2 and 3 in Table C.44) are similar but nonetheless the proportionate difference is used to adjust the scale of quality of life burden while alive based on GBD information (equivalent to adjusting estimates of incidence). Notable exceptions are PBC1 and PBC18+19 where the discrepancies are likely to be due to imperfect mapping from U-code to PBC via ICD codes. Summaries of the ICD specific values of the adjusted burden of disease while alive are depicted in column 1 of Table C.45. Total burden (for the population with disease in a particular year) is the sum of the two components of burden (Table C.46 presents a few examples for illustration).

Table C.46: Examples of QALY burden of disease for the population with disease in a particular year

Ucode	QALY burden	(QALY $lost_{death}$ + QALY $lost_{alive}$)
U037 (Other infectious diseases)*	0.20	(0.09+0.11)
U016 (Tetanus)	2.73	(2.73+0.00)
U061 (Mouth and oropharynx cancers)	2.97	(2.87+0.10)
U141 (Spina bifida)	0.65	(0.18+0.46)

Note: QALY burden of disease reflects burden while alive in one year and mortality burden in one year. Any mortality effects of disease in one year can lead to the loss of more than one life year, and for this reason burden due to premature death may assume values bigger than 1.

*Note that differential adjustments have been made to YLL (affecting QALY $lost_{death}$) and to the incidence (affecting QALY $lost_{alive}$), thus implied ratios from these burden estimates may differ from ratios presented is Section C.2.3.1.

The implications for the cost per QALY threshold of using information about the QALY burden of disease for all PBCs, rather than QALY ratios for those where an outcome elasticity can be estimated, are reported summarily in Table C.47 and in detail in Table C.48. The QALY effects of a change in PBC expenditure are a weighted average of the QALY effects within each of the ICDs that contribute to the PBC. The figures reported in column 2 are based on weighing the effects at ICD level by the proportion of the total PBC population within each contributing ICD code, rather than the contribution to variance in HES costs .[13]

Table C.47: Summary of the cost per QALY threshold

	Cost per QALY gained	
	QALY ratios (HoDAR and MEPs) [1]	QALY burden (HoDAR and MEPs) [2]
big 4 PBC's	£5,990	£3,036
11 PBCs (with mortality)	£10,297	£5,128
All 23 PBCs	£11,638*	£15,701*

* Preferred analysis

The cost per QALY threshold for the 11PBCs with outcome elasticities is a little lower using a measure of QALY burden (£5 128) rather than the QALY ratios (£10,297) described in Section C.3.2.1. This is in part because the way GBD calculates YLL overestimates net YLL (which accounts for counterfactual deaths, as detailed in Section C.2.2.3). This will make no difference to the first term in the QALY ratio (R_{death}) used in Section C.2.3.1 since an overestimate of YLL affects both denominator and numerator of the ratio. However, the second term (R_{alive}) is likely to be underestimated. Therefore the ratios will tend to underestimate the QALY effects of expenditure and overestimate the cost per QALY threshold (see Table C.47). We are able to adjust the GBD based measure of QALY burden for this overestimation in calculating the QALY threshold reported in column 2.

Since the purpose of this Section is to use the estimated mortality and life year effects as a surrogate for a more complete measure of likely health effects, it is the cost per QALY threshold for all 23 PBCs that is of most relevance. The cost per QALY threshold for all 23 PBCs is based on applying the proportionate effects on the QALY burden of disease, based on the observed effects of changes in expenditure on mortality in the 11 PBC with outcome elasticities,[14] to the QALY burden of disease in the other PBCs. This generates a much higher cost per QALY threshold (£15 701) than the one based on applying the estimated QALY effects of changes in expenditure, using QALY ratios for the 11 PBC with outcome elasticities, to changes in expenditure in the others (£11,638). The reason is that the QALY burden of disease in the other PBCs is, in general, lower than the QALY burden of disease across those PBCs where outcome elasticities can be estimated (see Table C.45 above).

[13] IHES costs are a much smaller proportion of total PBC expenditure for the 11 PBCs where a mortality effects could not be estimated (HES costs account for less than 15% of total PBC expenditure) and account for very little of a the variability in PBC costs across PCTs (the contribution that variance in HES costs makes to variance in PBC expenditure in this group of PBCs is less than 8%). Therefore, allocating PBC level effects to ICDs based on contribution to variance in HES costs is less appropriate when information about QALY burden in this groups of PBCs is used to inform the estimate of the overall threshold.

[14] Note that this is the ratio of total change in health to total change in expenditure across these PBC (rather than an average ratio) and the contribution that each of these PBCs make to these total effects on health and expenditure depends on the estimated expenditure as well as outcome elasticities.

Therefore, applying the same proportionate effects to a lower QALY burden generates a smaller health effect of a change in expenditure.[15] In essence the difference between these estimates is that in column 1 of Table C.47 the absolute effect on health associated with an absolute change in expenditure is extrapolated to the other PBCs, where as in column 2 it is the relative effect on health of an absolute change in expenditure that is extrapolated. Since we know that QALY burden differs between (and within) PBCs and especially between the groups of PBCs with and without estimated outcome elasticities (see Table C.45), it is the values based on QALY burden in column 2 of Table C.47 that are regarded as most credible and represent our central or best estimate.

A detailed breakdown of changes in expenditure and changes in QALYs across all PBCs is shown in Table C.48 when the analysis is based on QALY ratios and on QALY burden of disease. A comparison of these values confirms that QALY effects for the other PBC are lower and therefore the cost per QALY for each of these PBCs are in general much higher when based on a proportionate effect on QALY burden. Of course, we have not directly observed quality of life effects in these PBC but inferred them from the proportionate effects that we can observe. Insofar as investment and disinvestment opportunities in these PBCs might have been more valuable (offered greater improvement in quality of life) than suggested by the implied PBC thresholds, then overall QALY effects will tend to be underestimated and the cost per QALY threshold overestimated.

Table C.48: Breakdown of the cost per QALY threshold

			QALY ratios (HoDAR and MEPs)		QALY burden (HoDAR and MEPs)	
PBC	PBC description	Change in spend, £m [1]	Change in QALY [2]	Cost per QALY gained, £ [3]	Change in QALY [4]	Cost per QALY gained, £ [5]
2	Cancer	£19	1 699	£11,283	1 501	£12,772
10	Circulatory problems	£33	6 713	£4,956	5 908	£5,631
11	Respiratory problems	£22	3 215	£6,937	19 869	£1,123
13	Gastro-intestinal problems	£17	3 605	£4,577	2 776	£5,944
	Big 4			£5,990		£3,036
1	Infectious diseases	£8	27	£305,724	53	£158,349
4	Endocrine problems	£18	2 036	£8,673	4 887	£3,613
7	Neurological problems	£17	342	£50,295	963	£17,844
17	Genito-urinary problems	£32	12	£2,623,379	24	£1,320,516
16	Trauma & injuries	£10	0	NA	0	NA
18+19	Maternity & neonates	£8	273	£29,327	10	£813,578
	First 11 PBC's			£10,297		£5,128
3	Disorders of Blood	£11	1 087	£10,297	689	£16,257
5	Mental Health Disorders	£204	19 828	£10,297	3 397	£60,111
6	Learning Disability	£31	2 990	£10,297	125	£247,001
8	Problems of Vision	£24	2 348	£10,297	240	£100,871
9	Problems of Hearing	£6	621	£10,297	434	£14,718
12	Dental problems	£23	2 282	£10,297	489	£48,002
14	Skin	£11	1 021	£10,297	107	£98,620
15	Musculo skeletal system	£15	1 469	£10,297	1 697	£8,913
20	Poisoning and adverse effects	£4	426	£10,297	54	£81,782
21	Healthy Individuals	£18	1 781	£10,297	23	£811,562
22	Social Care Needs	£68	6 566	£10,297		NA
23	Other	£78	0	NA		NA
	All (23 PBCs)			£11,638		£15,701

For the reasons discussed in previous sections, we regard all the cost per QALY threshold reported in column 2 of Table C.47 to be on balance conservative with respect to overall health effects of a change in expenditure. However, the estimate of £15 701 may be especially conservative with respect to health effects (i.e., overestimated) based, as it is on an extrapolation of the proportionate effects to measures of burden on these PBC, rather than observations of the direct impact of changes in expenditure on quality of life in these types of disease. This is especially so in PBC 5, Mental Health Disorders, which accounts

[15] Applying the absolute health effect of expenditure from the 11 PBCs with outcome elasticities implies different (higher) proportionate effects in the other PBCs

for a large proportion of the change in overall expenditure (30%) and where a review of the evidence suggests that the investment and disinvestment opportunities in this PBC are likely to have been more valuable than the implied PBC cost per QALY of £60,111 (see Addendum 3 to this Appendix). The lower cost per QALY threshold for the 11PBCs with outcome elasticities (£5 128) might be regarded as more secure in this respect but they only account for a proportion (28%) of any change in overall expenditure (see Table C.53). [16]

Table C.49 reports how the estimated QALY effects based on measures of QALY burden for each PBC can be decomposed into that part associated with life year effects adjusted for quality and that part associated with 'pure' quality of life effects. These results are very similar to those reported in Table C.40 which were based on QALY ratios for the 11 PBCs with an estimated outcome elasticity. Those PBCs for which mortality is the major concern have a much greater share of total QALY effects associated with avoidance of premature death (e.g., PBC2 and PBC10) compared to those where quality of life is the major concern (e.g., PBC 7). The differences tend to favour QALYs gained though avoidance of disability, which reflects the underestimation of the effects on 'pure' quality of life when using QALY ratios based on estimates of YLL from GBD (see the discussion above). The exceptions are PBC 1 and PBC 18 &19. The reason is that there are significant adjustments made based on differences in observed and recorded mortality (to adjust for differences in recording) as well as differences in YLL due to the GBD method of calculation (see Table C.42).

Table C.49: Decomposing estimated QALY effects by PBC

PBC		QALY change (total) [1]	QALY change (death) [2]	% QALY gained for premature death [3]	% QALY gained for disability while alive [4]
2	Cancer	1,501	1,393	93%	7%
10	Circulatory	5,908	4,054	69%	31%
11	Respiratory	19,869	758	4%	96%
13	Gastro-intestinal	2,776		37%	63%
1	Infectious diseases	53	9	18%	82%
4	Endocrine	4,887	269	5%	95%
7	Neurological	963	43	4%	96%
17	Genito-urinary	24	5	22%	78%
16	Trauma & injuries*	0	0	NA	NA
18+19	Maternity & neonates*	10	7	69%	31%
3	Disorders of Blood	689	35	5%	95%
5	Mental Health	3,397	296	9%	91%
6	Learning Disability	125	25	20%	80%
8	Problems of Vision	240	9	4%	96%
9	Problems of Hearing	434	3	1%	99%
12	Dental problems	489	0	0%	100%
14	Skin	107	39	37%	63%
15	Musculo skeletal	1,697	84	5%	95%
20	Poisoning and AE	54	9	16%	84%

[16] It is not possible to estimate expenditure equations for all 23 PBCs simultaneously (see Section 5.8), so the 23 independently estimated expenditure elasticities may not necessarily account for all of a change in overall spend, i.e., the sum of changes in PBC expenditure based on a 1% change in total spend and the estimated PBC expenditure elasticities is less than a 1% change in total spend. Previously in Chapter 3 and Section 4.2, 4.3 and 4.4.1 any remaining change in total spend was assigned to the other 11 PBCs where outcome elasticities could not be estimated (in these Sections expenditure elasticities for these PBCs were not estimated because the same health effect of expenditure was assumed for these PBCs so it did not matter how spend was allocated between them). However, in this section it does matter how the remaining change in expenditure is allocated between the other 11 PBCs as they have different QALY burdens so different implied health effects of expenditure. Therefore, the remaining change in total spend is allocated between these 11 PBCs reflecting the relative share of changes in expenditure based on their estimated expenditure elasticities. This does mean that a greater proportion of a change in overall expenditure tends to be allocated to this group of PBCs. Since these PBCs tend to have lower QALY burden and a higher implied PBC cost per QALY this will tend to overestimate the overall cost per QALY threshold.

21	Healthy Individuals	23	4	16%	84%
22	Social Care Needs	0	0	NA	NA
23	Other	0	0	NA	NA

The implied QALY per life year gained and death averted are reported in Table C.50. As expected, the implied QALY per PBC death averted across all 11PBCs with outcome elasticities is higher (12.6 QALY) than reported in Section C.2.3.1 (6.3 QALY) because of the previous bias against quality of life effects.

Table C.50: Implied QALY per excess death averted: using burden, contribution to variance

PBC	PBC description	QALY per LY gained [1]	Implied QALY per excess death averted [2]	Implied QALY per PBC death averted [3]
2	Cancer	0.70	9.86	7.21
10	Circulatory problems	0.93	9.63	4.77
11	Respiratory problems	16.40	151.10	17.05
13	Gastro-intestinal problems	1.78	26.66	16.78
	Big 4	2.66	30.02	10.82
1	Infectious diseases	3.83	50.62	20.35
4	Endocrine problems	11.89	161.59	90.04
7	Neurological problems	14.86	202.05	90.92
17	Genito-urinary problems	2.85	23.80	4.87
16	Trauma & injuries	NA	NA	NA
18+19	Maternity & neonates	0.54	40.33	40.33
	First 11 PBC's	3.05	34.65	12.63

In Section C.2.3.1, the ratios of QALYs to YLL due to disease in those PBCs where outcome elasticities could not be estimated could not be used to inform estimates of the threshold or indicate how any total health effects in these other PBCs are likely to be 'shared' between life year effects adjusted for quality and that part associated with 'pure' quality of life effects (see Table C.42). By applying the observed proportionate effects of changes in expenditure to measures of QALY burden of disease in these other PBCs the likely share of any effects on QALYs between avoidance of premature mortality and avoidance of disability more closely reflect the nature of these types of diseases (see Table C.49). As expected, a much greater proportion of QALY effects are associated with quality of life during the disease compared to the 11PBCs where mortality based outcome elasticities could be estimated. The share of effects in particular PBCs are also much more credible. For example, in PBC5 Mental Health Disorders the overwhelming share of QALY effects are associated with quality of life itself and for others, such as PBC12 Dental problems, PBC9 Problems of Hearing and PBC8 Problems of Vision; almost all effects are associated with quality of life rather than mortality and life years. For this, and the other reasons discussed above, the analysis based on measures of QALY burden are regarded as the best estimate of a cost per QALY ratio that reflects a more complete picture of the likely health effects of changes in overall expenditure.

C.2.3.3 Summary of the cost per QALY threshold

The results of the three sequential steps of analysis described in this Chapter are summarised in Table C.51. In Section C.2.1 we explored ways in which the estimated effects on mortality from the econometrics work in Appendix B might be better translated in to life year effects by overcoming some of the limitations of mortality data available at PCT level and taking account of counterfactual deaths. The results of this analysis were reported in Table C.21 and are repeated in column 1 of Table C.51. These results can be interpreted as cost per QALY thresholds conditional on the assumption that all life years are lived in full health and the quality of life with disease is zero (equivalent to death).

Table C.51: Summary of cost per QALY threshold estimates

	[1] (Table C.20)	[2] (Table C.34)	[3]	
QoL associated with life extension:	*1*	*Norm*		
QoL during disease:	*0*	*0*	*Based on burden*	
			Best estimate	
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per death averted:	*~ 4.1 YLL*	*~ 4.1 YLL*	*~ 4.1 YLL*	
QALYs per death averted:	*~ 4.1 QALY*	*~ 3.5 QALY*[1]	*~ 12.6 QALY*	
big 4 PBC's	£8,080	£9,631	£3,036	[1]
11 PBCs (with mortality)	£15,628	£18,622	£5,128	[2]
All 23 PBCs	£17,663	£21,047	**£15,701**	[3]
			Lower bound	
Effect of expenditure on mortality:	*Remainder of disease duration*	*Remainder of disease duration*	*Remainder of disease duration*	
YLL per death averted:	*~ 4.1 YLL*	*~ 4.1 YLL*	*~ 4.1 YLL*	
QALYs per death averted:	*~ 4.1 QALY*	*~ 3.5 QALY*	*~ 12.6 QALY*	
big 4 PBC's	£3,846	£4,252	£674	[4]
11 PBCs (with mortality)	£6,106	£6,852	£860	[5]
All 23 PBCs	£6,901	£7,744	**£2,785**	[6]
			Upper bound	
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per death averted:	*2 YLL*	*2 YLL*	*2 YLL*	
QALYs per death averted:	*~ 2 QALY*	*~ 1.9 QALY*	*~ 6.1 QALY*	
big 4 PBC's	£16,432	£17,456	£6,292	[7]
11 PBCs (with mortality)	£32,387	£34,492	£10,626	[8]
All 23 PBCs	£36,604	£38,983	**£32,537**	[9]

In Section C.2.2 we considered how the estimated life year effects might be adjusted for the quality of life in which they are likely to be lived, taking account of the gender and the age at which life years are gained or lost (see Table C.34). The results of this analysis are repeated in column 2 below. Finally, in the current Section, C.2.3, we explored ways to also take account of the likely effects of changes in expenditure on quality of life during disease as well as the effects associated with mortality and life years (see column 3). These estimates provide our central estimate of a cost per QALY threshold, because they make best use of available information while the assumptions required, which on balance are likely conservative with respect to health effects, appear more reasonable than the other alternatives available.[17]

The estimate of £5,128 per QALY (line 2) is restricted to the effects of changes in expenditure in the 11PBCs where outcome elasticities can be estimated. Although this might be regarded as more secure these PBCs only account for a proportion of the change in overall expenditure (approximately 28%, see Section C.2.4). The threshold of £15,701 uses the estimated proportionate effects of expenditure on the QALY burden of disease in these PBC as a surrogate for proportionate effects in the others, i.e., assuming that the effects that can be observed will be similar to those that cannot. As discussed in Section C.2.3.2, there are reasons to suspect that this may underestimate health effects in these PBCs which have most influence on the overall threshold. As in previous sections, no health effects are assigned to PBC23 (General Medical Services) on the basis that any health effects of this expenditure would be recorded in the other PBCs. [18] Therefore, the best or central estimate of cost per QALY threshold is £15,701 (column 3, line 3).

[17] Note that the proportionate difference between the estimates in column 3 and columns 1 and 2 are greater in lines 1 and 2, reflecting the additional health effects from considering the likely impact of changes in expenditure on quality of life during disease. These differences are less marked in line 3 because the effects in those PBCs where an outcome elasticity can be estimated are extrapolated to the other PBCs using proportionate effect on QALY burden and measures of QALY burden in these other.

[18] It would be inappropriate to assign all the change in GMS expenditure to the estimate of cost per QALY based only on the 11 PBCs with outcome elasticities because it would imply that GMS only contributes to these PBCs. Restricting attention to the 11 PBCs with outcome elasticities but allocating part of the change in GMS expenditure to them based on their proportional share of changes in overall expenditure would yield a slightly higher cost per QALY than reported in line 2. It should be noted that including changes in GMS expenditure but not assigning health effects to this PBC is likely to overestimate the threshold because any health effects associated with GMS (or

This estimate reflects changes in undiscounted QALYs associated with changes in expenditure. Although all the health effects of a change in expenditure are restricted to one year (so no discounting is necessary) some of the quality adjusted life year effects of a change in mortality in that year will occur in future years, so in principle should be discounted. However, discounting these life year effects, even at the higher rate of 3.5% recommended by NICE, only increases the cost per QALY threshold to £15,940 (Table C.52).

Table C.52: Summary of QALY threshold, discounted.

	[1] undiscounted	[2] discounted[2]	
	Best estimate		
big 4 PBC's	£3,036	£3,097	[1]
11 PBCs (with mortality)	£5,128	£5,218	[2]
All 23 PBCs [1]	£15,701	£15,940	[3]

[1] in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

[2] Only quality adjusted net YLL were discounted, and thus QALYs associated with gains in QoL during disease were not. The discounting factor has been calculated by applying a 3.5% discount rate to each year of life lost in the PBCs – the estimate of years of life lost used was the implied YLL per death averted in each PBC (in Table C.18 column 4 and reproduced in Tables 28 column 2 and Table 35 column 2). This discounting factor was applied to net YLLs, before applying the outcome elasticity to calculate YLL averted.

As in previous Sections of this Chapter, the upper and lower bounds for the cost per QALY thresholds in column 3 are based on making the necessary assumptions about duration of health effects and how long a death might be averted optimistic (providing the lower bound for the threshold) or conservative (an upper bound for the threshold). The lower bound (lines 4 to 6) is based on assuming that health effects are not restricted to one year but apply to the whole of the remaining disease duration of the population at risk in PBCs during one year. Although this combines optimistic assumptions, it is possible that at least some part of a change in expenditure may prevent disease so will have an impact on populations that are incident to PBCs in the future. Such effects are not captured in any of the estimates presented in this Chapter so all are conservative with respect to this type of health effects of expenditure. The upper bound (lines 7 to 9) is based on the combination of assuming that health effects are restricted to one year for the population currently at risk and that any death averted is only averted for 2 years (see Section C.2.1.5).

C.2.4. Which PBCs matter most?

Which PBCs have the greatest influence on the overall threshold depends, to a large extent, on how a change in overall expenditure is allocated to the different PBCs (see column 1 in table C.53), i.e., those that account for a greater share of the change in expenditure will tend to have the greater influence. [19] However, the overall threshold also depends on the proportionate effect of a change in PBC expenditure on the QALY burden associated with the PBC[20] and the scale of the QALY burden (for the population at risk) associated with the type of diseases that make up each PBC[21]. These determine the cost per QALY associated with each PBC (see column 4 below). The share, attributable to each PBC, of the total health effects of a change in overall expenditure (see column 2) is the combined effect of all of these. The proportionate impact on the overall cost per QALY threshold of a 10% change in PBC health effects in gives an indication of how sensitive the overall threshold is to the estimate of health effects associated with each PBC (see column 3).

PBC 22 see Footnote 48 and 56) will not be reflected in the estimated outcome elasticities of other PBCs unless the effects happen to be correlated with changes in expenditure in those PBCs.

[19] Which are determined by the estimated expenditure elasticities (the proportionate change in PBC expenditure due to a change in overall expenditure) and total PBC expenditure (see Chapter 3 and section B11 in Appendix B).

[20] Which are determined by the outcome elasticities (the proportionate effects on mortality and YLL of a proportionate change in PBC expenditure (see section C.2.3 for details of how these estimates can be applied to measures of QALY burden in all PBCs).

[21] See section C.2.3 for how PBC level effects can be allocated to the contributing ICD codes and how measures of QALY burden for each ICD code can be established

Table C.53: Impact of each PBC on the overall cost per QALY threshold

PBC		% share of change in overall expenditure [1]	% share of total health effects (QALY) [2]	Elasticity of the threshold* [3]	PBC cost per QALY [4]
2	Cancer	2.82	3.47	0.35	£12,772
10	Circulatory	4.90	13.66	1.37	£5,631
11	Respiratory	3.28	45.95	4.60	£1,123
13	Gastro-intestinal	2.43	6.42	0.64	£5,944
1	Infectious diseases	1.23	0.12	0.01	£158,349
4	Endocrine	2.60	11.30	1.13	£3,613
7	Neurological	2.53	2.23	0.22	£17,844
17	Genito-urinary	4.68	0.06	0.01	£1,320,516
16	Trauma & injuries*	1.52	0	0	NA
18+19	Maternity & neonates*	1.18	0.02	<0.01	£813,578
3	Disorders of Blood	1.65	1.59	0.16	£16,257
5	Mental Health	30.07	7.85	0.79	£60,111
6	Learning Disability	4.53	0.29	0.03	£247,001
8	Problems of Vision	3.56	0.55	0.06	£100,871
9	Problems of Hearing	0.94	1.00	0.10	£14,718
12	Dental problems	3.46	1.13	0.11	£48,002
14	Skin	1.55	0.25	0.02	£98,620
15	Musculo skeletal	2.23	3.93	0.39	£8,913
20	Poisoning and AE	0.65	0.12	0.01	£81,782
21	Healthy Individuals	2.70	0.05	0.01	£811,562
22	Social Care Needs	9.96	0	0	NA
23	Other	11.52	0	0	NA

* Calculated using the effect on the threshold of a 10% increase (or decrease) in QALY change of the PBC.

Although the 11PBCs where outcome elasticities could be estimated only account for 27% of the change in overall expenditure they account for 83% of the overall health effects. Within this group some PBCs contribute more than others. For example, PBC11 (Respiratory) accounts for a greater share of total health effects and has a higher elasticity (4.60%) than PBC10 (Circulatory) even though it accounts for a greater part of a change in overall expenditure. The reason is that the cost per QALY associated with changes in expenditure in PBC11 is lower than PBC10 and much lower than the overall threshold (so generates more health effects for the same, or even smaller, change in expenditure). The elasticities in column 3 are instructive, e.g., the elasticity for PBC11 suggests that even if the health effects of a change in expenditure in this PBC were overestimated by 30% the overall threshold would only increase by 13.8% to £17,867. All other PBCs have much less influence in this respect. Nonetheless PBC10 is important compared to others as it does contribute a large share of total health effects and has one of the highest elasticities (1.37%).

The other 12 PBCs, where outcome elasticities could not be estimated, account for the greater part of a change in overall expenditure (73%) but only 17% of the overall health effects, i.e., the cost per QALYs associated with a change in expenditure in these PBCs are, in general, much higher. Of course, we have not directly observed quality of life effects in these PBCs but inferred them from the proportionate effects that we can observe. Insofar as investment and disinvestment opportunities in these PBCs might have been more valuable (offered greater improvement in quality of life) than suggested by the implied PBC thresholds in column 4, the overall QALY effects will tend to be underestimated and the overall cost per QALY threshold will be overestimated.

The overall threshold of £15,701 may be especially conservative (i.e., likely to be overestimated) with respect to health effects in PBC5 (Mental Health Disorders), which accounts for a large proportion of the change in overall expenditure (30%) and contributes most to the overall health effects (7.85%) compared to these other PBCs. The cost per QALY associated with this PBC (£60,111) is based on an extrapolation of estimated proportionate effects to a population based measures of QALY burden in this PBC, rather than observations of the direct impact of changes in expenditure on quality of life in the types of diseases that make up the PBC. Evidence that is available suggests that the investment and disinvestment opportunities in this PBC are likely to have been much more valuable than this implied

cost per QALY (Addendum 3 to this Appendix). A search for evidence about interventions in those ICD codes that contribute most to the PBC (based on prevalence or the contribution to the variance in PBC costs), suggests that pharmacological, psychological and social interventions for depression are all more cost effective (in general much less than £10,000 per QALY) than the overall threshold and significantly more valuable than the implied QALY threshold for this PBC. Based on the contribution that each ICD makes to variance in PBC costs across PCTs, it is schizophrenia that contributes most. Although interventions that may have been invested or disinvested in schizophrenia are, in general, less cost effective (in general less than £24,000 per QALY) than those available for depression, they are still much more valuable than the implied cost per QALY of this PBC in Table C.52.[22]

C.2.5. How uncertain are the estimates?

There are a number of sources of uncertainty which may contribute to an assessment of how uncertain a central or best estimate of the cost per QALY threshold might be. There are three reasons why uncertainty in the estimate of the threshold might be of policy interest: i) the uncertainty in the parameters that determine the threshold might influence the mean or expected value of the threshold if they have a non linear relationship to the threshold or when they have a multi linear relationship but are correlated with each other; ii) the consequences of over or underestimating the threshold differ so the uncertainty may have an influence on the extent to which a policy threshold (a single value that can be compared to the incremental cost effectiveness ratio of a new technology) should differ from the mean or expected value of the central or best estimate; and iii) in conjunction with other methods of analysis it can indicate the potential value of gathering more information to improve these estimates in the future. Such analysis, known as value of information analysis, has firm foundations in statistical decision theory and has been applied to health care decisions. A form of these analyses could be applied in subsequent research, ideally capturing some of the other sources of uncertainty. More recently it has been applied to the decisions faced by NICE when considering whether there is sufficient evidence to support the approval of a new technology.[7] Of course, hypothesis testing and the traditional rules of inference associated with it , such as statistical significance, p-values and confidence intervals, have no relevance when making unavoidable decisions about policy relevant quantities based on information currently available and the best use thereof.[8]

An assessment of parameter uncertainty

Two sets of parameters are critical to the threshold, the expenditure elasticities estimated for each of the 23 PBCs, and the outcome elasticities estimated for 11 of these. These parameters are estimated with uncertainty, indicated by the standard errors on the relevant coefficients in the econometric analysis detailed in Appendix B. Since these statistical models estimate coefficients using normality on the relevant scale, normal distributions can be assigned to each of these estimated coefficients, each with a mean and standard deviation based on the results of the econometric analysis. These distributions represent the uncertainty in the mean estimate of each of the parameters and can be propagated through the various calculations required to estimate and overall cost per QALY threshold (i.e., through the sequence of analysis detailed in Section C.2.2 to C.2.4) using Monte Carlo simulation which randomly samples from the assigned distributions. The use of Monte Carlo simulation in this context is in essence Bayesian, where the standard errors from the frequentist econometric analysis are used to assign normal prior distributions with means equal to the point estimates and a standard deviation equal to the estimated standard errors. This is equivalent to a fully Bayesian analysis with initially uninformative priors which are updated through the analysis of expenditure and mortality data.

[22] Insofar as measures of contribution to variance based on HES data (see Section C.2.3.2) will tend to introduce a bias against those ICD codes where costs are more likely to be recorded in primary care and community services (e.g., more common mental health problems such as depression) then the potential underestimation of health effects is likely to be greater (since these interventions appear more cost effective) and the likelihood that the overall threshold of £15,701 is overestimated will tend to be greater

The results of each random sample from the Monte Carlo simulation represent one possible realisation of the overall threshold, given the uncertainty in estimates of the mean parameter values that determine it. By repeatedly sampling, a distribution of potential values that the overall threshold might take can be revealed. The results of this simulation are illustrated in Figure C.6 showing a histogram of threshold values, and in Figure C.7 showing the cumulative probability density function for a cost per QALY threshold based only on the 11 PBCs with estimated outcome elasticities and for all 23 PBCs. It represents the probability (on the y axis) that the threshold lies below a particular value.

Figure C.6 Distribution of the cost per QALY threshold (all 23 PBCs)

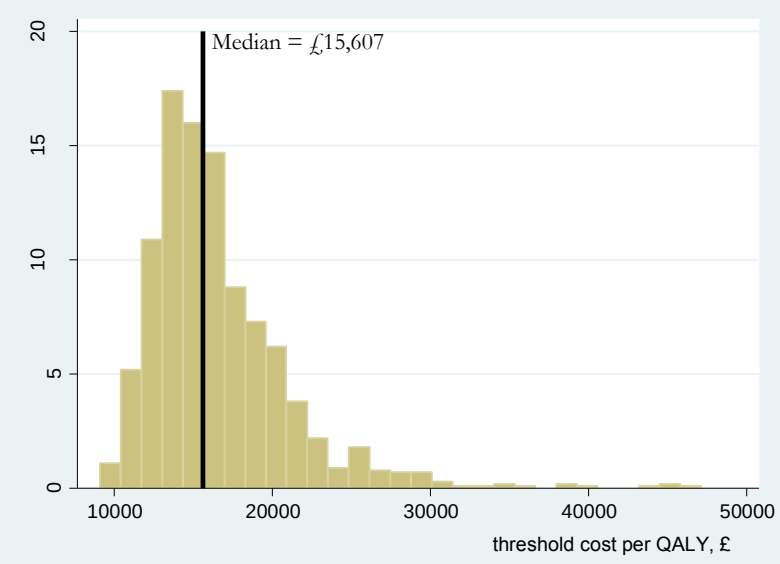
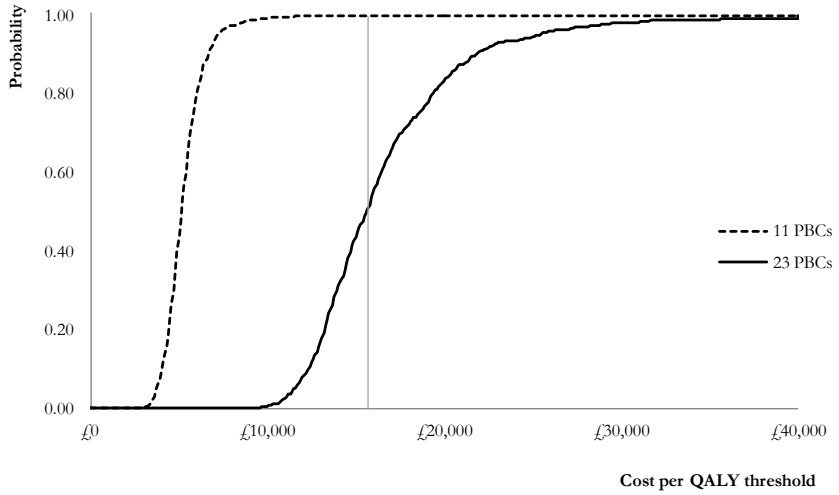


Figure C.7 Cumulative probability density function for the cost per QALY threshold



It has already been noted that restricting attention only to changes in expenditure in those 11PBC where an outcome elasticity can be estimated results in a much lower estimate of the threshold than considering all changes in expenditure across all PBCs. This lower estimate of £5,144 per QALY is much less uncertain but these PBCs only account for 27% of a change in overall expenditure, so it is the higher estimate, for all 23PBCs, that is of most relevance for policy (see Sections C.2.3.3 and C.2.4). The fact that this estimate is more uncertain simply reflects the quality and quantity of data currently available. Since useful analysis should endeavour to faithfully characterise uncertainty in policy relevant quantities, rather than select those quantities or questions for which precise estimates are possible, it is the more

uncertain estimate for all 23 PBCs that should be of primarily interest. The values that are used to generate Figure C.6 are available in column 2 of Table C.54. They indicate that the probability that the overall threshold is less than £20,000 per QALY is 0.84 and the probability that is less than £30,000 is 0.99.

Table C.54: Uncertainty over the QALY threshold.

	11PBCs [1]	All 23 PBCs [2]
Best estimate (deterministic)	£5,128	£15,701
Mean estimate (from the simulations)	£5,114	£15,634
Threshold value at the probability of (from the simulations):		
2.5%	£3,553	£10,963
5.0%	£3,776	£11,591
50.0%	£5,144	£15,607
95.0%	£7,154	£25,159
97.5%	£7,812	£28,212
Probability (from the simulations)of the threshold being smaller than:		
£3,000 per QALY	0.00	0.00
£4,000 per QALY	0.09	0.00
£5,000 per QALY	0.44	0.00
£6,000 per QALY	0.79	0.00
£7,000 per QALY	0.94	0.00
£8,000 per QALY	0.98	0.00
£9,000 per QALY	0.99	0.00
£10,000 per QALY	1.00	0.01
£15,000 per QALY	1.00	0.44
£20,000 per QALY	1.00	0.84
£25,000 per QALY	1.00	0.95
£30,000 per QALY	1.00	0.99
£35,000 per QALY	1.00	0.99
£40,000 per QALY	1.00	1.00

C.3 Re-estimating the cost per QALY threshold using 2008 expenditure data

The same methods of analysis can be applied to the econometric analysis of the 2008/09 expenditure and 2008 to 2010 mortality data (see Section 3.5.3 in Chapter 3 and Section B11 in Appendix B). The differences between the 2006 analysis reported above and the analysis of expenditure in 2008 reported below are the: i) total PBC expenditure ii) estimated expenditure elasticities; iii) estimated outcome elasticities; iv) observed PBC deaths by age and gender; and v) life expectancy by age and gender. The other information about quality of life norms (see Section C.2.2.2), disease related decrements (see Section C.2.2.3) and the information from GBD about incidence (by age and gender) and duration of disease (C.2.3) remain unchanged between 2006 and 2008.

It should be noted that important improvements were made to the classification and collection of PBC expenditure data that took place after the 2006 data were collected. Therefore, the differences in threshold estimates for 2006 and 2008 partly reflect this (see Section 3.5.4 and B11.4 in Appendix B) so should not be over interpreted. The results of the analysis of 2007 and 2008 expenditure are comparable in this respect, providing insights into how the threshold might change over time and with changes in the overall budget. For the purposes of this methodological research the 2008 expenditure and 2008 to 2010 mortality data were the latest to be analysed.

Table C.55: Outcome and spend elasticities (2008)

PBC	PBC description	Total spend 2008/09, (£) [1]	Spend elasticities unadjusted [2]	Spend elasticities adjusted [2]	Outcome elasticities* [3]
2	Cancer	£4,843	0.525	0.525	0.307
10	Circulatory problems	£6,655	0.648	0.648	1.319
11	Respiratory problems	£3,994	0.652	0.652	1.808
13	Gastro-intestinal problems	£3,989	0.456	0.456	1.364
	Big 4	£19,481			
1	Infectious diseases	£1,201	1.545	1.545	0.504
4	Endocrine problems	£2,222	0.484	0.484	1.170
7	Neurological problems	£3,466	0.980	0.980	0.417
17	Genito-urinary problems	£3,779	0.697	0.697	1.615
16	Trauma & injuries*	£3,255	1.344	1.344	-
18+19	Maternity & neonates*	£3,978	0.975	0.975	0.125
	First 11 PBC's	£37,382			
3	Disorders of Blood	£998	1.171	2.291	-
5	Mental Health Disorders	£9,794	1.036	2.027	-
6	Learning Disability	£2,874	0.205	0.401	-
8	Vision	£1,688	0.654	1.279	-
9	Hearing	£417	1.191	2.330	-
12	Dental problems	£3,198	0.513	1.003	-
14	Problems of the Skin	£1,657	0.674	1.318	-
15	Musculo-skeletal system	£4,081	0.505	0.988	-
20	Poisoning and AE	£938	0.562	1.099	-
21	Healthy Individuals	£1,831	1.097	2.146	-
22	Social Care Needs	£1,874	0.911	1.782	-
23	Other	£11,666	0.494	0.494	-
	All (23 PBCs)	£78,398			

* without the negative sign

C.3.1 From mortality to life years

In this section we summarise report the calculation of net YLL, which take account of the fact that some of the observed deaths would have occurred anyway (had the same population not been at risk in the particular PBC) when estimating YLL (unobserved counterfactual deaths). In summary, to obtain net YLL, all observed deaths - both those that occur below and those that occur above LE (Table C.56) – are taken into account. Those deaths occurring below LE generate YLL and those that occur at ages above LE generate life years 'gained' (YLG). By subtracting YLG from YLL to generate net YLL we take account of the fact that not all deaths below LE are excess deaths but some deaths above LE are.

Table C.56. Number of deaths above LE in 2008/9/10, by PBC

PBC		<LE 2008 [1]	>LE 2008 [2]	<LE 2009 [3]	>LE 2009 [4]	<LE 2010 [5]	>LE 2010 [6]	Annual N deaths <LE [7]	Annual N deaths > LE [8]
1	Infectious diseases	3,406	2,586	3,044	2,190	2,667	1,894	3,039	2,223
2	Cancer	94,873	37,029	94,276	37,151	94,309	38,198	94,486	37,459
4	Endocrine	4,033	2,877	3,834	2,826	3,816	2,902	3,894	2,868
7	Neurological	9,638	6,859	9,445	6,939	9,951	7,480	9,678	7,093
10	Circulatory	80,894	76,292	76,048	73,342	74,035	73,719	76,992	74,451
11	Respiratory	32,083	35,180	29,912	33,304	29,691	33,176	30,562	33,887
13	Gastro-intestinal	15,945	8,259	15,361	8,161	15,595	8,372	15,633	8,264
17	Genito-urinary	4,471	6,667	4,378	6,900	4,453	7,166	4,434	6,911
18+19	Maternity & neonates	267	0	281	1	247	0	265	0

The estimates of net YLL calculated considering estimates of the life expectancy for each PBC are detailed in Table C.57.

Table C.57. Net YLL using life expectancy for each PBC (2008)

PBC		LE of Males [1]	LE of Females [2]	Average2006-2008 Deaths <LE [3]	Deaths >LE [4]	YLL [5]	YLG [6]	Net YLL [7]
1	Infectious diseases	79.6	83.6	2,919	2,344	53,926	15,132	38,794
2	Cancer	83.0	84.7	100,487	31,459	1,456,255	134,089	1,322,166
4	Endocrine	81.0	84.7	3,945	2,818	65,800	15,983	49,817
7	Neurological	79.6	83.3	9,112	7,659	137,791	47,722	90,069
10	Circulatory	83.0	86.5	89,434	62,009	1,049,459	278,421	771,038
11	Respiratory	80.3	84.0	29,828	34,621	306,838	229,403	77,434
13	Gastro-intestinal	80.6	84.5	15,612	8,286	271,395	46,141	225,254
17	Genito-urinary	83.5	85.6	5,058	6,287	49,036	32,528	16,508
18+19	Maternity & neonates	78.7	83.1	265	0	19,783	1	19,781

The impact on the cost per life year threshold is summarised in column 2 of Table C.58, and a detailed breakdown in Table C.59.

Table C.58. Summary of cost per life year threshold (2008)

	2006 [1]	**2008** [2]
big 4 PBC's	£8,080	£10,220
11 PBCs (with mortality)	£15,628	£23,360
All 23 PBCs (zero health effects for remaining 12 PBCs)	£57,497	£64,275
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£17,663	£25,214

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

Table C.59: Breakdown of the cost per life year threshold (2008).

PBC	PBC description	Change in spend, £m [1]	Using LE of the PBC population: Net YLL [5]	Change in net YLL [6]	Cost per LY gained, £ [7]
2	Cancer	£25	1 322 166	2 131	£11,931
10	Circulatory problems	£43	771 038	6 590	£6,544
11	Respiratory problems	£26	77 434	913	£28,528
13	Gastro-intestinal problems	£18	225 254	1 401	£12,983
	Big 4				£10,220
1	Infectious diseases	£19	38 794	302	£61,425
4	Endocrine problems	£11	49 817	282	£38,122
7	Neurological problems	£34	90 069	368	£92,282
17	Genito-urinary problems	£26	16 508	186	£141,746
16	Trauma & injuries*	£44	NA	0	NA
18+19	Maternity & neonates*	£39	19 781	24	£1,608,817
	First 11 PBC's				£23,360
3	Disorders of Blood	£23		979	£23,360
5	Mental Health Disorders	£198		8 496	£23,360
6	Learning Disability	£12		493	£23,360
8	Problems of Vision	£22		924	£23,360
9	Problems of Hearing	£10		416	£23,360
12	Dental problems	£32		1 374	£23,360
14	Skin	£22		935	£23,360
15	Musculo skeletal system	£40		1 726	£23,360
20	Poisoning and AE	£10		441	£23,360
21	Healthy Individuals	£39		1 682	£23,360
22	Social Care Needs	£33		1 430	£23,360
23	Other	£58		0	NA
	All (23 PBCs)				£25,214

Note that we have been unable to obtain a satisfactory outcome model for trauma & injuries and have assumed a zero outcome elasticity.
Note that, for expenditure in 2006/7, the neonate category has been merged with maternity to obtain plausible outcome and expenditure models.

The estimates of net YLL imply a number of excess deaths required to generate them in each PBC. The implied excess deaths associated with net YLL are reported in Table C.60.

Table C.60: Excess deaths implied by net YLL (2008).

PBC		Net YLL [1]	YLL per observed death [2]	Excess deaths [3]	Total deaths [4]	% excess deaths [5]
1	Infectious diseases	38,794	13.4	2,934	5,262	56%
2	Cancer	1,322,166	14.1	93,917	131,945	71%
4	Endocrine	49,817	13.7	3,663	6,762	54%
7	Neurological	90,069	13.6	6,642	16,771	40%
10	Circulatory	771,038	10.5	74,217	151,443	49%
11	Respiratory	77,434	9.2	8,432	64,449	13%
13	Gastro-intestinal	225,254	15.2	15,049	23,897	63%
17	Genito-urinary	16,508	8.3	1,978	11,345	17%
18+19	Maternity & neonates	19,781	74.1	265*	265	100%

Excess deaths are calculated for each gender by dividing net YLLs by the YLL per death (column [3] = column [1] / column [2])
* The number of excess deaths estimated in PBC18&19 was initially estimated to be 265, higher than the number of total deaths. This is due to the use of approximations (i.e. in the life expectancy, or in using the net YLL) thus, for consistency, we assumed this to be 100% of the total deaths.

The cost per excess death and the cost per PBC death averted are reported in Table C.61, and a detailed breakdown of changes in spend and excess or total deaths across PBCs is shown in Table C.62. The cost per PBC death averted is, of course; significantly lower than the cost per excess death as excess deaths are only a proportion of total deaths (see Table C.61).

Table C.61. Summary of the cost per death averted threshold (2008)

	2006-2008		2008-2010	
	Cost per excess death averted, £ [1]	Cost per PBC death averted, £ [2]	Cost per excess death averted, £ [3]	Cost per PBC death averted, £ [4]
big 4 PBC's	£91,129	£32,864	£115,234	£46,692
11 PBCs (with mortality)	£177,691	£64,774	£265,784	£105,872
All 23 PBCs (zero health effects for remaining 12 PBCs)	£653,744	£238,310	£731,301	£291,305
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£200,828	£73,208	£286,872	£114,272

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.

Table C.62: Breakdown of the cost per death averted threshold (2008).

			PBC deaths			Excess deaths		
PBC	PBC description	Change in spend, £m [1]	Total PBC deaths [2]	Change in PBC deaths [3]	Cost per PBC death averted, £ [4]	Excess deaths [5]	Change in excess deaths [6]	Cost per excess death averted, £ [7]
2	Cancer	£25	131 945	212.66	£119,559	93 917	151.37	£167,969
10	Circulatory problems	£43	151 443	1294.40	£33,316	74 217	634.34	£67,983
11	Respiratory problems	£26	64 449	759.74	£34,276	8 432	99.40	£261,992
13	Gastro-intestinal problems	£18	23 897	148.64	£122,379	15 049	93.60	£194,332
	Big 4				£46,692	0		£115,234
1	Infectious diseases	£19	5 262	40.97	£452,858	2 934	22.84	£812,249
4	Endocrine problems	£11	6 762	38.29	£280,856	3 663	20.74	£518,533
7	Neurological problems	£34	16 771	68.54	£495,603	6 642	27.14	£1,251,391
17	Genito-urinary problems	£26	11 345	127.71	£206,253	1 978	22.27	£1,182,744
16	Trauma & injuries*	£44	NA	0	NA	NA	0	NA
18+19	Maternity & neonates*	£39	265	0.32	£120,090,566	265	0.32	£120,090,566
	First 11 PBC's				£105,872			£265,784
3	Disorders of Blood	£23		215.92	£105,872		86.01	£265,784
5	Mental Health Disorders	£198		1874.69	£105,872		746.76	£265,784
6	Learning Disability	£12		108.86	£105,872		43.36	£265,784
8	Problems of Vision	£22		203.97	£105,872		81.25	£265,784
9	Problems of Hearing	£10		91.76	£105,872		36.55	£265,784
12	Dental problems	£32		303.11	£105,872		120.74	£265,784
14	Skin	£22		206.34	£105,872		82.19	£265,784
15	Musculo skeletal system	£40		380.77	£105,872		151.68	£265,784
20	Poisoning and AE	£10		97.40	£105,872		38.80	£265,784
21	Healthy Individuals	£39		371.11	£105,872		147.83	£265,784
22	Social Care Needs	£33		315.43	£105,872		125.65	£265,784
23	Other	£58		0	NA		0	NA
	All (23 PBCs)				£114,272			£286,872

The number of life years gained associated with each excess death averted are reported for each PBC in Table C.63 (column 1) and range from 74.6 years for PBC18&19 to 8.3 years for PBC17. On average, across all 11 PBCs each excess death averted is associated with 11.4 life years gained. The life years associated with each observed death are reported for each PBC in (column 2) and range from 74.6 years in PBC 18 & 19 to 1.2 for PBC17. On average across all 11 PBCs each PBC death averted is associated with 4.5 life years gained.

Table C.63: Implied YLL per death averted for each PBC (2008)

PBC	PBC description	Implied YLL per excess death averted [1]	Implied YLL per PBC death averted [2]
2	Cancer	14.1	10.0
10	Circulatory problems	10.4	5.1
11	Respiratory problems	9.2	1.2
13	Gastro-intestinal problems	15.0	9.4
	Big 4	11.3	4.6
1	Infectious diseases	13.2	7.4
4	Endocrine problems	13.6	7.4
7	Neurological problems	13.6	5.4
17	Genito-urinary problems	8.3	1.5
16	Trauma & injuries	NA	NA
18+19	Maternity & neonates	74.6	74.6
	First 11 PBC's	11.4	4.5

Summary of cost per life year estimates

The cost per life year threshold in lines 1 to 4 in Table C.64 are regarded as the central or best estimates given the evidence available and the credibility of alternative assumption that could be made. As explained in Section C.1, these are based on the conservative assumption that any health effects of changes in expenditure are restricted to one year, which, to some extent, may be offset by the more optimistic assumption any death averted returns the individual to the mortality risk face by the general population, matched for age and gender. See Section C.2.5 for guidance in the interpretation of the upper and lower bound estimates.

Table C.64: Summary of the cost per life year threshold with upper and lower bounds (2008)

	2006-2008 [1]	**2008-2010** [2]	
	Best estimate		
Effect of expenditure on mortality:	*1 year*	*1 year*	
YLL per PBC death averted:	*~ 4.1 YLL* **	*~ 4.5 YLL* **	
big 4 PBC's	£8,080	£10,220	[1]
11 PBCs (with mortality)	£15,628	£23,360	[2]
All 23 PBCs (zero health effects for remaining 12 PBCs)	£57,497	£64,275	[3]
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£17,663	£25,214	[4]
	Lower bound		
Effect of expenditure on mortality:	*Remainder of disease*	*Remainder of disease*	
YLL per PBC death averted:	*~ 4.1 YLL* **	*~ 4.5 YLL* **	
big 4 PBC's	£3,846	£5,083	[5]
11 PBCs (with mortality)	£6,106	£8,579	[6]
All 23 PBCs (zero health effects for remaining 12 PBCs)	£22,463	£23,605	[7]
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£6,901	£9,260	[8]
	Upper bound		
Effect of expenditure on mortality:	*1 year*	*1 year*	
YLL per PBC death averted:	*2 YLL*	*2 YLL*	
big 4 PBC's	£16,432	£23,346	[9]
11 PBCs (with mortality)	£32,387	£52,936	[10]
All 23 PBCs (zero health effects for remaining 12 PBCs)	£119,155	£145,653	[11]
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£36,604	£57,136	[12]

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal.
** see Table C.63

C.3.2 Adjusting life years for quality of life

The central or best estimates of the cost per life year threshold, which were presented in Table C.64 (lines 2 and 4) take no account of the health related quality of life in which years of life, expected to be gained or lost through changes in expenditure, are likely to be lived. In this Section we examine the ways in which the life years reported in Section C.3.2 can be adjusted for quality, taking account of information that is available about: i) how quality of life differs by age and gender, and ii) how the quality of life years associated with mortality changes might be affected by the types of diseases that make up each PBC.

Quality of life based on the general population

Quality of life norms (in Figure C.4) can be applied to the YLL associated with all observed deaths in each PBC, taking account of gender and age at death. The results are reported in column 4 to 6 of Table C.65.

Table C.65: Net YLL adjusted for quality of life 'norms' (2008)

PBC		Unadjusted life years			Quality adjusted life years		
		YLL	YLG	YLL	YLG	YLL	YLG
		[1]	[2]	[3]	[4]	[5]	[6]
1	Infectious diseases	53,926	15,132	38,794	43,703	10,187	33,516
2	Cancer	1,456,255	134,089	1,322,166	1,129,191	89,231	1,039,960
4	Endocrine	65,800	15,983	49,817	52,465	10,598	41,867
7	Neurological	137,791	47,722	90,069	110,532	32,262	78,270
10	Circulatory	1,049,459	278,421	771,038	807,893	183,796	624,097
11	Respiratory	306,838	229,403	77,434	237,981	154,300	83,680
13	Gastro-intestinal	271,395	46,141	225,254	214,756	30,811	183,945
17	Genito-urinary	49,036	32,528	16,508	37,178	21,190	15,989
18+19	Maternity & neonates	19,783	1	19,781	17,176	1	17,175

The implications of the quality adjustment to a cost per QALY threshold that only accounts for the direct health effects of mortality are summarised in Table C.66, and detailed in Table C.67.

Table C.66: Summary of cost per QALY threshold based on population norms and mortality effects (2008)

	2006-2008		**2008-2010**	
	Cost per life year threshold	Cost per QALY threshold Population norms	Cost per life year threshold	Cost per QALY threshold Population norms
	[1]	[2]	[3]	[4]
big 4 PBCs	£8,080	£9,631	£10,220	£12,338
11 PBCs	£15,628	£18,622	£23,360	£28,045
All 23 PBCs	£17,663	£21,047	£25,214	£30,270

Table C.67: A breakdown of the cost per QALY threshold based on population norms (2008)

PBC	PBC description	Change in spend, £m [1]	YLL using LE of PBC Change in QALY [2]	YLL using LE of PBC Cost per QALY gained, £ [3]
2	Cancer	£25	1 676	£15,169
10	Circulatory problems	£43	5 334	£8,084
11	Respiratory problems	£26	986	£26,399
13	Gastro-intestinal problems	£18	1 144	£15,899
	Big 4			£12,338
1	Infectious diseases	£19	261	£71,098
4	Endocrine problems	£11	237	£45,361
7	Neurological problems	£34	320	£106,193
17	Genito-urinary problems	£26	180	£146,347
16	Trauma & injuries*	£44	0	NA
18+19	Maternity & neonates*	£39	21	£1,852,926
	First 11 PBC's			£28,045
3	Disorders of Blood	£23	815	£28,045
5	Mental Health Disorders	£198	7 077	£28,045
6	Learning Disability	£12	411	£28,045
8	Problems of Vision	£22	770	£28,045
9	Problems of Hearing	£10	346	£28,045
12	Dental problems	£32	1 144	£28,045
14	Skin	£22	779	£28,045
15	Musculo skeletal system	£40	1 437	£28,045
20	Poisoning and adverse effects	£10	368	£28,045
21	Healthy Individuals	£39	1 401	£28,045
22	Social Care Needs	£33	1 191	£28,045
23	Other	£58	0	NA
	All (23 PBCs)			£30,270

Table C.68 depicts the judgements over life years, quality of life weights and total QALYs implicit in calculations of the threshold cost per QALY in Table C.64.

Table C.68: Implied YLL per excess death averted and implied QoL score per YLL gained, for each PBC (2008)

PBC	PBC description	Implied YLL per excess death averted [1]	Implied YLL per PBC death averted **[2]**	Implied QALYs gained per excess death averted [3]	Implied QALYs gained per PBC death averted [4]
2	Cancer	14.08	10.02	11.07	7.88
10	Circulatory	10.39	5.09	8.41	4.12
11	Respiratory	9.18	1.20	9.92	1.30
13	Gastro-intestinal	14.97	9.43	12.22	7.70
	Big 4	11.28	4.57	9.34	3.78
1	Infectious diseases	13.22	7.37	11.42	6.37
4	Endocrine	13.60	7.37	11.43	6.19
7	Neurological	13.56	5.37	11.78	4.67
17	Genito-urinary	8.34	1.46	8.08	1.41
16	Trauma & injuries*	NA	NA	NA	NA
18+19	Maternity & neonates*	74.65	74.65	64.81	64.81
	First 11 PBC's	11.38	4.53	9.48	3.78

Adjusting age related quality of life for disease decrements

By using age related quality of life disease decrements (exemplified in Figure C.5) YLL can be adjusted for quality of life of disease. The results are reported in column 4 to 6 of Table C.69.

Table C.69: Net YLL adjusted for disease and age related quality of life (2008)

PBC		Unadjusted life years			Quality adjusted life years		
		YLL [1]	YLG [2]	YLL [3]	YLG [4]	YLL [5]	YLG [6]
1	Infectious diseases	53,926	15,132	38,794	34,108	7,524	26,584
2	Cancer	1,456,255	134,089	1,322,166	943,650	72,197	871,452
4	Endocrine	65,800	15,983	49,817	43,063	8,334	34,729
7	Neurological	137,791	47,722	90,069	69,520	18,084	51,436
10	Circulatory	1,049,459	278,421	771,038	625,150	135,622	489,527
11	Respiratory	306,838	229,403	77,434	173,953	106,200	67,754
13	Gastro-intestinal	271,395	46,141	225,254	162,441	22,060	140,380
17	Genito-urinary	49,036	32,528	16,508	30,770	16,949	13,820
18+19	Maternity & neonates	19,783	1	19,781	16,100	1	16,099

The implications of the quality adjustment to a cost per QALY threshold that only accounts for the direct health effects of mortality are summarised in Table C.70, and detailed in Table C.71.

Table C.70: Summary of cost per QALY threshold based on disease and age related quality of life and mortality effects (2008)

	2006		**2008**	
	Cost per life year threshold [1]	Cost per QALY threshold Disease related disutility [2]	Cost per life year threshold [3]	Cost per QALY threshold Disease related disutility [4]
big 4 PBC's	£8,080	£12,109	£10,220	£15,534
11 PBCs (with mortality)	£15,628	£23,395	£23,360	£35,397
All 23 PBCs (zero health effects for remaining 12 PBCs)	£57,497	£86,072	£64,275	£97,395
All 23 PBCs (non-zero health effects for remaining 12 PBCs, except GMS)*	£17,663	£26,441	£25,214	£38,206

* in PBCs without a mortality signal, health effects were estimated by valuing changes in expenditure at the same rate as observed in PBCs for which there was a mortality signal except GMS.

Table C.71: A breakdown of the cost per QALY threshold based on disease and age related quality of life and mortality effects (2008)

PBC	PBC description	Change in spend, £m [1]	YLL using LE of PBC Change in QALY [2]	YLL using LE of PBC Cost per QALY gained, £ [3]
2	Cancer	£25	1 405	£18,102
10	Circulatory problems	£43	4 184	£10,307
11	Respiratory problems	£26	799	£32,604
13	Gastro-intestinal problems	£18	873	£20,833
	Big 4			£15,534
1	Infectious diseases	£19	207	£89,638
4	Endocrine problems	£11	197	£54,685
7	Neurological problems	£34	210	£161,594
17	Genito-urinary problems	£26	156	£169,315
16	Trauma & injuries*	£44	0	NA
18+19	Maternity & neonates*	£39	20	£1,976,769
	First 11 PBC's			£35,397
3	Disorders of Blood	£23	646	£35,397
5	Mental Health Disorders	£198	5 607	£35,397
6	Learning Disability	£12	326	£35,397
8	Problems of Vision	£22	610	£35,397
9	Problems of Hearing	£10	274	£35,397
12	Dental problems	£32	907	£35,397
14	Skin	£22	617	£35,397
15	Musculo skeletal system	£40	1 139	£35,397
20	Poisoning and adverse effects	£10	291	£35,397
21	Healthy Individuals	£39	1 110	£35,397
22	Social Care Needs	£33	943	£35,397
23	Other	£58	0	NA
	All (23 PBCs)			£38,206

Table C.72 depicts the judgements over life years, quality of life weights and total QALYs implicit in calculations of the threshold cost per QALY in Table C.70.

Table C.72: Implied YLL per excess death averted and implied QoL score per YLL gained, for each PBC (2008)

PBC	PBC description	Implied YLL per excess death averted [1]	Implied YLL per PBC death averted [2]	Implied QALYs gained per excess death averted [3]	Implied QALYs gained per PBC death averted [4]
2	Cancer	14.08	10.02	9.28	6.60
10	Circulatory	10.39	5.09	6.60	3.23
11	Respiratory	9.18	1.20	8.04	1.05
13	Gastro-intestinal	14.97	9.43	9.33	5.87
	Big 4	11.28	1.80	7.42	3.01
1	Infectious diseases	13.22	7.37	9.06	5.05
4	Endocrine	13.60	7.37	9.48	5.14
7	Neurological	13.56	5.37	7.74	3.07
17	Genito-urinary	8.34	1.46	6.99	1.22
16	Trauma & injuries*	NA	NA	NA	NA
18+19	Maternity & neonates*	74.65	74.65	60.75	60.75
	First 11 PBC's	11.38	4.53	6.77	2.99

Summary of the cost per QALY threshold based only on mortality effects

The analysis to this point is summarised in Table C.73. The three estimates of a cost per QALY threshold are based on assuming that each life year gained is either: lived in full health (see column 1), lived in a quality of life that reflects age and gender norms of the general population (column 2); or lived in a quality of life that reflects the original disease state (column 3).

Table C.73: Summary of QALY threshold estimates based only on mortality effects (2008)

	[1] *(QoL score =1)*	[2] *(QoL norm)*	[3] *(QoL diseased)*	
	Best estimate			
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
*YLL per death averted *:*	*~ 4.5 YLL*	*~ 4.5 YLL*	*~ 4.5 YLL*	
*QALYs per death averted *:*	*~ 4.5 QALY*	*~ 3.8 QALY*	*~ 3.0 QALY*	
big 4 PBC's	£10,220	£12,338	£15,534	[1]
11 PBCs	£23,360	£28,045	£35,397	[2]
All 23 PBCs	£25,214	£30,270	£38,206	[3]
	Lower bound			
Effect of expenditure on mortality:	*Remainder of disease*	*Remainder of disease*	*Remainder of disease*	
*YLL per death averted *:*	*~ 4.5 YLL*	*~ 4.5 YLL*	*~ 4.5 YLL*	
*QALYs per death averted *:*	*~ 4.5 QALY*	*~ 3.8 QALY*	*~ 3.0 QALY*	
big 4 PBC's	£5,083	£5,811	£7,305	[4]
11 PBCs	£8,579	£9,861	£12,720	[5]
All 23 PBCs	£9,260	£10,644	£13,729	[6]
	Upper bound			
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
*YLL per death averted *:*	*2 YLL*	*2 YLL*	*2 YLL*	
*QALYs per death averted *:*	*~ 2 QALY*	*~ 1.8 QALY*	*~ 1.4 QALY*	
big 4 PBC's	£23,346	£26,138	£32,797	[7]
11 PBCs	£52,936	£59,151	£74,183	[8]
All 23 PBCs	£57,136	£63,844	£80,069	[9]

* see Table C.72

C.3.3. Including quality of life effects during disease

In this section we explore how estimates of effects of expenditure that can be observed (i.e., on mortality) can be used to infer the likely effects on what cannot be directly observed (quality of life), rather than making extreme assumptions that are not credible (e.g., assuming that changes in expenditure will have no effects on quality of life outcomes). In Section C.2.3.2, we described the use of ratios of QALYs lost

to life years lost due to disease and explored how the use of the QALY burden of disease is preferable to inform estimates of the threshold. We here present only the results for the QALY burden approach.

In Table C.74, deaths and YLL from ONS (2008 to 2010 mortality data) compare to those from GBD. The factors used to adjust GBD information are reported in columns 4 and 7.

Table C.74: Comparing deaths and YLL from ONS and GBD. (2008)

			deaths			YLL		
		Excess deaths ONS	All deaths ONS	All deaths GBD*	adjustment factor (deaths)	Net estimates ONS	Total YLL GBD*	adjustment factor (YLL)
		[1]	[2]	[3]	[4]	[5]	[6]	[7]
1	Infectious diseases	2,934	5,262	1,408	3.737	38,794	25,142	1.543
2	Cancer	93,917	131,946	140,124	0.942	1,322,166	1,932,637	0.684
4	Endocrine	3,663	6,762	7,509	0.901	49,817	95,401	0.522
7	Neurological	6,642	16,771	12,854	1.305	90,069	164,796	0.547
10	Circulatory	74,217	151,443	178,454	0.849	771,038	1,750,608	0.440
11	Respiratory	8,432	64,449	67,441	0.956	77,434	594,529	0.130
13	Gastro-intestinal	15,049	23,897	28,329	0.844	225,254	396,829	0.568
17	Genito-urinary	1,978	11,345	8,606	1.318	16,508	77,338	0.213
18+19	Maternity & neonates	265	265	2,211	0.120	19,781	149,868	0.132
	Total	*207,097*	*412,140*	*446,936*	*0.92*	*2,610,861*	*5,187,148*	*0.50*

The threshold cost per QALY based on burden associated with one year of disease derived from GBD are summarised in Table C.75 and detailed in Table C.76.

Table C.75: Summary of the cost per QALY threshold (2008)

	2006	2008
	[1]	[2]
big 4 PBC's	£3,036	£4,872
11 PBCs (with mortality)	£5,128	£8,308
All 23 PBCs	£15,701	£18,317

* Preferred analysis

Table C.76: Breakdown of the cost per QALY threshold (2008)

			QALY burden (HoDAR and MEPs)	
PBC	PBC description	Change in spend, £m	Change in QALY	Cost per QALY gained, £
		[1]	[4]	[5]
2	Cancer	£25	1 496	£16,997
10	Circulatory problems	£43	6 127	£7,038
11	Respiratory problems	£26	13 032	£1,998
13	Gastro-intestinal problems	£18	2 494	£7,293
	Big 4			£4,872
1	Infectious diseases	£19	891	£20,829
4	Endocrine problems	£11	3 442	£3,124
7	Neurological problems	£34	6 198	£5,480
17	Genito-urinary problems	£26	601	£43,813
16	Trauma & injuries*	£44	0	NA
18+19	Maternity & neonates*	£39	13	£2,969,208
	First 11 PBC's			£8,308
3	Disorders of Blood	£23	808	£28,305
5	Mental Health Disorders	£198	3 983	£49,835
6	Learning Disability	£12	146	£78,854
8	Problems of Vision	£22	281	£76,850
9	Problems of Hearing	£10	509	£19,070
12	Dental problems	£32	574	£55,916
14	Skin	£22	125	£174,775
15	Musculo skeletal system	£40	1 990	£20,254
20	Poisoning and adverse effects	£10	63	£163,766
21	Healthy Individuals	£39	26	£1,483,012
22	Social Care Needs	£33	0	NA
23	Other	£58	0	NA
	All (23 PBCs)			£18,317

Summary of the cost per QALY threshold

The results of the three sequential steps of analysis are summarised in Table C.77, for this year of analysis. They include: i) the cost per life year (column 1) based on the methods of analysis outlined in Section C.2.1; ii) the cost per life year adjusted for quality of life (column 2) based on the methods of analysis outlined in Section C.2.2; and iii) the cost per QALY (column 3) based on the methods of analysis outlined in Section C.2.3. These estimates, in column 3, take account of the likely effects of changes in expenditure on quality of life during disease as well as the effects associated with mortality and life years; making best use of available information, while the assumptions required appear more reasonable than the other alternatives available. For this reason these estimates remain our central or best estimates for all the waves of expenditure and mortality data.

Table C.77: Summary of cost per QALY threshold estimates (2008)

	[1]	[2]	[3]	
QoL associated with life extension:	*1*	*Norm*		
QoL during disease:	*0*	*0*	*Based on burden*	
			Best estimate	
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per death averted:	*~ 4.5 YLL*	*~ 4.5 YLL*	*~ 4.5 YLL*	
QALYs per death averted:	*~ 4.5 QALY*	*~ 3.8 QALY*	*~ 12.7 QALY*	
big 4 PBC's	£10,220	£12,338	£4,872	[1]
11 PBCs (with mortality)	£23,360	£28,045	£8,308	[2]
All 23 PBCs	£25,214	£30,270	**£18,317**	[3]
			Lower bound	
Effect of expenditure on mortality:	*Remainder of disease duration*	*Remainder of disease duration*	*Remainder of disease duration*	
YLL per death averted:	*~ 4.5 YLL*	*~ 4.5 YLL*	*~ 4.5 YLL*	
QALYs per death averted:	*~ 4.5 QALY*	*~ 3.8 QALY*	*~ 12.7 QALY*	
big 4 PBC's	£5,083	£5,811	£1,194	[4]
11 PBCs (with mortality)	£8,579	£9,861	£1,175	[5]
All 23 PBCs	£9,260	£10,644	**£2,832**	[6]
			Upper bound	
Effect of expenditure on mortality:	*1 year*	*1 year*	*1 year*	
YLL per death averted:	*2 YLL*	*2 YLL*	*2 YLL*	
QALYs per death averted:	*~ 2 QALY*	*~ 1.4 QALY*	*~ 5.6 QALY*	
big 4 PBC's	£23,346	£26,138	£11,040	[7]
11 PBCs (with mortality)	£52,936	£59,151	£18,827	[8]
All 23 PBCs	£57,136	£63,844	**£41,507**	[9]

The estimate of £8,308 per QALY (line 2) is restricted to the effects of changes in expenditure in the 11PBCs where outcome elasticities can be estimated. However, these PBCs only account for a proportion of a change in overall expenditure (approximately 35%, see Table C.80 below). As was explained in Section C.2.3 the QALY threshold of £18,317 (column 3, line 3) uses the estimated proportionate effects of expenditure on the QALY burden of disease in the 11PBCs as a surrogate for proportionate effects in the others, (i.e., assuming that the effects that can be observed will be similar to those that cannot) and represents our central or best estimate. As in previous sections, no health effects are assigned to PBC23 or 22 (General Medical Services and Social Care) on the basis that any health effects of this expenditure would be recorded in the other PBCs. Although this estimate of £18,317 reflects changes in undiscounted QALYs associated with changes in expenditure, discounting the quality adjusted life year effects only increases the cost per QALY threshold to £18,613 (Table C.78). The effects of discounting are modest because: i) the health effects of a change in expenditure are restricted to one year (where no discounting is necessary); ii) most of the total QALY effect occurs in that year; iii) it is only some of the life year effects (adjusted for quality) of a change in mortality in that year that occur in future years that need to be discounted; and iv) these need to be discounted only over 4.5 years on average.

Table C.78: Summary of QALY threshold, discounted (2008).

	2008-2010 discounted	
	Best estimate	
big 4 PBC's	£4,998	[1]
11 PBCs	£8,467	[2]
All 23 PBCs	£18,613	[3]

[2] Only quality adjusted net YLL were discounted, and thus QALYs associated with gains in QoL during disease were not. The discounting factor has been calculated by applying a 3.5% discount rate to each year of life lost in the PBCs – the estimate of years of life lost used was the implied YLL per death averted in each PBC (in Table C.18 column 4 and reproduced in Tables 28 column 2 and Table 35 column 2). This discounting factor was applied to net YLLs, before applying the outcome elasticity to calculate YLL averted.

The upper and lower bounds for the cost per QALY thresholds in column 3 are based on making the necessary assumptions about duration of health effects and how long a death might be averted optimistic (providing the lower bound for the threshold) or conservative (an upper bound for the threshold). The lower bound (lines 4 to 6) is based on assuming that health effects are not restricted to one year but apply to the whole of the remaining disease duration of the population at risk in PBCs during one year. Although this combines optimistic assumptions, it is possible that at least some part of a change in expenditure may prevent disease so will have an impact on populations that are incident to PBCs in the future. Such effects are not captured in any of the estimates presented in this report so all are conservative in this respect. The upper bound (lines 7 to 9) is based on the combination of assuming that health effects are restricted to one year for the population currently at risk and that any death averted is only averted for 2 years (see Section C.2.1.5).

The estimated QALY effects associated with each PBC can be decomposed into that part due to life year effects adjusted for quality and that part associated with effects on quality during disease (Table C.79). Those PBCs for which mortality is the major concern have a much greater share of total QALY effects associated with avoidance of premature death (e.g., PBC2 and PBC10) compared to those where quality of life is the major concern (e.g., PBC 7).

Table C.79: Decomposing estimated QALY effects by PBC (2008)

PBC		QALY change (total) [1]	QALY change (death) [2]	% QALY gained for premature death [3]	% QALY gained for disability while alive [4]
2	Cancer	1,453	1,393	96%	4%
10	Circulatory	5,125	4,054	79%	21%
11	Respiratory	10,947	758	7%	93%
13	Gastro-intestinal	2,087	1,024	49%	51%
1	Infectious diseases	14	9	67%	33%
4	Endocrine	2,921	269	9%	91%
7	Neurological	441	43	10%	90%
17	Genito-urinary	13	5	40%	60%
16	Trauma & injuries*	0	0	NA	NA
18+19	Maternity & neonates*	22	7	30%	70%
3	Disorders of Blood	689	35	5%	95%
5	Mental Health	3,397	296	9%	91%
6	Learning Disability	125	25	20%	80%
8	Problems of Vision	240	9	4%	96%
9	Problems of Hearing	434	3	1%	99%
12	Dental problems	489	0	0%	100%
14	Skin	107	39	37%	63%
15	Musculo skeletal	1,697	84	5%	95%
20	Poisoning and AE	54	9	16%	84%
21	Healthy Individuals	23	4	16%	84%
22	Social Care Needs	0	0	NA	NA
23	Other	0	0	NA	NA

C.3.4. Which PBCs matter most?

Table C.80: Impact of each PBC on the overall cost per QALY threshold (2008)

PBC		% share of change in overall expenditure [1]	% share of total health effects (QALY) [2]	Elasticity of the threshold* [5]	PBC cost per QALY
2	Cancer	3.24	3.50	0.35%	£16,997
10	Circulatory	5.50	14.32	1.43%	£7,038
11	Respiratory	3.32	30.45	3.05%	£1,998
13	Gastro-intestinal	2.32	5.83	0.58%	£7,293
1	Infectious diseases	2.37	2.08	0.21%	£20,829
4	Endocrine	1.37	8.04	0.80%	£3,124
7	Neurological	4.33	14.48	1.45%	£5,480
17	Genito-urinary	3.36	1.40	0.14%	£43,813
16	Trauma & injuries*	5.58	0	0.00%	NA
18+19	Maternity & neonates*	4.95	0.03	0.00%	£2,969,208
3	Disorders of Blood	2.92	1.89	0.19%	£28,305
5	Mental Health	25.32	9.31	0.93%	£49,835
6	Learning Disability	1.47	0.34	0.03%	£78,854
8	Problems of Vision	2.75	0.66	0.07%	£76,850
9	Problems of Hearing	1.24	1.19	0.12%	£19,070
12	Dental problems	4.09	1.34	0.13%	£55,916
14	Skin	2.79	0.29	0.03%	£174,775
15	Musculo skeletal	5.14	4.65	0.47%	£20,254
20	Poisoning and AE	1.32	0.15	0.01%	£163,766
21	Healthy Individuals	5.01	0.06	0.01%	£1,483,012
22	Social Care Needs	4.26	0	0.00%	NA
23	Other	7.35	0	0.00%	NA

* Calculated using the effect on the threshold of a 10% increase (or decrease) in QALY change of the PBC.

C.3.5. How uncertain are the estimates?

In Section 2.2.5, the impact of uncertainty over the spend and outcome elasticities on estimates of the cost per QALY threshold has been illustrated and interpreted in detail. We here repeat this analysis using expenditure data from 2008/09 and mortality data from 2008 to 2010. Figure C.8 shows the histogram of threshold values from the Monte Carlo simulation (where each random sample from the simulation represents one possible realisation of the overall threshold), and Figure C.9 shows the cumulative probability density function for a cost per QALY threshold based on the 11 PBC with estimated outcome elasticities and for all 23 PBCs.

Figure C.8: Histogram of simulation of undiscounted threshold (all 23 PBCs) (2008)

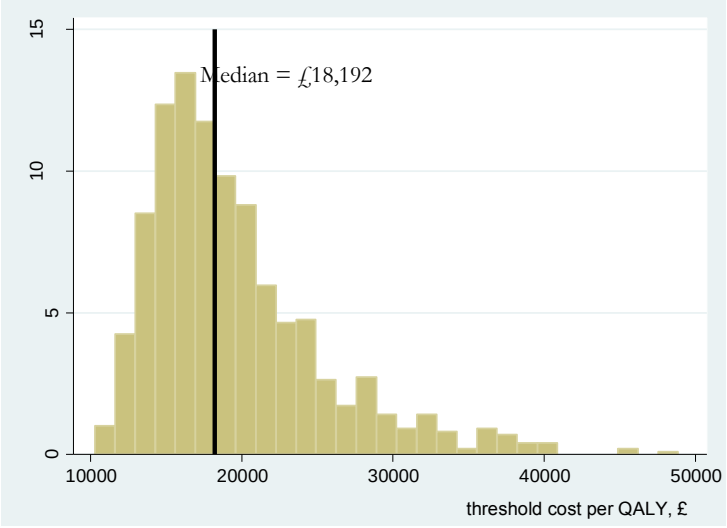


Figure C.9: Cumulative probability density function for the cost per QALY threshold (2008)

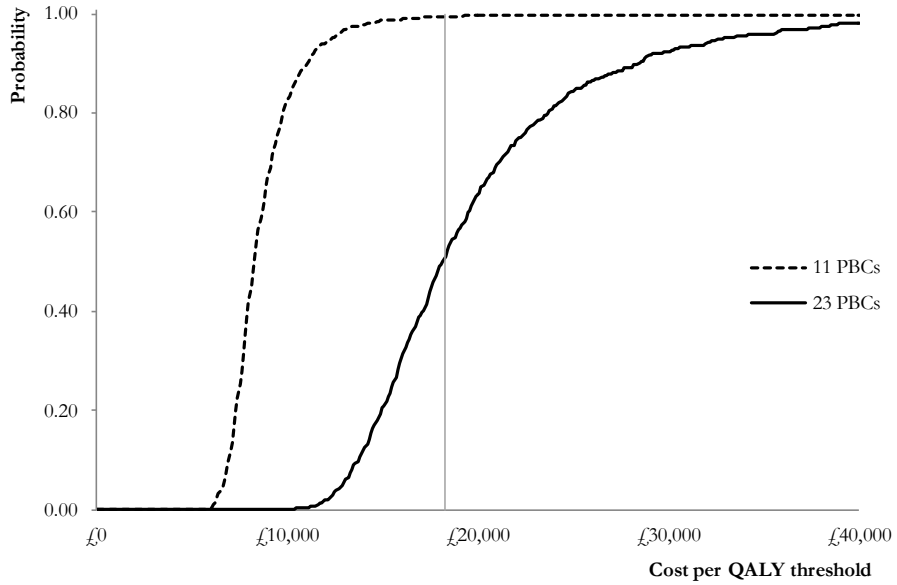


Table C.81: Uncertainty over the QALY threshold (2008).

	11PBCs [1]	All 23 PBCs [2]
Best estimate (deterministic)	£8,308	£18,317
Mean estimate (from the simulations)	£8,330	£18,310
Threshold value at the probability of (from the simulations):		
2.5%	£6,329	£12,232
5.0%	£6,670	£12,907
50.0%	£8,266	£18,192
95.0%	£12,272	£32,845
97.5%	£13,602	£38,099
Probability (from the simulations)of the threshold being smaller than:		
£5,000 per QALY	0%	0%
£6,000 per QALY	1%	0%
£7,000 per QALY	12%	0%
£8,000 per QALY	42%	0%
£9,000 per QALY	67%	0%
£10,000 per QALY	82%	0%
£15,000 per QALY	99%	21%
£20,000 per QALY	100%	64%
£25,000 per QALY	100%	85%
£30,000 per QALY	100%	92%
£35,000 per QALY	100%	96%
£40,000 per QALY	100%	98%
£45,000 per QALY	100%	98%
£50,000 per QALY	100%	99%

C.4 Re-estimating the cost per QALY threshold using 2007 expenditure data

The same methods of analysis were applied to the econometric analysis of the 2007/08 expenditure and 2007 to 2009 mortality data (see Section B10 in Appendix B). Given the detailed reporting of the methods and interpretation of the analyses for other expenditure years (see Sections C.2 and C.3), we will here only present the necessary Tables of results.

Table C.82: Outcome and spend elasticties (2007)

PBC	PBC description	Total spend 2007/08, (£) [1]	Spend elasticities unadjusted [2]	Spend elasticities adjusted [2]	Outcome elasticities* [3]
2	Cancer	£4,573	0.890	0.890	0.365
10	Circulatory problems	£6,325	0.293	0.293	1.277
11	Respiratory problems	£3,431	0.536	0.536	2.205
13	Gastro-intestinal problems	£3,805	0.622	0.622	1.328
	Big 4	£18,134			
1	Infectious diseases	£1,119	1.436	1.436	0.548
4	Endocrine problems	£1,997	0.264	0.264	0.566
7	Neurological problems	£3,165	1.035	1.035	0.339
17	Genito-urinary problems	£3,439	1.004	1.004	1.855
16	Trauma & injuries*	£2,918	1.686	1.686	0.369 +
18+19	Maternity & neonates*	£3,662	0.514	0.514	0.110
	First 11 PBC's	£34,434			
3	Disorders of Blood	£986	1.83	2.879	-
5	Mental Health Disorders	£9,171	1.145	1.801	-
6	Learning Disability	£2,748	0.44	0.692	-
8	Vision	£1,556	1.17	1.841	-
9	Hearing	£409	1.029	1.619	-
12	Dental problems	£3,014	0.424	0.667	-
14	Problems of the Skin	£1,542	0.428	0.673	-
15	Musculo-skeletal system	£3,848	0.806	1.268	-
20	Poisoning and AE	£803	0.668	1.051	-
21	Healthy Individuals	£1,594	0.986	1.551	-
22	Social Care Needs	£1,789	1.852	2.913	-
23	Other	£11,763	0.563	0.563	-
	All (23 PBCs)	£73,656			

* without the negative sign

\+ Estimated 0.369 but not used in the threshold calculations for consistency with other years of analysis

Table C.83. Number of deaths above LE in 2007/8/9, by PBC

PBC		<LE 2007	>LE 2007	<LE 2008	>LE 2008	<LE 2009	>LE 2009	Annual N deaths <LE	Annual N deaths > LE
1	Infectious diseases	3,906	3,731	3,404	2,588	3,042	2,192	3,451	2,837
2	Cancer	95,385	35,401	94,814	37,088	94,218	37,209	94,806	36,566
4	Endocrine	3,970	2,747	4,031	2,879	3,832	2,828	3,944	2,818
7	Neurological	8,852	6,494	9,632	6,865	9,439	6,945	9,308	6,768
10	Circulatory	80,687	78,404	80,834	76,352	75,993	73,397	79,172	76,051
11	Respiratory	29,571	35,029	32,059	35,204	29,890	33,326	30,507	34,520
13	Gastro-intestinal	15,667	8,367	15,937	8,267	15,354	8,168	15,653	8,267
17	Genito-urinary	4,077	6,553	4,468	6,670	4,375	6,903	4,307	6,709
18+19	Maternity & neonates	216	0	267	0	281	1	255	0

Table C.84: Net YLL using LE of the PBC (2007)

PBC		LE of Males	LE of Females	Average2007-2009 Deaths <LE	Deaths >LE	YLL	YLG	Net YLL
1	Infectious diseases	79.6	83.6	3,280	3,008	57,715	19,085	38,629
2	Cancer	83.0	84.7	100,810	30,561	1,464,726	129,810	1,334,916
4	Endocrine	81.0	84.7	4,004	2,759	66,575	15,386	51,189
7	Neurological	79.6	83.3	8,719	7,357	135,760	44,925	90,835
10	Circulatory	83.0	86.5	92,729	62,494	1,069,632	276,368	793,264
11	Respiratory	80.3	84.0	29,668	35,359	304,168	230,245	73,922
13	Gastro-intestinal	80.6	84.5	15,640	8,280	271,092	45,500	225,593
17	Genito-urinary	83.5	85.6	5,008	6,007	47,656	30,931	16,725
18+19	Maternity & neonates	78.7	83.1	255	0	18,844	1	18,843

Table C.85: Comparing deaths and YLL from ONS and GBD (2007).

		Excess deaths ONS	deaths All deaths ONS	deaths All deaths GBD*	deaths adjustment factor (deaths)	YLL Net estimates ONS	YLL Total YLL GBD*	YLL adjustment factor (YLL)
1	Infectious diseases	2,925	6,288	1,408	4.47	38,629	25,142	1.54
2	Cancer	94,827	131,372	140,124	0.94	1,334,916	1,932,637	0.69
4	Endocrine	3,765	6,762	7,509	0.90	51,189	95,401	0.54
7	Neurological	6,692	16,076	12,854	1.25	90,835	164,796	0.55
10	Circulatory	76,322	155,223	178,454	0.87	793,264	1,750,608	0.45
11	Respiratory	8,034	65,027	67,441	0.96	73,922	594,529	0.12
13	Gastro-intestinal	15,064	23,920	28,329	0.84	225,593	396,829	0.57
17	Genito-urinary	2,005	11,016	8,606	1.28	16,725	77,338	0.22
18+19	Maternity & neonates	255	255	2,211	0.12	18,843	149,868	0.13
	Total	*209,890*	*415,939*	*446,936*	*0.93*	*2,643,916*	*5,187,148*	*0.51*

Table C.86: Summary of the cost per QALY threshold (2007)

	2006 [1]	**2007** [2]	2008 [3]
big 4 PBC's	£3,036	**£4,549**	£4,872
11 PBCs (with mortality)	£5,128	**£8,513**	£8,308
All 23 PBCs	£15,701	**£18,624**	£18,317

Table 87: Breakdown of the cost per QALY threshold (2007)

PBC	PBC description	Change in spend, £m [1]	QALY burden (HoDAR and MEPs) Change in QALY [4]	Cost per QALY gained, £ [5]
2	Cancer	£41	3 041	£13,384
10	Circulatory problems	£19	2 756	£6,724
11	Respiratory problems	£18	13 152	£1,398
13	Gastro-intestinal problems	£24	3 316	£7,137
	Big 4			£4,549
1	Infectious diseases	£16	1 035	£15,530
4	Endocrine problems	£5	910	£5,796
7	Neurological problems	£33	5 111	£6,409
17	Genito-urinary problems	£35	974	£35,449
16	Trauma & injuries*	£49	0	NA
18+19	Maternity & neonates*	£19	6	£3,250,386
	First 11 PBC's			£8,513
3	Disorders of Blood	£28	878	£32,310
5	Mental Health Disorders	£165	4 331	£38,145
6	Learning Disability	£19	159	£119,676
8	Problems of Vision	£29	306	£93,716
9	Problems of Hearing	£7	554	£11,960
12	Dental problems	£20	624	£32,214
14	Skin	£10	136	£76,382
15	Musculo skeletal system	£49	2 164	£22,545
20	Poisoning and adverse effects	£8	68	£123,247
21	Healthy Individuals	£25	29	£858,150
22	Social Care Needs	£52	0	NA
23	Other	£66	0	NA
	All (23 PBCs)			£18,624

Table C.88: Decomposing estimated QALY effects by PBC (2007)

PBC		QALY change (total) [1]	QALY change (death) [2]	% QALY gained for premature death [3]	for disability while alive [4]
2	Cancer	3,041	2,820	93%	7%
10	Circulatory	2,756	1,886	68%	32%
11	Respiratory	13,152	547	4%	96%
13	Gastro-intestinal	3,316	1,223	37%	63%
1	Infectious diseases	1,035	206	20%	80%
4	Endocrine	910	50	5%	95%
7	Neurological	5,111	210	4%	96%
17	Genito-urinary	974	197	20%	80%
16	Trauma & injuries*	0	0	NA	NA
18+19	Maternity & neonates*	6	4	69%	31%
3	Disorders of Blood	878	45	5%	95%
5	Mental Health	4,331	378	9%	91%
6	Learning Disability	159	31	20%	80%
8	Problems of Vision	306	12	4%	96%
9	Problems of Hearing	554	4	1%	99%
12	Dental problems	624	1	0%	100%
14	Skin	136	50	37%	63%
15	Musculo skeletal	2,164	108	5%	95%
20	Poisoning and AE	68	11	16%	84%
21	Healthy Individuals	29	5	16%	84%
22	Social Care Needs	0	0	NA	NA
23	Other	0	0	NA	NA

Summary of the cost per QALY threshold

Table C.89: Summary of cost per QALY threshold estimates (2007)

QoL associated with life extension: / *QoL during disease:*	*Based on burden*	
	Best estimate	
Effect of expenditure on mortality:	*1 year*	
YLL per death averted:	*~ 4.6 YLL*	
QALYs per death averted:	*~ 12.7 QALY*	
big 4 PBC's	£4,549	[1]
11 PBCs (with mortality)	£8,513	[2]
All 23 PBCs	**£18,624**	[3]
	Lower bound	
Effect of expenditure on mortality:	*Remainder of disease duration*	
YLL per death averted:	*~ 4.6 YLL*	
QALYs per death averted:	*~ 12.7 QALY*	
big 4 PBC's	£1,116	[4]
11 PBCs (with mortality)	£1,361	[5]
All 23 PBCs	**£3,247**	[6]
	Upper bound	
Effect of expenditure on mortality:	*1 year*	
YLL per death averted:	*2 YLL*	
QALYs per death averted:	*~ 5.6 QALY*	
big 4 PBC's	£10,965	[7]
11 PBCs (with mortality)	£20,517	[8]
All 23 PBCs	**£44,889**	[9]

Table C.90: Summary of QALY threshold, discounted (2007)

	2006-2008 [1]	**2007-2009** [2]	2008-2010 [3]	
big 4 PBC's	£3,036	**£4,690**	£4,998	[1]
11 PBCs	£5,218	**£8,718**	£8,467	[2]
All 23 PBCs	£15,701	**£18,987**	£18,613	[3]

Appendix C: Addendum 1

DATA SOURCES

Contents

References

A. General Practice Research Database (GPRD)

GPRD contains over 3 million active patient records drawn from approximately 400 primary care practices in the UK. The Medicine Control Agency manages the dataset. The database has clinical and prescription data and can provide information to support pharmaco-vigilance (indication, utilization, and risk/benefit profiles of drugs) and formal pharmaco-epidemiologic studies, including information on demographics, medical symptoms, therapy (medicines, vaccines, devices), and treatment outcomes.

As of 29th March 2012 GPRD has become the Clinical Practice Research Database (CPRD), an expanded dataset that represents 'The All England Data and Interventional Research Service'. GPRD was approached to provide information on the prevalence of disease by ICD-10 disease code. A sample set of data was analysed by researchers at Pharmatelligence[23] who were tasked with extracting data on prevalence of each disease state by ICD-10.

We were provided with access to data comprising of 22,313,086 rows/patient–ICD10 events (3 digit)[24] representing 4,229,910 patients with data on new diagnosis of diseases observed between 1 Jan 2006 and 24 June 2011. Multiple events per patient are thus possible, and all patients are active in the dataset, i.e. patients had at least one new diagnosis in the period of interest. Newly diagnosed (incident) events were defined using a wash-in period of 24 months (or from registration to index date if lower than 24 months). The sample contains 1873 unique ICD codes in the dataset. 70 ICD codes account for 50% of the total number of events, 166 for 75% and 306 for 90%.

Diagnoses are collected in GPRD using Read codes. These were mapped into three-character ICD-10 codes. Cross-mappings from Read V2 and Read V3 to ICD-10 were used in order to maximize the number of GPRD Read and ICD-10 codes included (33.2% of Read codes; 99.7% of ICD codes)[25].

---

[23] Prof. Craig Currie and Sara Jenkins-Jones

[24] This represents six fewer than the incidence data as in these instances the end dates for the disease were beyond the end of the data collection period.

[25] Mapping algorithms were provided by the NHS Connecting for Health group, see http://www.connectingforhealth.nhs.uk/systemsandservices/data/clinicalcoding/crossmap for more details

Unfortunately due to the short collection period of GPRD it was not possible to directly observe prevalence only incidence over a period. Attempts were made to elicit prevalence estimate through observed incidence data from GPRD coupled with clinical expertise on expected disease duration (provided by Dr. Charlotte Haylock, Hull and East Yorkshire Hospitals NHS Trust). Our approach classified expected duration for all ICD-10 diseases by 3-digit code into one of five duration 'buckets'[26]. However, the limitations of the data were deemed too extensive to provide sufficient accuracy of estimates to represent a stronger estimate of prevalence than provide by GBD.

B. Global Burden of Disease (GBD)

The WHO GBD project draws on a wide range of data sources to quantify global and regional effects of diseases, injuries and risk factors on population health. We were provided with access to the beta version of the WHO's National Burden of Disease (NBD) toolkit for the United Kingdom which represents a set of metrics on World Health Organisation (WHO) prior estimates of mortality and burden of disease for WHO Member states for 2004 (based on the Global Burden of Disease: 2004 update[9])[27].

The metrics of interest to our analysis included disease incidence, prevalence, duration and mortality. These metrics were provided by U-code disease code which were mapped to ICD-10 using direct WHO mapping algorithms[10]. In addition, in many cases each U-code was sub-divided by disease sequela which represent disease sub-categories of each U-code[10]. As an individual may be represented in multiple sequela in a single U-code to avoid double counting in the event of multiple sequela in a given U-code our analysis uses prevalence estimates based on the sequela with the largest prevalent population.

Our analysis uses two forms of prevalence data, "point prevalence" and "annual prevalence". "Point Prevalence" represents the instantaneous prevalence of a disease whereas "annual prevalence" represents the extent of the prevalence population over a given year. To calculate "annual prevalence" incidence of a disease was multiplied by expected disease duration rounded up to the nearest year.

All data was provided by age, given for both genders in fixed age buckets (either eight or nineteen buckets depending on the data of interest), as a result it was necessary to assume the relevant population could be represented by the mid-point of that bucket for the relevant metric.

C. Health Survey for England (HSE)

The Health Survey for England (HSE) comprises a series of annual surveys beginning in 1991. This survey is now commissioned and published by The NHS Information Centre. It is designed to provide regular information on various aspects of the nation's health. All surveys have covered the adult population aged 16 and over living in private households in England.[28]

In order to define the quality of life norms for the population of the UK required for the analysis detailed in section 4.3.1 data from six Health Surveys for England (1996, 2003-2006 and 2008) were pooled. Self-reported health status and EQ-5D data were extracted and used to generate mean health state utility values for the 'normal' population.

---

[26] Representing instantaneous, one month, one year, five years and life-long

[27] For more information on access to the Toolkit see http://www.who.int/healthinfo/global_burden_disease/tools_nbd_toolkit/en/index.html

[28] For more information on the surveys and the data they collect see http://www.dh.gov.uk/en/Publicationsandstatistics/PublishedSurvey/HealthSurveyForEngland/index.htm

Surveys are not completed for people age under 16; as a result we have assumed that all persons age 0 to 15 have the same quality of life norms as a person age 16. In addition the number of surveys recorded for persons over 91 years of age is relatively small, as a result all persons over 91 are assumed to have the same quality of life norm as a person age 91. The quality of life norms for each age and gender are shown in figure 1 in section 4.3.1.

D. Health Outcome Data Repository (HODaR)

HODaR represents a supplement of routine clinically coded data from the Cardiff and Vale NHS Hospitals Trust, UK, with survey data covering socio-demographic characteristics, QoL, utility, and resource use information[4]. HODaR data was collected for subjects treated at Cardiff and Vale NHS hospital from 2002 to 2004. Inpatients were surveyed 6 weeks post-discharge whilst outpatients are handed a survey package when they attend. More than 30,000 observations (aged above 18) are available relating to approximately 2,000 diagnoses of disease by ICD-10.

We used HODaR to estimate Health Related Quality of Life (HRQoL) by ICD-10 diagnoses codes and age using EQ-5D. If data on a patient was provided with multiple diagnoses the primary condition was used.

E. Medical Expenditure Panel Survey (MEPS)

MEPS is a national representative survey of the US civilian non-institutionalised population, collecting information on health care utilisation which began in 1996.[5] EQ-5D was employed to measure HRQoL of the population in years 2000 to 2002. There are about 38,000 adults (aged above 18) completing EQ-5D relating to 700 ICD-10 diagnoses. MEPS consists of a household component and an insurance component, both aimed at identifying the medical usage of individuals as well as how they are funded, their cost, and the scope and breadth of health insurance held and available.

As with HODaR, MEPS allowed us to estimate the HRQoL by ICD-10 code and age. If data on a patient was provided with multiple conditions the primary condition was used.

F. Hospital Episode Statistics (HES)

HES represents a collection of data with details on all admissions to the NHS hospital in England. It contains admitted patient care data from 1989 onwards, with more than 12 million new records added each year, and outpatient attendance data from 2003 onwards, with more than 40 million new records added each year.

Expenditure by ICD-10 codes and PCT was used to estimate the contribution to variance of each PBC. This was done by calculating the contribution of that ICD to the variance in expenditure between PCTs within a PBC (total costs allocated to individual ICDs were divided by the number of patients using services in the PCT). For our analysis we make use of HES data on the year 2007/08.

G. Patient Reported Outcome Measures (PROMs)

Introduced in 2009, the English NHS Patient Reported Outcomes (PROMs) programme routinely collects self-reported health status of patients receiving surgery for four elective procedures: knee and hip replacement, groin hernia repair, and varicose vein surgery. Patients are invited to complete a questionnaire prior to surgery, and again six (or three) months after surgery.[11] Differences in their self-reported health status are used to explore differences between provider performance in improving patient health.[12] The data that are collected include both condition specific questions (the Oxford Hip Score,

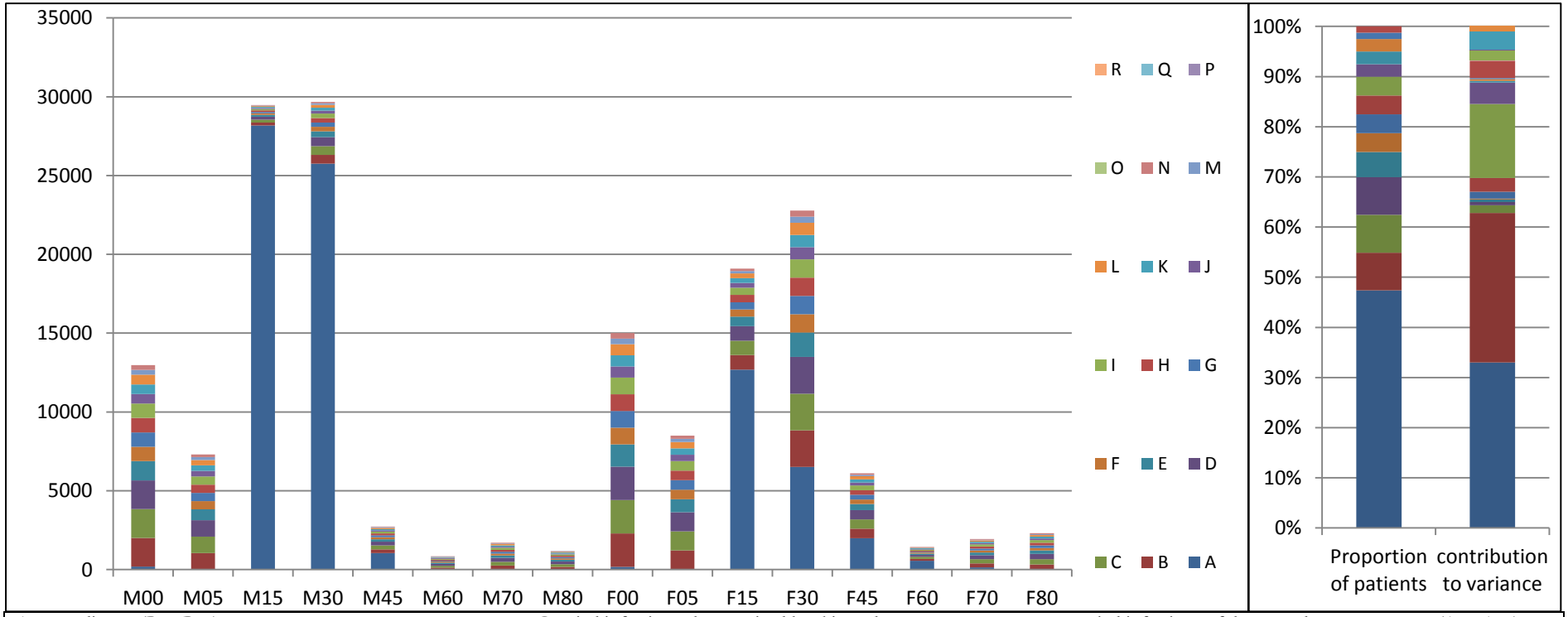
Oxford Knee Score and the Aberdeen Varicose Vein score; no condition specific instrument is available for hernia) as well as the generic instrument, the EQ-5D (both the EQ-5D profile, and the patient's global assessment of their health, the EQ-VAS - see.[13] All NHS patients receiving these surgical procedures are invited to complete the PROMs questionnaires - in practice, for a variety of reasons, some patients do not participate, or complete only the pre-surgery, or the post-surgery, questionnaire - so the data do not cover 100% of patients. However, good coverage rates have been achieved - for example, the response rate from hip surgery patients to April 2012 was 78% for the pre-surgery questionnaire, and 81% on the post-surgery questionnaire.[14]

Patient-level data from the PROMs programme are freely available to download in anonymised form. Those data can also be linked to further information in the HES database, via requests to the NHS Information Centre. Standardised reports on the PROMs data, including the average (case-mix adjusted) performance of providers, is regularly published by the Information Centre, currently on a quarterly basis.

There are plans to extend the PROMs programme in the future, in keeping with the Government's NHS Outcomes Framework, and a number of pilot studies have been commissioned by the Department of Health in order to inform the roll out to other NHS services. There is currently work underway or being planned around the potential use of PROMs in a wide range of long term conditions; primary care; in cancer survivorship; cardiovascular services; muscular skeletal; and cosmetic surgery.

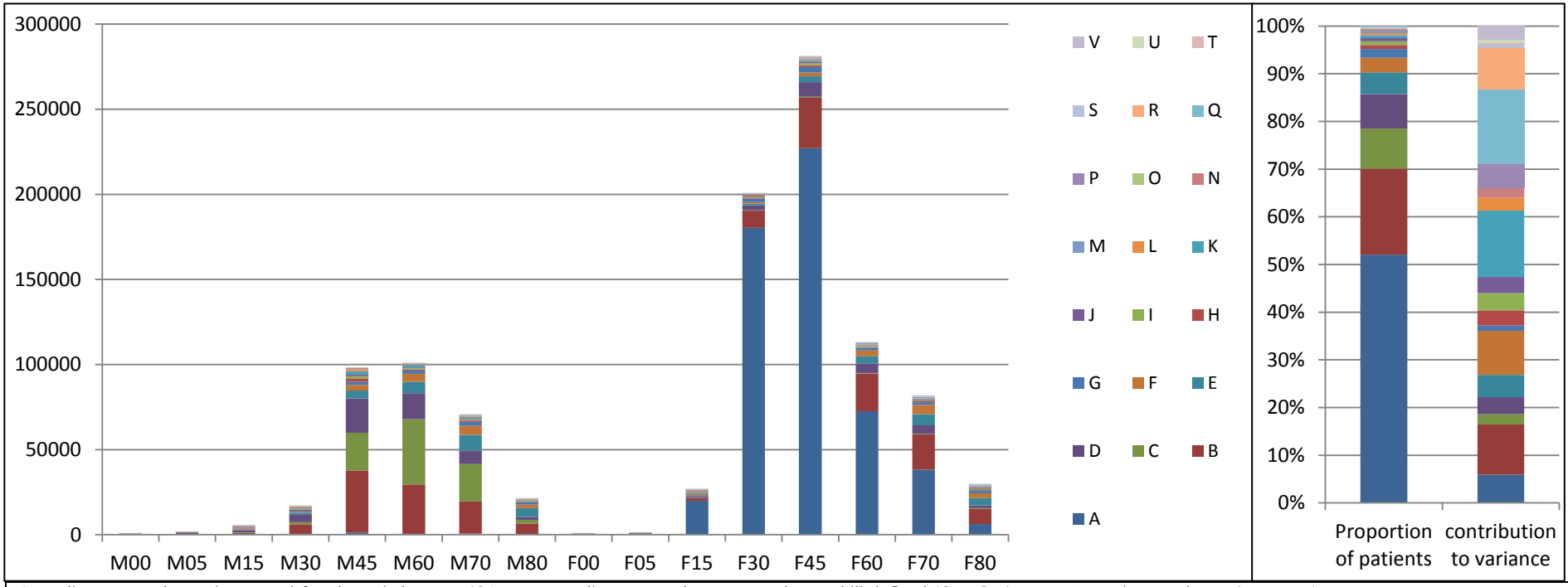
G. Tables of prevalence distribution within PBCs

Table C1.1: Distribution of PBC1 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- HIV disease (B20-B24)	G- Viral infections characterized by skin and mucous membrane lesions (B00-B09)	M- Viral infections of the central nervous system (A80-A89)
B- Other bacterial diseases (A30-A49)	H- Mycoses (B35-B49)	N- Other infectious diseases (B98-B99)
C- Certain zoonotic bacterial diseases (A20-A28)	I- Protozoal diseases (B50-B64)	O- General symptoms and signs (R50-R69)
D- Arthropod-borne viral fevers and viral haemorrhagic fevers (A90-A99)	J- Other viral diseases (B25-B34)	P- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)
E- Rickettsioses (A75-A79)	K- Helminthiases (B65-B83)	Q- Persons with potential health hazards related to communicable diseases (Z20-Z29)
F- Other spirochaetal diseases (A65-A69)	L- Pediculosis, acariasis and other infestations (B85-B89)	R- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

Table C1.2: Distribution of PBC2 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Malignant neoplasms, breast and female genital organs (C50-C58)

B- Malignant neoplasms, digestive organs (C15-C26)

C- Malignant neoplasms, human male genital organsmale genital organs (C60-C63)

D- Malignant neoplasms, urinary organs (C64-C68)

E- Malignant neoplasms, respiratory system and intrathoracic organs (C30-C39)

F- Malignant neoplasms, stated or presumed to be primary, of lymphoid, haematopoietic and related tissue (C81-C96)

G- Malignant neoplasms, skin (C43-C44)

H- Malignant neoplasms, connective and soft tissue (C45-C49)

I- Malignant neoplasms, secondary and ill-defined (C76-C80)

J- Malignant neoplasms, eye, brain and central nervous system (C69-C72)

K- Malignant neoplasms, lip, oral cavity and pharynx (C00-C14)

L- Malignant neoplasms, endocrine glands and related structures (C73-C75)

M- Other diseases of the respiratory system (J95-J99)

N- Malignant neoplasms, bone and articular cartilage (C40-C41)

O- Malignant neoplasms of independent (primary) multiple sites (C97-)

P- In situ neoplasms (D00-D09)

Q- Benign neoplasms (D10-D36)

R- Neoplasms of uncertain or unknown behaviour (D37-D48)

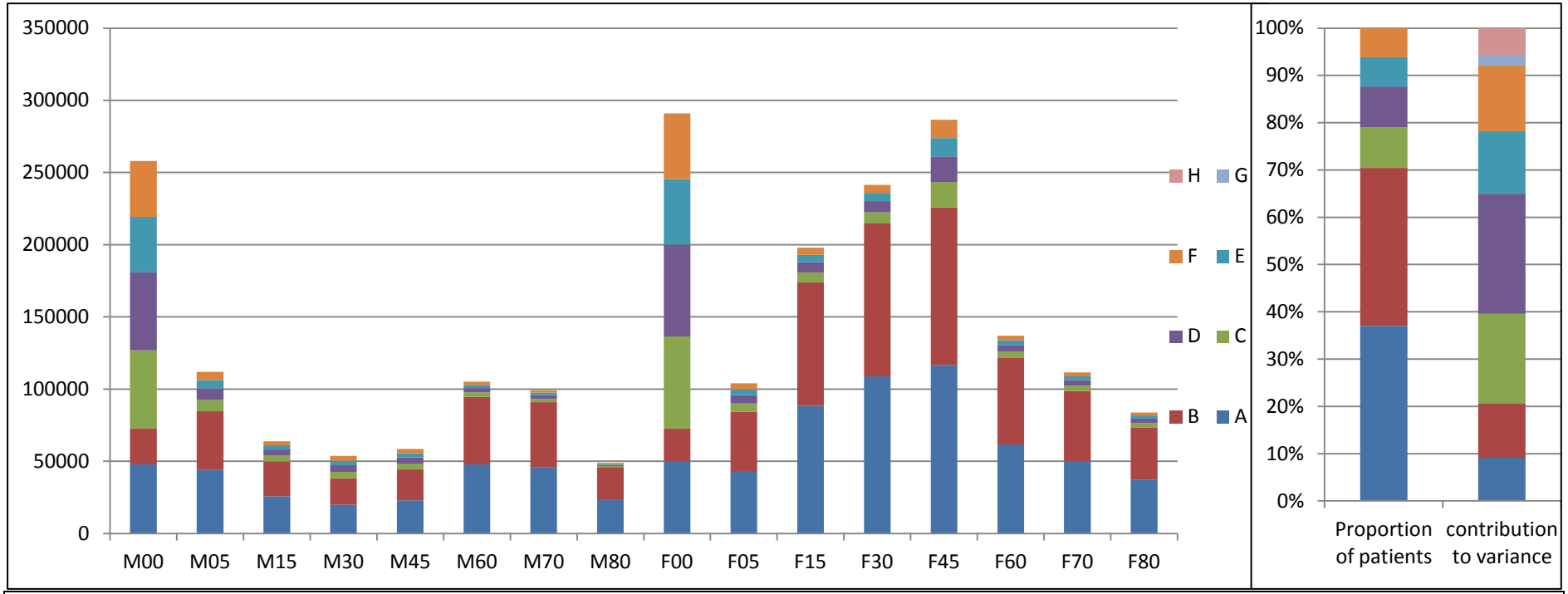
S- Other diseases of urinary system (N30-N39)

T- Persons encountering health services for examination and investigation (Z00-Z13)

U- Persons encountering health services for specific procedures and health care (Z40-Z54)

V- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

Table C1.3: Distribution of PBC3 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Aplastic and other anaemias (D60-D64)

B- Nutritional anemias (D50-D53)

C- Other diseases of blood and blood-forming organs (D70-D77)

D- Certain disorders involving the immune mechanism (D80-D89)

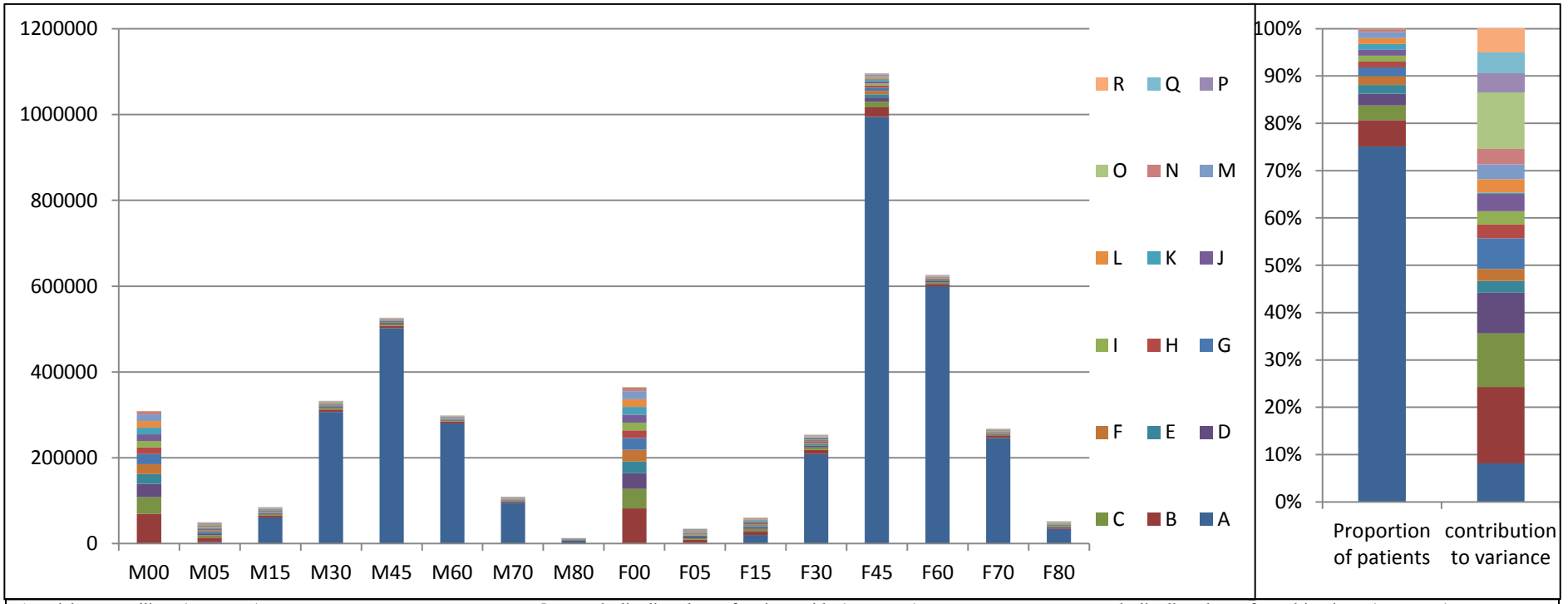
E- Haemolytic anaemias (D55-D59)

F- Coagulation defects, purpura and other haemorrhagic conditions (D65-D69)

G- Congenital malformations and deformations Other (Q80-Q89)

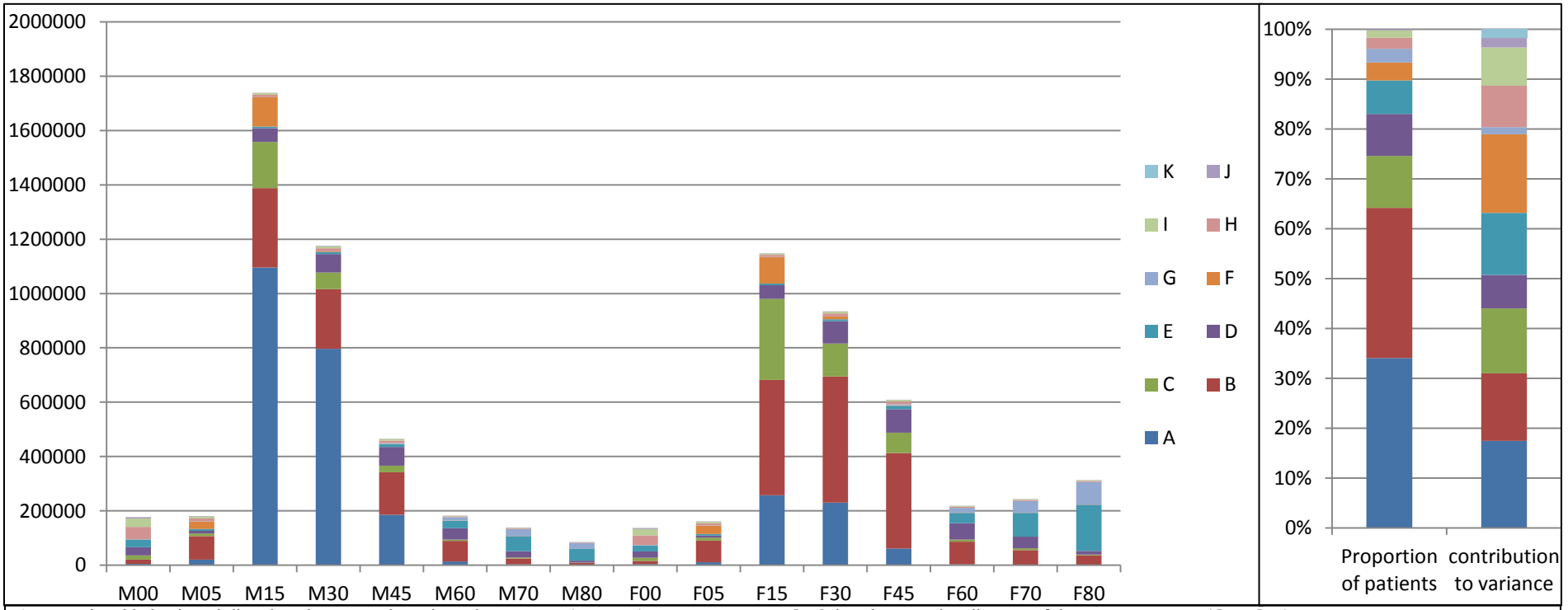
H- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)

Table C1.4: Distribution of PBC4 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Diabetes mellitus (E10-E14)
B- Other metabolic disorders (E79-E90)
C- Thyroid gland / Thyroid hormone (E00-E07)
D- Adrenal gland / Aldosterone, cortisol, epinephrine, norepinephrine (E24-E27)
E- Gonads / Estrogen, androgens, testosterone, etc. (E28-E30)
F- Other endocrine diseases (E31-E35)
G- Metabolic disorders of amino-acids (E70-E72)
H- Other disorders of glucose regulation and pancreatic internal secretion (E15-E16)
I- Parathyroid gland / PTH (E20-E21)
J- Pituitary gland / ADH, oxytocin, GH, ACTH, TSH, LH, FSH, prolactin (E22-E23)
K- Obesity and other hyperalimentation (E65-E68)
L- Metabolic disorders of carbohydrates (E73-E74)
M- Metabolic disorders of combinations (E76-E78)
N- Metabolic disorders of lipids (E75-)
O- Other nutritional deficiencies (E50-E64)
P- Malnutrition (E40-E46)
Q- General symptoms and signs (R50-R69)
R- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)

Table C1.5: Distribution of PBC5 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Mental and behavioural disorders due to psychoactive substance use (F10-F19)

B- Mood (affective) disorders (F30-F39)

C- Neurotic, stress-related and somatoform disorders (F40-F48)

D- Behavioural syndromes associated with physiological disturbances and physical factors (F50-F59)

E- Organic, including symptomatic, mental disorders (F00-F09)

F- Schizophrenia, schizotypal and delusional disorders (F20-F29)

G- Other degenerative diseases of the nervous system (G30-G32)

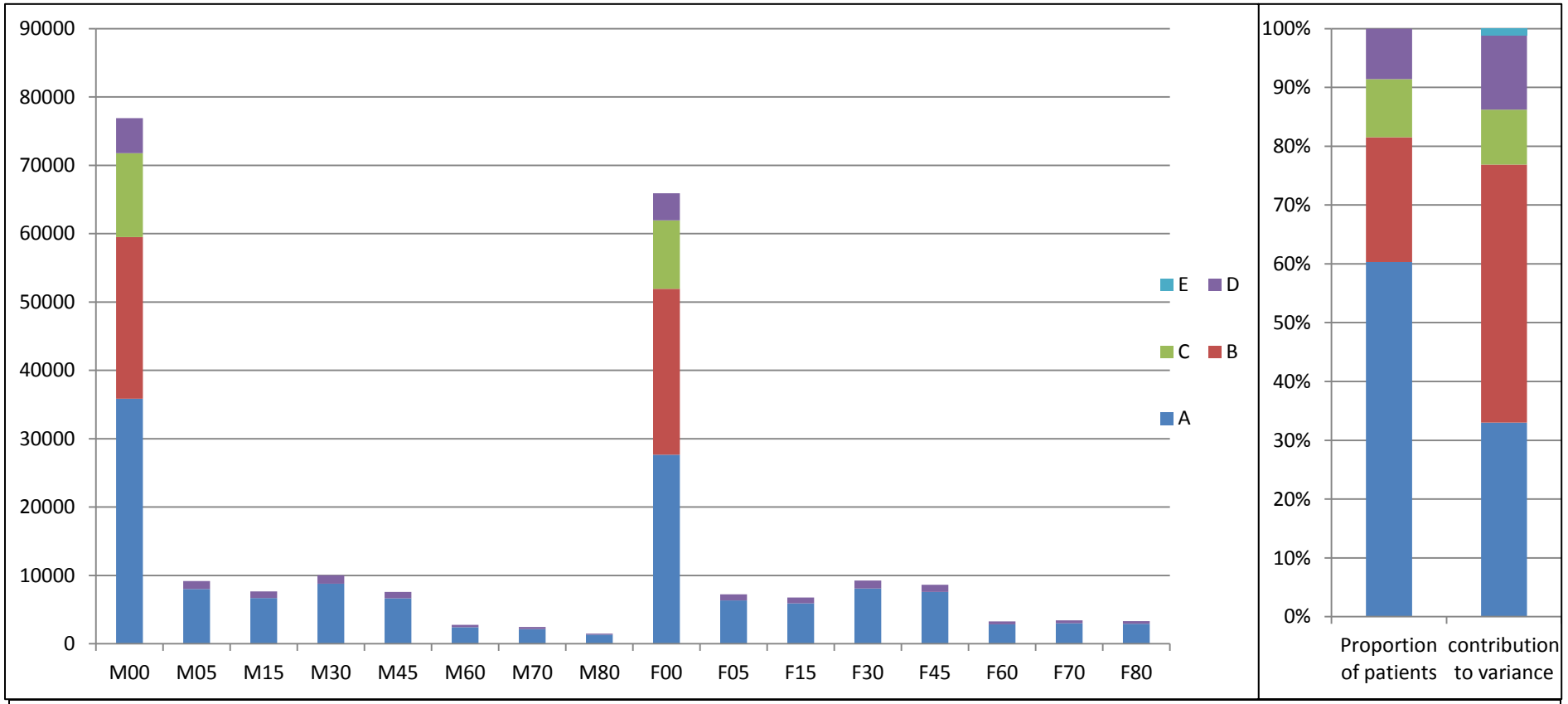
H- Disorders of adult personality and behaviour (F60-F69)

I- Behavioural and emotional disorders with onset usually occurring in childhood and (F90-F98)

J- Unspecified mental disorder (F99-)

K- Persons with potential health hazards related to socioeconomic and psychosocial circumstances (Z55-Z65)

Table C1.6: Distribution of PBC6 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Disorders of psychological development (F80-F89)

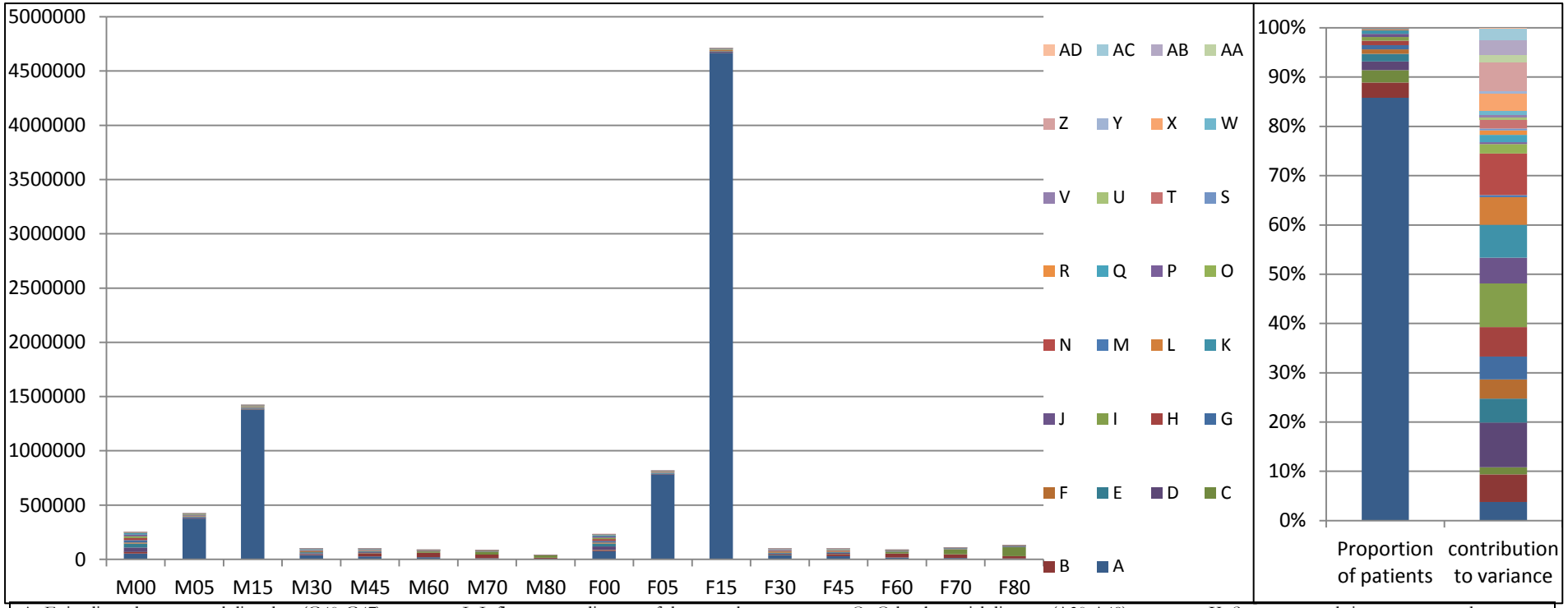
B- Mental retardation (F70-F79)

C- Chromosomal abnormalities, not elsewhere classified (Q90-Q99)

D- Behavioural and emotional disorders with onset usually occurring in childhood and (F90-F98)

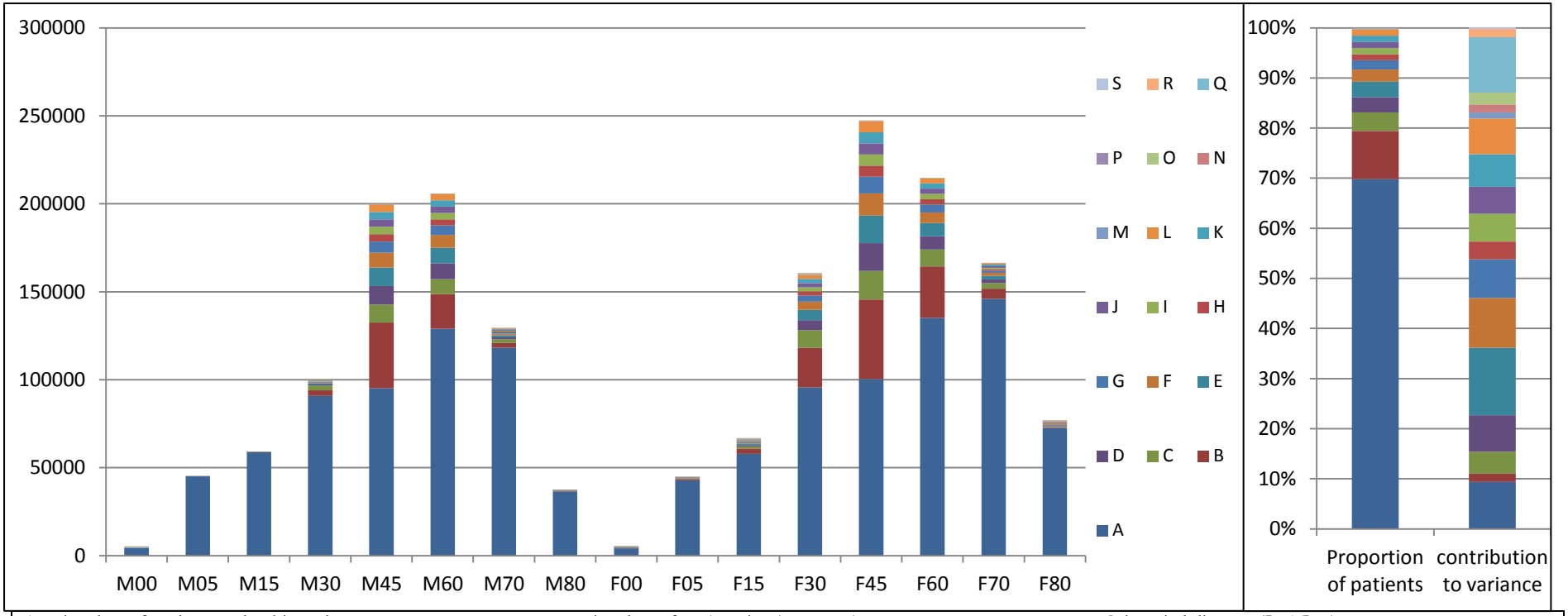
E- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

Table C1.7: Distribution of PBC7 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



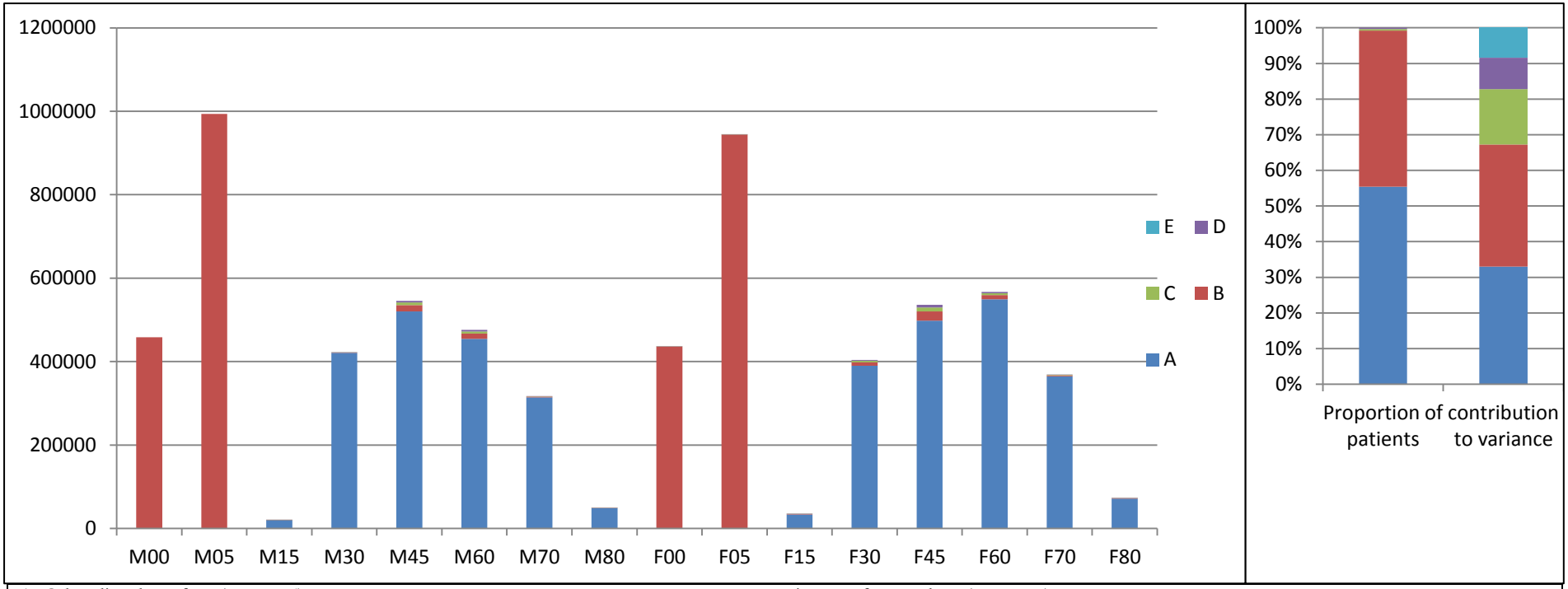
A- Episodic and paroxysmal disorders (G40-G47)	I- Inflammatory diseases of the central nervous system (G00-G09)	Q- Other bacterial diseases (A30-A49)	X- Symptoms and signs nervous and musculoskeletal systems (R25-R29)
B- Extrapyramidal and movement disorders (G20-G26)	J- Systemic atrophies primarily affecting the central nervous system (G10-G13)	R- Other viral diseases (B25-B34)	Y- Symptoms and signs urinary system (R30-R39)
C- Other degenerative diseases of the nervous system (G30-G32)	K- Diseases of myoneural junction and muscle (G70-G73)	S- Helminthiases (B65-B83)	Z- Symptoms and signs cognition, perception, emotional state and behaviour (R40-R46)
D- Other disorders of the nervous system (G90-G99)	L- Viral infections of the central nervous system (A80-A89)	T- Tuberculosis (A15-A19)	AA- Symptoms and signs speech and voice (R47-R49)
E- Nerve, nerve root and plexus disorders (G50-G59)	M- Other disorders of ear (H90-H95)	U- Protozoal diseases (B50-B64)	AB- General symptoms and signs (R50-R69)
F- Demyelinating diseases of the central nervous system (G35-G37)	N- Congenital malformations and deformations nervous system (Q00-Q07)	V- Symptoms and signs circulatory and respiratory systems (R00-R09)	AC- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)
G- Polyneuropathies and other disorders of the peripheral nervous system (G60-G64)	O- Viral infections characterized by skin and mucous membrane lesions (B00-B09)	W- Symptoms and signs skin and subcutaneous tissue (R20-R23)	AD- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)
H- Cerebral palsy and other paralytic syndromes (G80-G83)	P- Other diseases of urinary system (N30-N39)		

Table C1.8: Distribution of PBC8 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Disorders of ocular muscles, binocular movement, accommodation and refraction (H49-H52)
B- Glaucoma (H40-H42)
C- Disorders of lens (H25-H28)
D- Disorders of eyelid, lacrimal system and orbit (H00-H06)
E- Disorders of choroid and retina (H30-H36)
F- Disorders of sclera and cornea (H15-H19)
G- Other disorders of eye and adne (H55-H59)
H- Disorders of conjunctiva (H10-H13)
I- Disorders of iris and ciliary body (H20-H22)
J- Disorders of vitreous body and globe (H43-H45)
K- Disorders of optic nerve and visual pathways (H46-H48)
L- Visual disturbances and blindness (H53-H54)
M- Other diseases casused by chlamydiae (A70-A74)
N- Other viral diseases (B25-B34)
O- Protozoal diseases (B50-B64)
P- Helminthiases (B65-B83)
Q- Congenital malformations and deformation eye, ear, face and neck (Q10-Q18)
R- Persons encountering health services for specific procedures and health care (Z40-Z54)
S- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

Table C1.9: Distribution of PBC9 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Other disorders of ear (H90-H95)

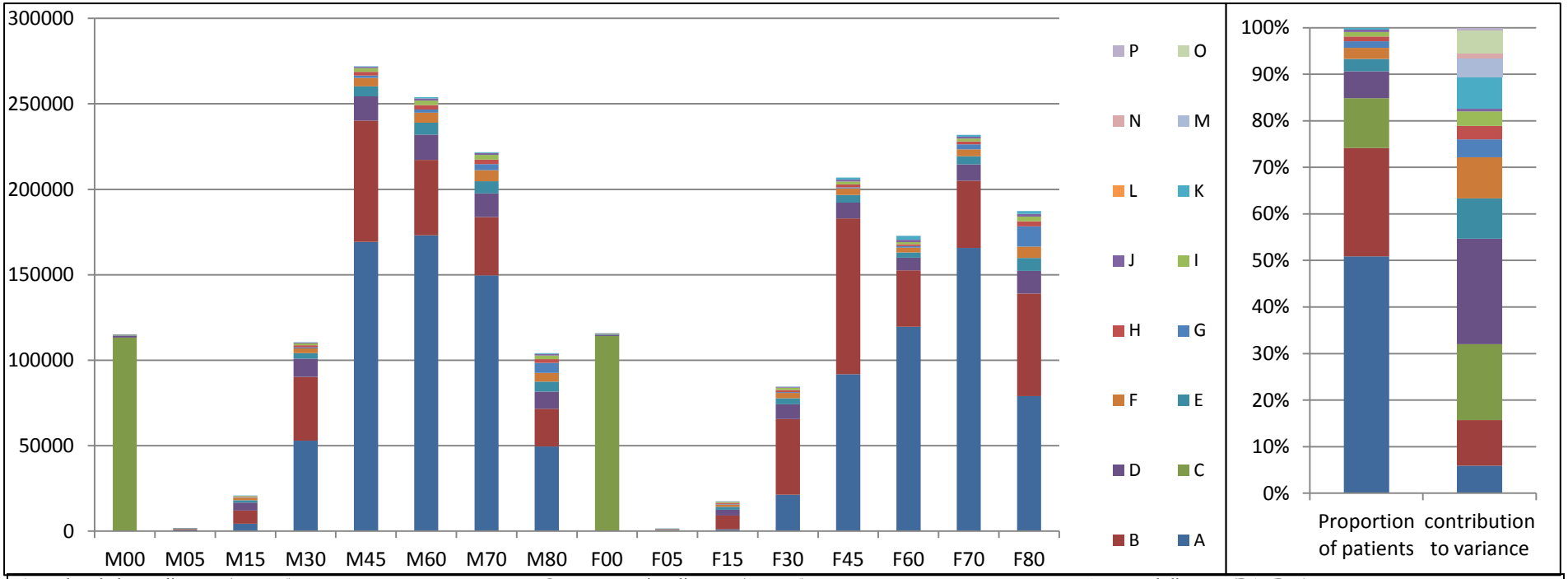
B- Diseases of middle ear and mastoid (H65-H75)

C- Diseases of inner ear (H80-H83)

D- Diseases of external ear (H60-H62)

E- Congenital malformations and deformation eye, ear, face and neck (Q10-Q18)

Table C1.10: Distribution of PBC10 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Ischemic heart diseases (I20-I25)
B- Cerebrovascular diseases (I60-I69)

C- Congenital malformations and deformations circulatory system (Q20-Q28)
D- Other forms of heart disease (I30-I52)

E- Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified (I80-I89)

F- Diseases of arteries, arterioles and capillaries (I70-I79)

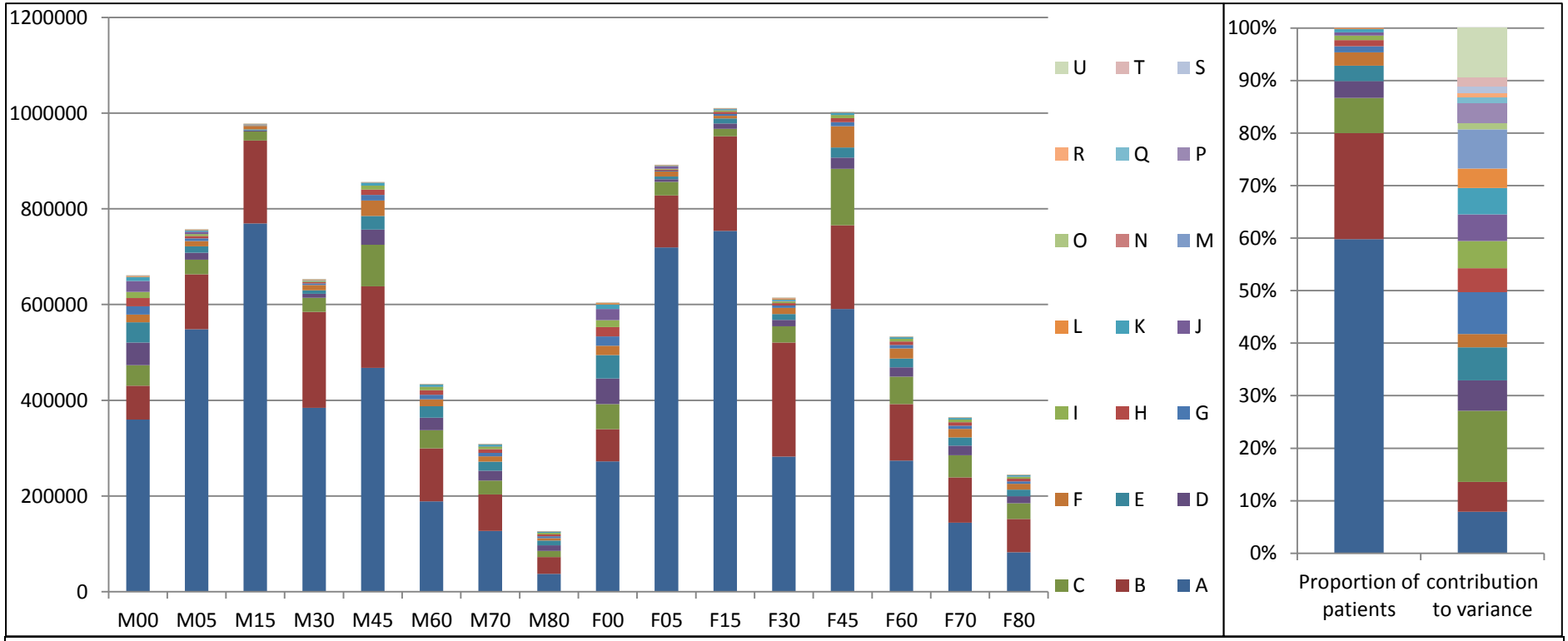
G- Hypertensive diseases (I10-I15)
H- Pulmonary heart disease and diseases of pulmonary circulation (I26-I28)
I- Other and unspecified disorders of the circulatory system (I95-I99)
J- Acute rheumatic fever (I00-I02)

K- Chronic rheumatic heart diseases (I05-I09)

L- Protozoal diseases (B50-B64)
M- Symptoms and signs circulatory and respiratory systems (R00-R09)
N- General symptoms and signs (R50-R69)

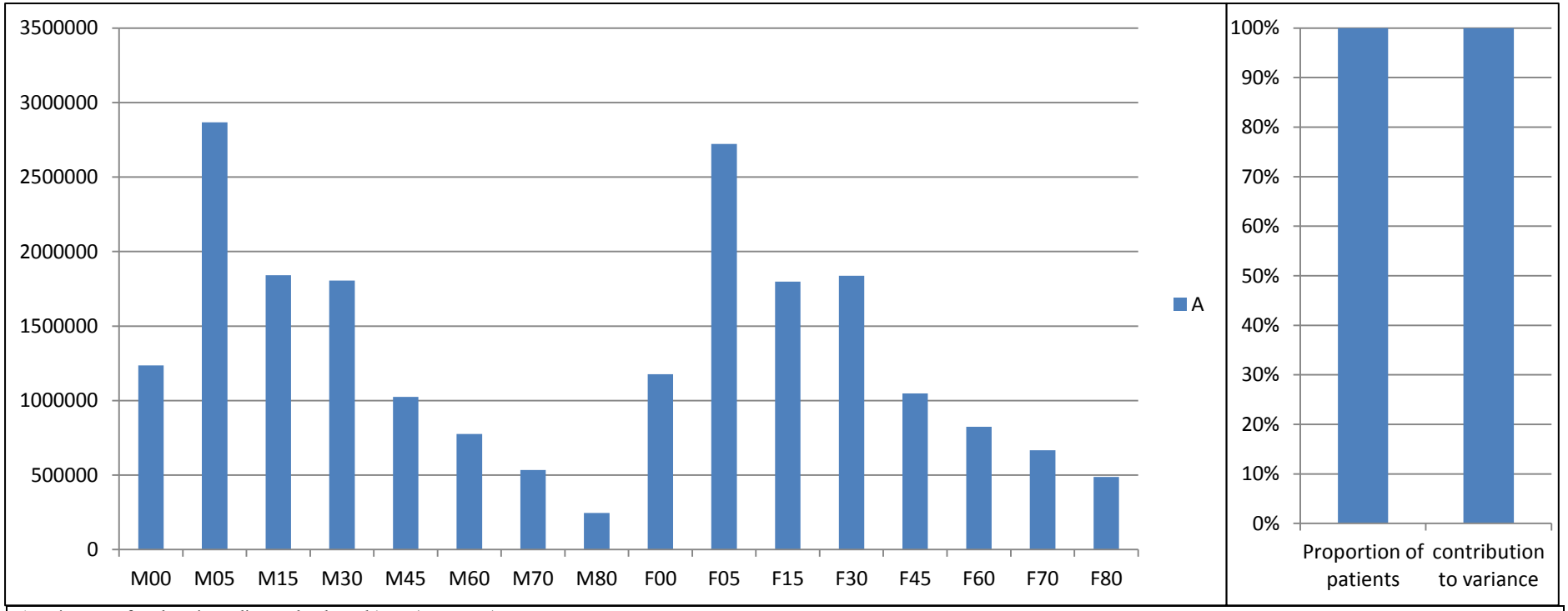
O- Persons encountering health services for specific procedures and health care (Z40-Z54)
P- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

Table C1.11: Distribution of PBC11 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



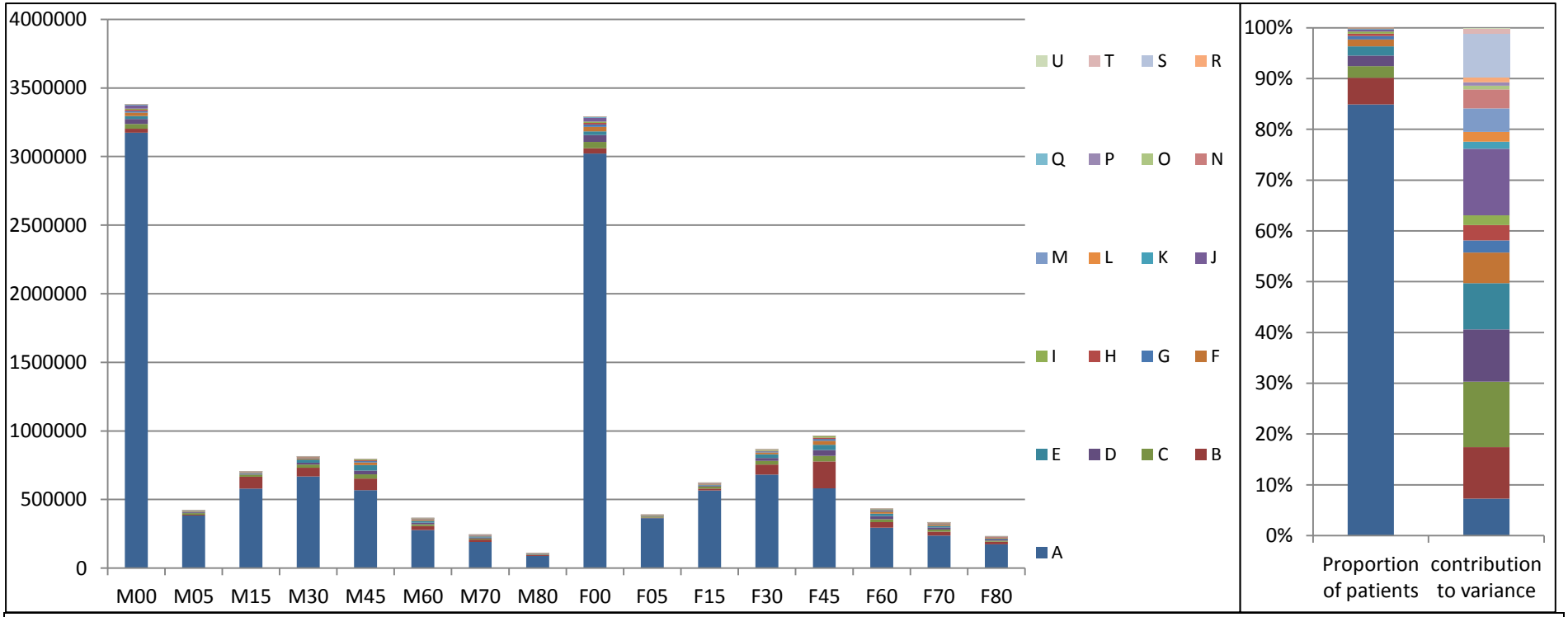
A- Chronic lower respiratory diseases (J40-J47)
B- Acute upper respiratory infections (J00-J06)
C- Influenza and Pneumonia (J09-J18)
D- Lung diseases due to external agents (J60-J70)
E- Other diseases of upper respiratory tract (J30-J39)
F- Other acute lower respiratory infections (J20-J22)
G- Other respiratory diseases principally affecting the interstitium (J80-J84)
H- Other diseases of pleura (J90-J94)
I- Other diseases of the respiratory system (J95-J99)
J- Other bacterial diseases (A30-A49)
K- Suppurative and necrotic conditions of lower respiratory tract (J85-J86)
L- Mycoses (B35-B49)
M- Tuberculosis (A15-A19)
N- Other diseases casused by chlamydiae (A70-A74)
O- Other viral diseases (B25-B34)
P- Symptoms and signs circulatory and respiratory systems (R00-R09)
Q- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)
R- Persons encountering health services for examination and investigation (Z00-Z13)
S- Persons encountering health services for specific procedures and health care (Z40-Z54)
T- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)
U- Congenital malformations and deformations respiratory system (Q30-Q34)

Table C1.12: Distribution of PBC12 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Diseases of oral cavity, salivary glands and jaws (K00-K14)

Table C1.13: Distribution of PBC13 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Intestinal infectious diseases (A00-A09)

B- Diseases of oesophagus, stomach and duodenum (K20-K31)

C- Diseases of appendix (K35-K38)

D- Other diseases of intestines (K55-K63)

E- Diseases of liver (K70-K77)

F- Disorders of gallbladder, biliary tract and pancreas (K80-K87)

G- Noninfective enteritis and colitis (K50-K52)

H- Diseases of peritoneum (K65-K67)

I- Other diseases of the digestive system (K90-K93)

J- Congenital malformations and deformations digestive system (Q35-Q45)

K- Helminthiases (B65-B83)

L- Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified (I80-I89)

M- Viral hepatitis (B15-B19)

N- Diseases of oral cavity, salivary glands and jaws (K00-K14)

O- Certain zoonotic bacterial diseases (A20-A28)

P- Other spirochaetal diseases (A65-A69)

Q- Arthropod-borne viral fevers adn viral haemorrhagic fevers (A90-A99)

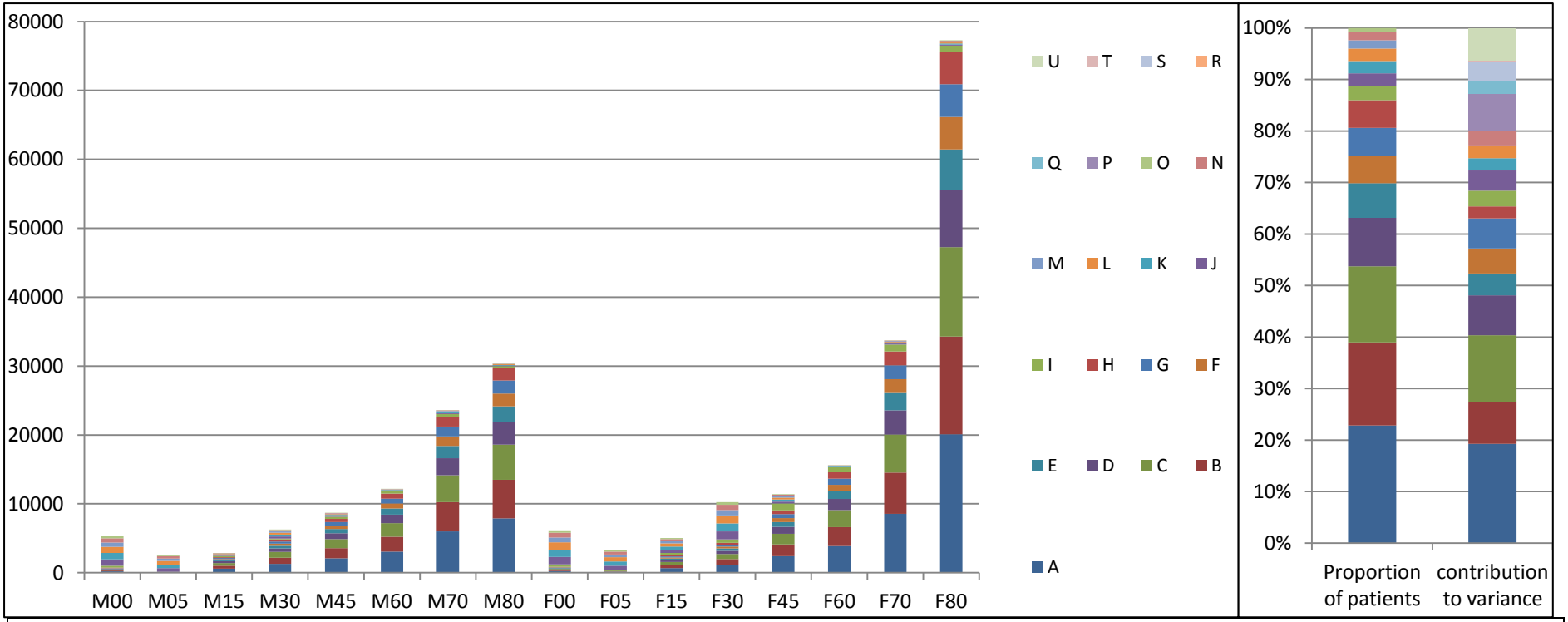
R- Mycoses (B35-B49)

S- Symptoms and signs digestive system and abdomen (R10-R19)

T- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)

U- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

Table C1.14: Distribution of PBC14 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Other disorders of the skin and subcutaneous tissue (L80-L99)

B- Disorders of skin appendages (L60-L75)

C- Dermatitis and eczema (L20-L30)

D- Infections of the skin and subcutaneous tissue (L00-L08)

E- Papulosquamous disorders (L40-L45)

F- Bullous disorders (L10-L14)

G- Urticaria and erythema (L50-L54)

H- Radiation-related disorders of the skin and subcutaneous tissue (L55-L59)

I- Other diseases of urinary system (N30-N39)

J- Viral infections characterized by skin and mucous membrane lesions (B00-B09)

K- Mycoses (B35-B49)

L- Pediculosis, acariasis and other infestations (B85-B89)

M- Certain zoonotic bacterial diseases (A20-A28)

N- Other bacterial diseases (A30-A49)

O- Other spirochaetal diseases (A65-A69)

P- Congenital malformations and deformations Other (Q80-Q89)

Q- Persons encountering health services for specific procedures and health care (Z40-Z54)

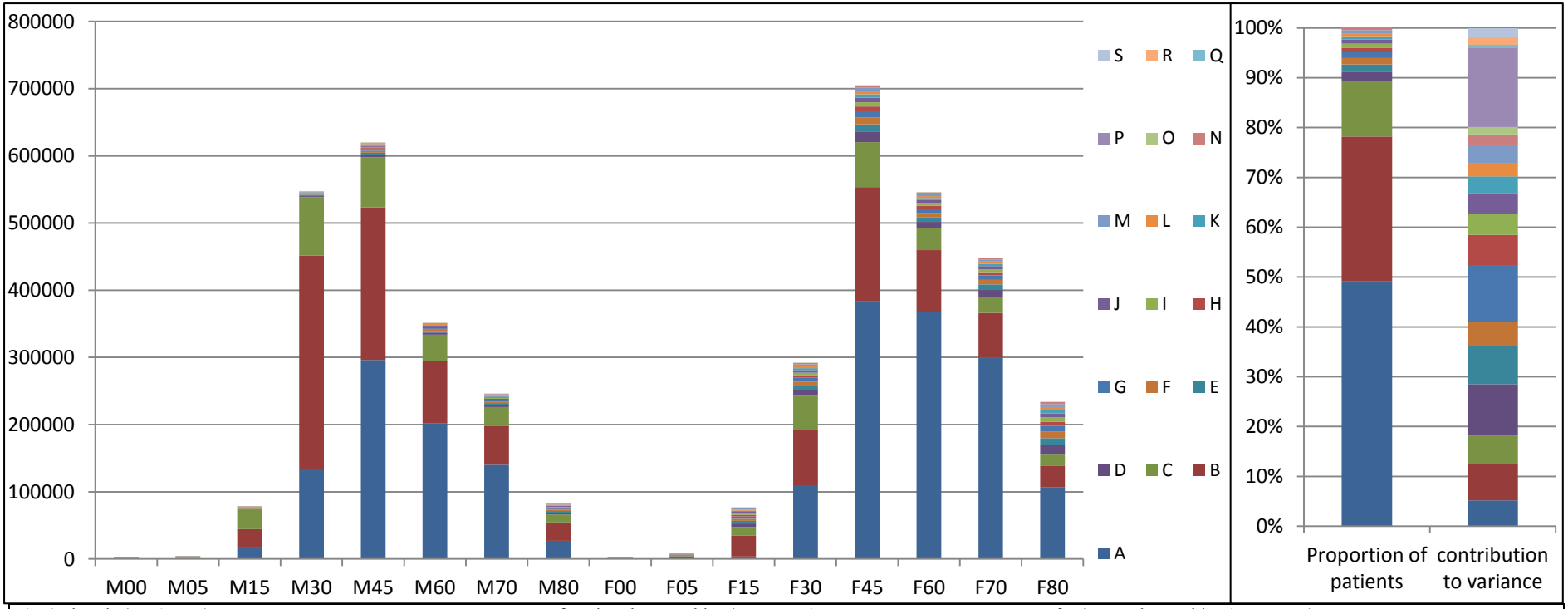
R- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

S- Symptoms and signs skin and subcutaneous tissue (R20-R23)

T- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)

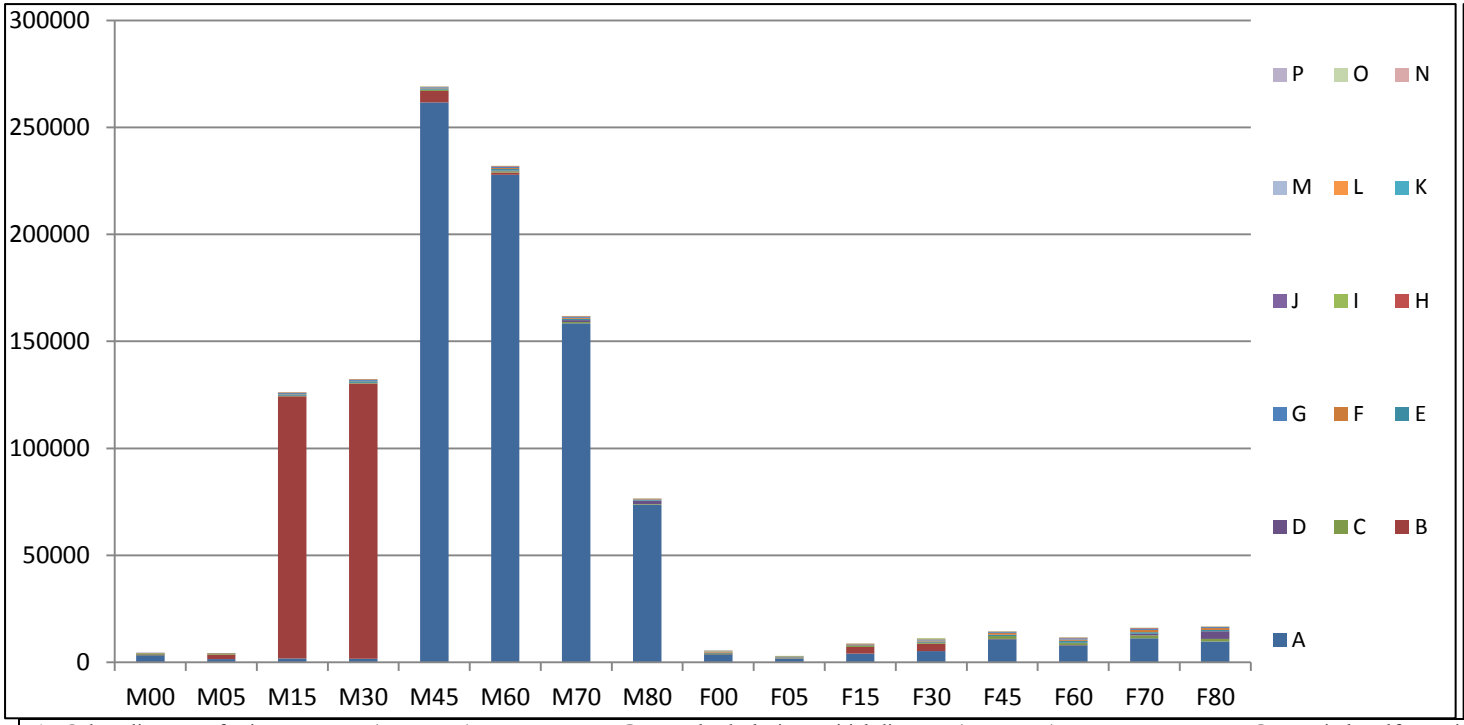
U- Burns and corrosions (T20-T32)

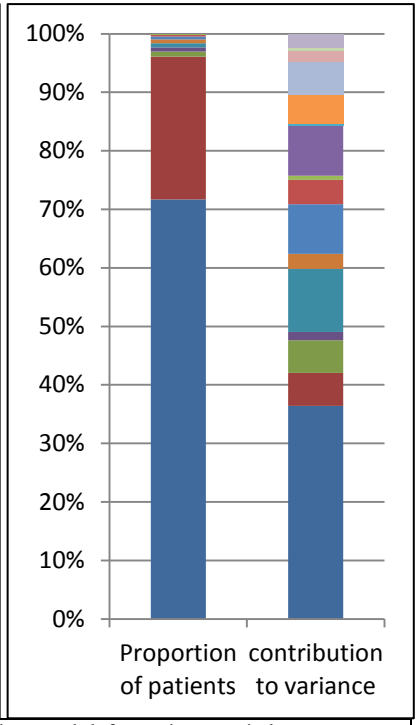
Table C1.15: Distribution of PBC15 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Arthrosis (M15-M19)
B- Inflammatory polyarthropathies (M05-M14)
C- Spondylopathies (M45-M49)
D- Osteopathies (M80-M90)
E- Other soft tissue disorders (M70-M79)
F- Other joint disorders (M20-M25)
G- Systemic connective tissue disorders (M30-M36)
H- Deforming dorsopathies (M40-M43)
I- Other dorsopathies (M50-M54)
J- Chondropathies (M91-M94)
K- Disorders of muscles (M60-M63)
L- Disorders of synovium and tendon (M65-M68)
M- Other disorders of the musculoskeletal system and connective tissue (M95-M99)
N- Infectious arthropathies (M00-M03)
O- Tuberculosis (A15-A19)
P- Congenital malformations and deformations musculoskeletal system (Q65-Q79)
Q- Congenital malformations and deformation eye, ear, face and neck (Q10-Q18)
R- Congenital malformations and deformations Other (Q80-Q89)
S- Persons encountering health services for specific procedures and health care (Z40-Z54)

Table C1.16: Distribution of PBC17 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD





A- Other diseases of urinary system (N30-N39)

B- Infections with a predominantly sexual mode of transmission (A50-A64)

C- Other disorders of kidney and ureter (N25-N29)

D- Hypertensive diseases (I10-I15)

E- Glomerular diseases (N00-N08)

F- Urolithiasis (N20-N23)

G- Renal tubulo-interstitial diseases (N10-N16)

H- Renal failure (N17-N19)

I- Other viral diseases (B25-B34)

J- Congenital malformations and deformations urinary system (Q60-Q64)

K- Helminthiases (B65-B83)

L- Congenital malformations and deformations genital organs (Q50-Q56)

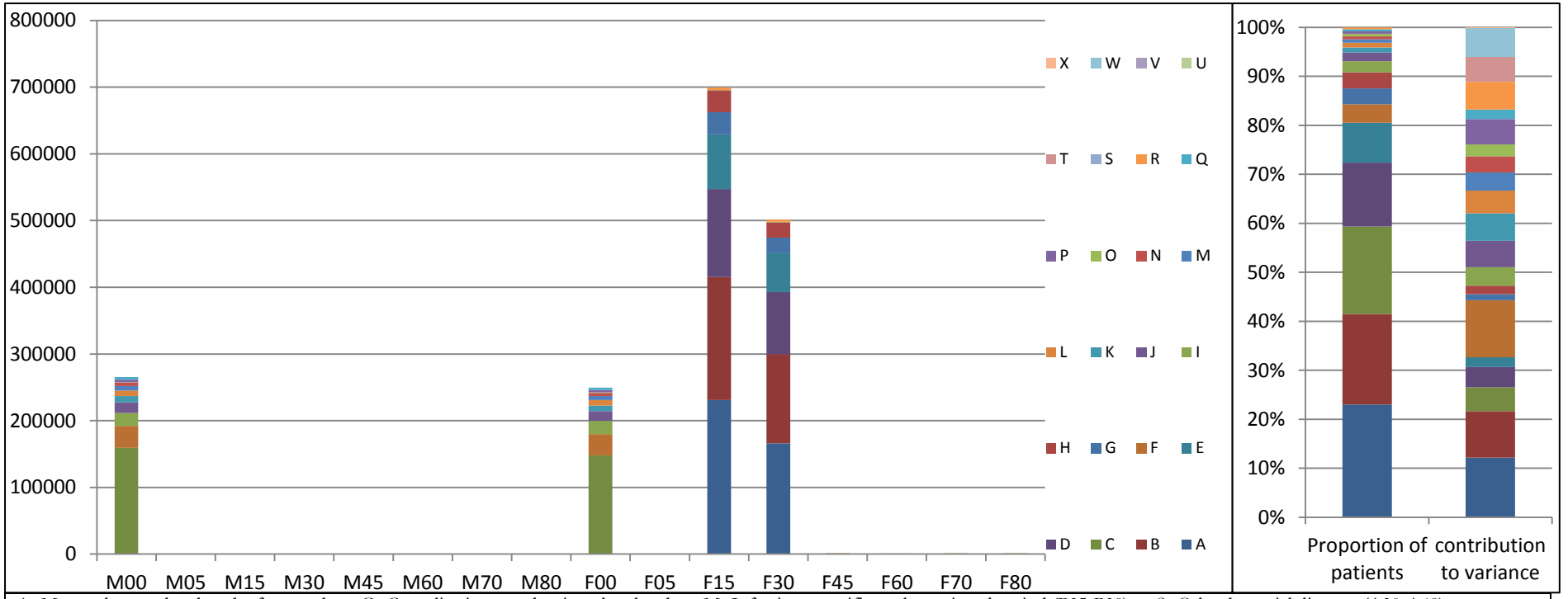
M- Symptoms and signs urinary system (R30-R39)

N- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)

O- Persons encountering health services for specific procedures and health care (Z40-Z54)

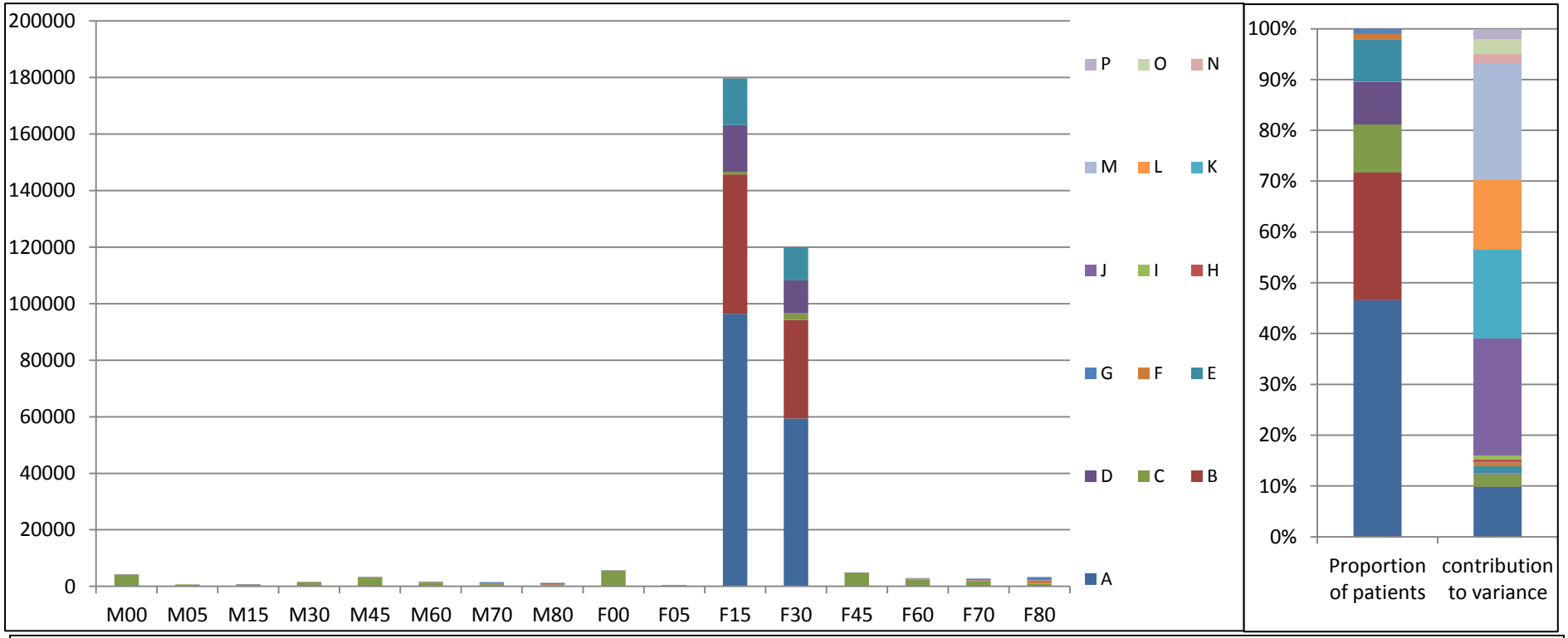
P- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

Table C1.17: Distribution of PBC18&19 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Maternal care related to the fetus and amniotic cavity and possible delivery problems (O30-O48)	G- Complications predominantly related to the puerperium (O85-O92)	M- Infections specific to the perinatal period (P35-P39)	S- Other bacterial diseases (A30-A49)
B- Complications of labour and delivery (O60-O75)	H- Other obstetric conditions, not elsewhere classified (O95-O99)	N- Transitory endocrine and metabolic disorders specific to fetus and newborn (P70-P74)	T- Pregnancy with abortive outcome (O00-O08)
C- Disorders related to length of gestation and fetal growth (P05-P08)	I- Birth trauma (P10-P15)	O- Other diseases of urinary system (N30-N39)	U- Other bacterial diseases (A30-A49)
D- Other maternal disorders predominantly related to pregnancy (O20-O29)	J- Haemorrhagic and haematological disorders of fetus and newborn (P50-P61)	P- Digestive system disorders of fetus and newborn (P75-P78)	V- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)
E- Delivery (O80-O84)	K- Other disorders originating in the perinatal period (P90-P96)	Q- Conditions involving the integument and temperature regulation of fetus and newborn (P80-P83)	W- Persons encountering health services in circumstances related to reproduction (Z30-Z39)
F- Respiratory and cardiovascular disorders specific to the perinatal period (P20-P29)	L- Fetus and newborn affected by maternal factors and by complications of pregnancy, labour and delivery (P00-P04)	R- Oedema, proteinuria and hypertensive disorders in pregnancy, childbirth and the puerperium (O10-O16)	X- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

Table C1.18: Distribution of PBC20 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Complications predominantly related to the puerperium (O85-O92)

B- Other obstetric conditions, not elsewhere classified (O95-O99)

C- Other diseases of the digestive system (K90-K93)

D- Other maternal disorders predominantly related to pregnancy (O20-O29)

E- Complications of labour and delivery (O60-O75)

F- Radiation-related disorders of the skin and subcutaneous tissue (L55-L59)

G- Disorders of skin appendages (L60-L75)

H- Congenital malformations and deformations Other (Q80-Q89)

I- Abnormal clinical and laboratory findings, not elsewhere classified (R70-R99)

J- Poisoning by drugs, medicaments and biological substances (T36-T50)

K- Toxic effects of substances chiefly nonmedicinal as to source (T51-T65)

L- Other and unspecified effects of external causes (T66-T78)

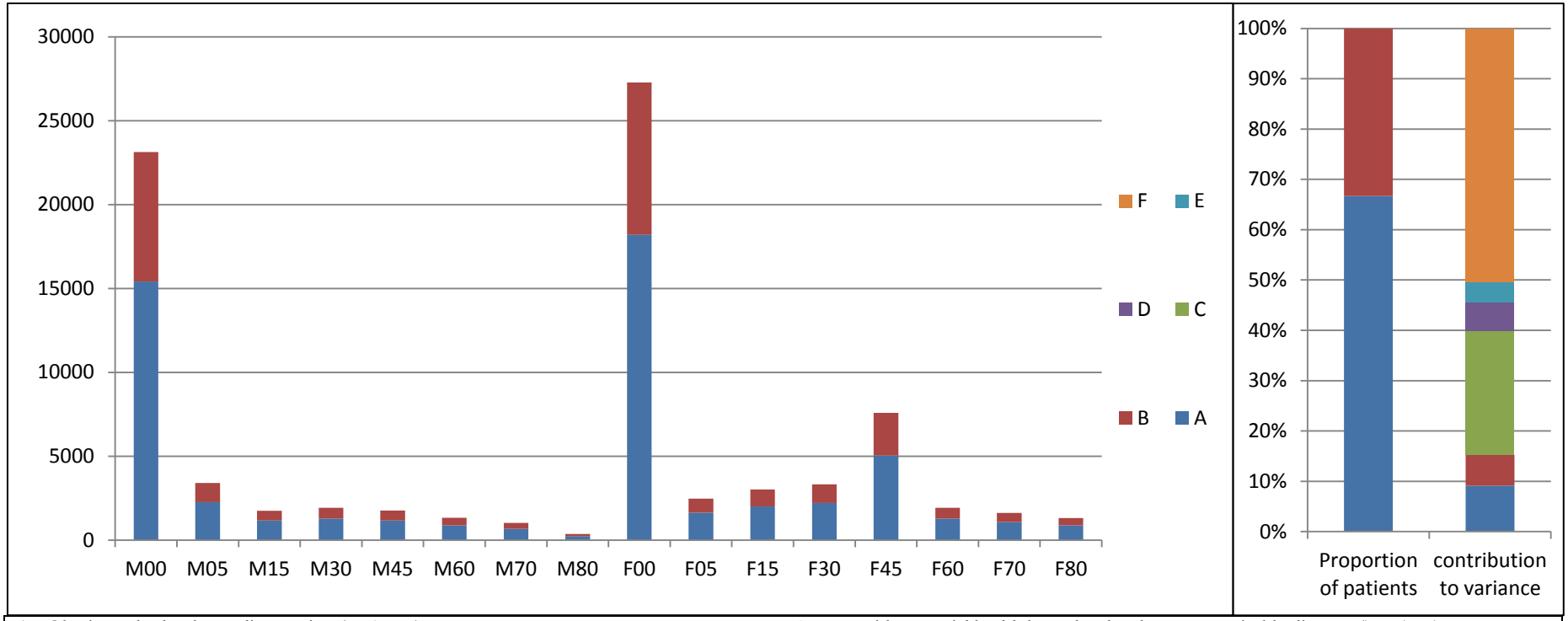
M- Complications of surgical and medical care, not elsewhere classified (T80-T88)

N- Persons with potential health hazards related to family and personal history and certain conditions influencing health status (Z80-Z99)

O- Pregnancy with abortive outcome (O00-O08)

P- Other diseases of urinary system (N30-N39)

Table C1.19: Distribution of PBC21 prevalence by age, gender and contributing ICD alongside proportion of prevalence patients and contribution to variance of each ICD



A- Obesity and other hyperalimentation (E65-E68)

B- Metabolic disorders of combinations (E76-E78)

C- Persons encountering health services for examination and investigation (Z00-Z13)

D- Persons with potential health hazards related to communicable diseases (Z20-Z29)

E- Persons encountering health services for specific procedures and health care (Z40-Z54)

F- Persons with potential health hazards related to socioeconomic and psychosocial circumstances (Z55-Z65)

Appendix C: Addendum 2

THE ROLE OF DATA ON LOCAL NHS DECISIONS

Contents

A. The role of local data in this study

The aim of this research project, noted in Chapter 1, is to develop and demonstrate methods for threshold estimation that make best use of routinely available NHS data. The principal focus of that methodological development, as reflected in the main body of this report, has been the use of econometric methods to exploit observed variations in spending and health outcomes between PCTs, at the programme budget level of aggregation.

However, we also aimed to investigate, as a complementary element of the project, the extent to which there may be other, more disaggregated sources of evidence on investment and disinvestment decisions made by local NHS organisations which might inform our analysis.

Specifically, we set out to (a) identify and evaluate what data might be routinely available from local NHS organisations with respect to their decisions to increase or decrease spending on specific services, and (b) consider whether and how such evidence might contribute directly toward the quantitative estimates of the threshold – for example, by providing more granular, contextual information on spending decisions that might assist in the interpretation of model estimates. For example, we wished to explore whether there were any routinely collected data from local NHS organisations that could tell us something about which ICDs within a given PBC might be the focus of investment and disinvestment.

The work which was undertaken was therefore focused on the potential use of local data *alongside* the econometric analysis – rather than their potential use as an alternative means of identifying the marginal cost of a QALY in the NHS[15].

B. Sources of publicly available data on PCT investment and disinvestment

To help us identify possible sources of data on NHS spending decisions, we began by consulting a number of experts within the NHS, identified for us by our collaborator, Professor David Parkin (Chief Economist at NHS South East Coast). These included Directors of Finance, Commissioning and Public Health. Those discussions helped direct us to a number of initiatives which involved the development of tools or evidence to inform resource allocation decisions, and helped to identify types and sources of documents published by PCTs that potentially contained relevant information on spending decisions. We then undertook a search for publicly available documents, in each case identifying what was available, and assessing its potential relevance for the purposes of this work outlined above. In evaluating each data source, the key considerations were:

(a) Whether the data were routinely collected: Routinely-collected data are preferred, as our overarching aim is to develop a set of methods to estimate the threshold, which can be readily updated from data routinely generated by the NHS.

(b) Whether the data were in the public domain: Published data are preferred to data that can only be obtained on request, because this would increase the cost and effort required in obtaining data from all relevant organisations.

(c) Whether the data were collected and reported in a systematic and consistent manner that would facilitate comparisons between PCTs, and with sufficient detail to enable us to link spending decisions to specific programme budgets or ICDs. This aspect of the work was undertaken during 2010.

The following were identified as potential sources of data:

B.1. Programme Budgeting tools – quadrant analysis – spend outcome tool (SPOT)

Data are available for three years, 06/07, 07/08 and 08/09[29] under the Spend Outcome Tool (SPOT) which is available to download. Expenditure data are organised by Programme Budget Category only,

[29] from http://www.yhpho.org.uk/resource/view.aspx?RID=49488

with no lower level of disaggregation. The data shows, for each PCT, the spend per head this year, the Z-score of that spend, and the PCTs national ranking based on their Z-score.

Outcomes data have also been captured, with different outcome measures within each PB category. Again, for each outcome there is a related Z-score and the PCTs national ranking based on that Z-score.

The tool enables users to see graphically how one PCT compares to others nationally, by SHA and by those PCTs similar to it by cluster (eg. other PCTs in manufacturing towns). The quadrant analysis tool has to origin as the mean PCT for that PB category, with Z-score for both expenditure and outcome equal to zero. The y-axis shows outcome, and the y-axis expenditure, both by Z-score.

While a useful tool, this source added little to the data already used in the econometric analysis, as it does not provide any additional information on the allocation of resources within PBs.

B.2. Lists of interventions not normally funded

Most PCTs provide information about interventions not normally funded. However, these were of limited usefulness because most of the procedures listed are those that might be expected (cosmetic surgery; tattoo removal, etc), and are not particularly informative about the marginal cost per QALY in the NHS. We did not find any information regarding whether any previously funded treatments had been added to these lists.

B.3. Special therapeutic and cancer committees

These are regionally based (not PCT or SHA) specialised committees that make decisions regarding spending on new cancer medicines and other special therapeutic areas. While such decisions would be potentially of direct relevance, we were unable to find any public documentation on their processes or decision outcomes.

B.4. Quality Innovation Productivity and Prevention (QIPP) published data on efficiency savings in the NHS

Introduced in 2009, QIPP addresses the quality and productivity challenge faced by the NHS. Developed by NICE, the Cochrane Quality and Productivity (QP) topics identify areas where resources could be significantly reduced or stopped completely without reducing the quality of NHS care, releasing cash and/or resources to other areas in the NHS. Each Cochrane topic has been established from systematic reviews undertaken by reviewers at the Cochrane Collaboration.

Every month the Cochrane Collaboration informs NICE as to new or existing Cochrane reviews where they have found that the existing treatment options(s) are harmful or ineffective and should not be used, or where evidence is unavailable or insufficient to support widespread use of that treatment in the NHS. NICE then completes an assessment of a Cochrane topic, to evaluate the efficiency savings that are likely against the QIPP criteria of likely ease of implementation, impact on productivity, and on the quality of care.

Savings per 100,000 patients are calculated, and then efficiency gains per PCT can be calculated. Once a topic has been accepted as best practise, users (PCTs) are encourage to submit their experience of implementing the changes, and the users achieving the best efficiency gains become QIPP examples of best practise.

The data shows which procedures are considered inefficient use of resources, although to the extent these are based on means of achieving the same or improved outcomes but with lower resources, will not be revealing of the marginal cost of producing a QALY in the NHS. Further, there is incomplete information about the extent to which PCTs actually implement these recommendations.

B.5. NHS Right Care

This website[30], has a section on the NHS Atlas of variation in health care, which seeks to reduce unwarranted variations in health care, defined as "…variation in the utilisation of health care services that cannot be explained by variation in patients or patients preferences", to increase value and improve quality.

It also provides a Third Annual Population Value Review which uses programme budgeting and marginal analysis to deliver QIPP. This provides, amongst other things, a 10-step, structured approach for PCTs to follow to establish where investment and disinvestment decisions could be made.

Further, it provides a tool for NHS Foundation Trusts to improve efficiency via Service Line Management.

While these tools may be being used by PCTs and Foundation Trusts, it was not clear to what extent that was the case, and there is no routine data on their use by NHS organisations or the decisions that resulted from that.

B.6. Health Investment Network – case studies of PBMA

The NHS network, Health Investment Network, was established to provide the access to the latest knowledge and tools to help commissioners optimise their investment and disinvestment decisions. It provides case studies of PCTs which have used PBMA to identify efficiency gains. This includes examples of 'spend to save' decisions e.g. where an initial investment (eg in vascular checks for men in deprived areas) could be more than outweighed by savings. Such initiatives, while important, are not useful in identifying the marginal cost per QALY in the NHS. Other case studies identify 'wish lists' (areas which PCTs prioritise for additional spending, should budgets expand) and 'hit lists' (services that might be reduced, to free up resources for more cost effective services). These case studies provide useful selected examples – but do not provide a routine or systematic reporting of such decisions across all PCTs.

B.7. Annual operating plans and strategic commissioning plans

PCTs are required to publish, each year, operating plans and strategic commissioning plans detailing their planning for the coming year, including information on the way that PCTs have made decisions concerning resource allocation. Because these reports are published annually, we considered that they constituted the most promising source of data, as they are produced routinely, and cover all PCTs.

Contact details and websites were identified for all 142 PCTs. Strategic Commissioning Plans were obtained for an initial 70 of these. These were used to identify any information provided about programmes of care or specific services where spending was planned to be increased or decreased. Those data were extracted and recorded into a spreadsheet, along with any relevant contextual information eg relating to the process by which the decision had been made.

Our review of the data from the first 70 of these showed that there was considerable variation between the documents in terms of the level of detail and specificity about the services which were the subject of changes in spending. In many cases, the services were described in terms of broad initiatives which might have related to multiple programme budgets and ICDs. There was also variation in, and occasionally a lack of clarity about, the way in which spending changes were described: in some cases these were described in terms of absolute changes in spending; in others, as net changes, once estimates of offsetting savings elsewhere had been taken into account; and in others it was not stated.

Given those concerns, the data were considered unlikely to be useful to complement the econometric analysis, and the research team decided not to proceed with further data extraction for the remaining PCTs.

[30] Available at http://www.rightcare.nhs.uk/atlas/index.html

C. Conclusions

The context within which this element of the work took place may be relevant to note. While the NHS was not subject to the budget cuts imposed on other areas of government activity in response to the financial crisis, the NHS was required to make substantial productivity improvements within its existing budgets. This gave rise to a number of initiatives in response to the 'productivity challenge' and, generally, heightened interest in the identification of ways to improve efficiency; potential areas for disinvestment; and areas for investment which were motivated by 'spend to save'. This may have made it more likely that we would observe disinvestment decisions. The NHS was also, during the course of this project, undergoing a period of restructuring. The transition from PCTs to clinical commissioning groups, and the disestablishment of strategic health authorities, may have had an effect on the availability of data and information relating to decision making. It may also have broader implications for the availability of data in the future, given the change in administrative units.

Our review of local data sources suggested that there is very little routinely collected data on investment and disinvestment by local NHS organisations beyond the high-level aggregate data on spending by PB which are used in the econometric analysis. More disaggregated data on spending decisions about specific services could, of course, be obtained by other means – for example, by surveying PCTs, or by requesting such information from them using a Freedom of Information request. However, that would impose data collection costs and would need to be designed carefully to ensure that such efforts yielded complete and consistent information.

Appendix C: Addendum 3

Characterisation of the investment and disinvestment decisions in mental health: depression and schizophrenia

Contents

A. Introduction

As has been highlighted in the main body of this project, it was not possible to produce an outcome equation for PBC 5: mental health problems, because no relevant mortality data was available from the NHS IC by PCT. Mental health represents a significant incidence and expenditure within the NHS. As a result we investigated the direction of bias from the exclusion of mental health problems on our estimate of the cost-effectiveness threshold. To understand this bias we examined current investment decisions in mental health. Recent investments in treatments with ICERs above the estimated threshold would suggest that not including PBC 5 more directly in our calculation may underestimated the threshold, conversely if recent investment has ICERs below the estimated threshold it would suggest that its exclusion results in an overestimated threshold. We focussed on depression and schizophrenia because of their high prevalence and contribution to variance.

B. Method employed

To evaluate the direction of bias of the exclusion of PBC 5 we followed four steps to make the connection from the identification of the most significant ICDs of PBC 5 to considering the cost-effectiveness of the investment and disinvestment decisions made in the NHS around these disease areas. The strategy was as follows:

Step 1:

- Identify the mental health ICD codes that are most influential and suitable on which to focus our analysis
    - Done from number of patients and contribution to variance calculations using HES.

Step 2:

- Determine the medications or treatments used in the NHS to treat each of the significant ICDs
    - There is likely to be a large cross-over in the use of treatments for mental health areas, for example antipsychotics and cognitive behavioural therapy (CBT) are both widely used.
    - We made use of the NHS Choices website coupled with clinical expertise for this identification.

Step 3:

- Identify the cost-effectiveness of the current treatments and medications used in the NHS.
    - This identification will be done from a range of sources including: published HTAs, published guidance, TUFTs, NHS EED and Medline searches.
    - This step relies heavily on the literature published, literature tends to cover historical activities many of which represent treatments of interest for this analysis. The case could be made that historical treatments that have not been evaluated have escaped evaluation due to their apparent cost-effectiveness, and are as such unlikely to be marginal activities.
    - Further difficulties arose in the identification of the relevant cost-effectiveness figure. Ideally it would represent the cost-per QALY relative to what would be performed if that activity was no longer available to the NHS.

Step 4:

- Connecting the available literature on the cost-effectiveness to recent investment and disinvestment decisions made in the NHS.

C. Results of analysis

C.1. Step 1: identification of relevant ICDs

We first rank ICDs by prevalence and contribution to variance. Prevalence is estimated from HES data. The contribution to variance is calculated as the variance in expenditure across PCTs for each ICD compared to the total variance in expenditure across PCTs for all ICDs within PBC 5. The most prevalent ICD was for depressive episode (F32) at 25.07% of all ICDs within PBC 5 (Table C3.1). The ICD with the greatest contribution to variance was for schizophrenia (F20) with 45.16% (Table C3.2).

Depression (F32) and (F33) and schizophrenia (F20) have been chosen as the focus of our evaluation as they represent two of the largest mental health ICDs in terms of proportion of patients as well as proportion of variance in expenditure, as shown in Tables C3.1 and C3.2 below. In addition they represent ICDs that involve interventions by the NHS that can be more clearly defined (in contrast to, for example, unspecified dementia and mental and behavioural disorders due to the use of alcohol[31]).

Table C3.1: table showing ranking of mental health ICDs by prevalence from HES

ICD	Description	% of Mental health prevalence	Contribution to variance
F10	Mental and behavioural disorders due to use of alcohol	27.84%	9.70%
F20	**Schizophrenia**	**10.01%**	**45.16%**
F32	**Depressive episode**	**9.96%**	**6.91%**
F31	Bipolar affective disorder	6.19%	6.38%
F41	Other anxiety disorders	4.92%	0.26%
F60	Specific personality disorders	4.33%	14.11%
F03	Unspecified dementia	3.93%	3.29%
F01	Vascular dementia	3.32%	1.58%
G30	Alzheimer disease	3.30%	0.84%
F33	**Recurrent depressive disorder**	**2.83%**	**3.68%**

Table C3.2: table showing ranking of mental health ICDs by contribution to variance

ICD	Description	% of Mental health prevalence	Contribution to variance
F20	**Schizophrenia**	**10.01%**	**45.16%**
F60	Specific personality disorders	4.33%	14.11%
F10	Mental and behavioural disorders due to use of alcohol	27.84%	9.70%
F32	**Depressive episode**	**9.96%**	**6.91%**
F31	Bipolar affective disorder	6.19%	6.38%
F33	**Recurrent depressive disorder**	**2.83%**	**3.68%**
F03	Unspecified dementia	3.93%	3.29%
F01	Vascular dementia	3.32%	1.58%
G30	Alzheimer disease	3.30%	0.84%
F41	Other anxiety disorders	4.92%	0.26%

[31] This contrast was informed by our clinical representative

C.2. Step 2: determination of treatment employed

Table C3.3 provided an overview of the main treatments for depression and schizophrenia. This list of treatments was identified using the NHS Choices website[32] as well as discussion with our clinical representative for each of the respective illnesses. This list was used to inform a literature search of cost-effectiveness publications.

Table C3.3: table showing treatments for schizophrenia and depression in the NHS

ICD	Disease	Treatments
F20	Schizophrenia	1. Typical antipsychotics 2. Atypical antipsychotics 3. Cognitive behavioural therapy (CBT) 4. Crisis resolution teams (CRT)
F32 & F33	Depressive episode & recurrent depressive episode	1. Cognitive behavioural therapy (CBT) 2. Interpersonal therapy (IPT) 3. Selective serotonin reuptake inhibitors (SSRIs) .Serotonin–norepinephrine reuptake inhibitors (SNRIs) 5. Tricyclic antidepressants (TCAs) 6. Monoamine oxidase inhibitors (MAOIs) 7. Lithium 8. Electro-convulsive therapy (ECT)

C.3. Step 3: evaluation of the relevant cost-effectiveness literature

Using the treatment categories identified in step 2 of this work a systematic search was conducted to attempt to identify the range of literature on the cost-effectiveness of current NHS treatment of schizophrenia and depression. For both illnesses five online databases were searched: the Cost-Effectiveness Analysis (CEA) Registry of the Tufts Medical Centre, the NHS Economics Evaluation Database run by the Centre for Reviews and Dissemination (CRD) at the University of York, Medline, the NICE online database of Technical Appraisals (TA) and Clinical Guidelines (CG), as well as NIHR's Health Technology Assessments (HTA). All searches were conducted on the 19th October 2011.

The search strategies employed to search for relevant cost-effectiveness literature and details of the results can be found in the search strategy section at the end of this addendum. For both schizophrenia and depression five sources of information were searched sequentially: the Cost-Effectiveness Analysis (CEA) Registry, the NHS Economic Evaluation Database (EED), Medline, NICE's online library of Technical Appraisals (TA) and Clinical Guidelines (CG), and finally NIHR's online library of Health Technology Assessments (HTAs). For schizophrenia this approach identified 61 unique publications, five of which were deemed to be of broad relevance to this analysis. For depression 65 publications were discovered, ten of which were relevant. A paper of relevance to our analysis of mental health was deemed to be so if it presented cost-effectiveness results (in the form of a cost per QALY ICER) of a comparison of at least two of the treatments for either schizophrenia or depression identified in section C.2. These results could be from a de-novo analysis or from a systematic review of the relevant literature.

Table C3.4 reports the cost-effectiveness results of antipsychotics for schizophrenia as first line treatments. The NICE clinical guidelines for schizophrenia (CG82)[16] demonstrate that the differences in costs and effects of the 1st and 2nd generation treatments described are very similar with ICERs comparing each to no treatment ranging from £21,517 to £23,237 per QALY. Comparisons to active treatments result in ICERs of £5,156 to £33,240 per QALY[17, 18].

[32] Available at http://www.nhs.uk/Pages/HomePage.aspx accessed on 10/10/2011

Table C3.4: table showing cost-effectiveness studies of antipsychotics for schizophrenia

Study	Treatment	Comparator	Cost (£)	QALYs	ICER (£/QALY)
NICE CG82[16]	Zotepine (2nd)	No treatment	139,170	6.468	21,517
	Paliperidone (2nd)	No treatment	142,173	6.427	22,121
	Olanzapine (2nd)	No treatment	141,212	6.42	21,996
	Risperidone (2nd)	No treatment	149,112	6.417	23,237
	Haloperidol (1st)	No treatment	143,406	6.413	22,362
	Aripiprazole (2nd)	No treatment	145,697	6.4	22,765
	Amisulpride (2nd)	No treatment	147,920	6.392	23,141
Knapp et al 2008 [17]	Olanzapine (2nd)	other antipsychotics			5,156
Davies et al 2008 [18]	Clozapine (2nd)	Other 2nd gen antipsychotics			33,240
	Aripiprazole then risperidone	risperidone then olanzapine			9,440

The CG82 results are similar to the first line treatment results from Bagnall et al.[19], shown below in Table C3.5. The cost-effectiveness of antipsychotics compared to no treatment as second, third or final therapy are less than £20,000 per QALY.

Table C3.5: table showing cost-effectiveness studies of antipsychotics for schizophrenia

Bagnall et al., 2003 [19]				
ICER (£/QALY)	**Line of treatment**			
Antipsychotic	1st	2nd	3rd	Final
Chlorpromazine (1st)	21,989	15,185	15,419	15,303
Haloperidol (1st)	24,069	17,177	17,211	17,022
Clozapine (2nd)	24,500	17,595	17,577	17,402
Olanzapine (2nd)	25,719	18,869	18,808	18,865
Quetiapine (2nd)	26,316	19,090	18,751	19,096
Zotepine (2nd)	22,769	16,350	16,360	16,400
Risperidone (2nd)	22,255	15,596	15,599	15,700
Ziprasidone (2nd)	21,935	15,192	15,191	15,224
Amisulpride (2nd)	23,174	15,941	15,945	15,962
Sertindole (atypical)	23,181	16,297	16,308	16,354

Only one study reported the cost-effectiveness of a psychological or social intervention for schizophrenia. Barton et al. [20] conducted a randomised trial to estimate the clinical and cost-effectiveness of social recovery orientated cognitive behavioural therapy (SRCBT) against case management alone for people recently diagnosed with psychosis. SRCBT consisted of three stages of social recovery combined with CBT techniques including vocational case management. SRCBT was found to have an ICER of £18, 844 per QALY compared to case management. However, it is not clear that all forms of CBT are well represented by this one study or that these results relate well to schizophrenia since this study was for the use of social recovery CBT for psychosis disorders in general.

Table C3.6: table showing cost-effectiveness of psychological/social interventions for schizophrenia

Study	Treatment	Comparator	ICER (£/QALY)
Barton et al. 2009 [20]	SRCBT	case management	18,844

Table C3.7 reports the cost-effectiveness results of publications identified in the systematic search of drug treatments for depression in the NHS. As was highlighted in table C3.3 a range of drug treatments are available for depression, broadly falling into five categories: selective serotonin reuptake inhibitors (SSRIs), Serotonin–norepinephrine reuptake inhibitors (SNRIs), tricyclic antidepressants (TCAs), monoamine oxidase inhibitors (MAOIs) and lithium.

The NICE guideline CG90 tested the cost-effectiveness of numerous treatments for moderate and severe depression. It was found that across all the treatments tested the mean QALYs for moderate depression had a range of 0.053 and severe depression had a range of 0.065. The costs had a range of £408 for moderate and £396 for severe depression. The results suggest that mirtazapine has the lowest ICER for both moderate and severe depression. If mirtazapine is not a suitable treatment option than escitalopram or sertraline is preferred because escitalopram dominates venlafaxine and sertraline dominates the remaining antidepressants. The ICERs of escitalopram versus sertraline are £32,987 per QALY for moderate depression and £27,172 per QALY for severe. The authors thus suggest that according to these results escitalopram should be considered when mirtazapine and sertraline are not suitable. Other ICERs reported in CG90 can be found in Table C3.7. CG90 states that the economic evidence had limitations and these comparisons were considered insufficient to make specific recommendations for treatments.

ICERs in other studies range from £2,172 - £20,600 per QALY, with TCA being dominated by Lofepramine (TCA) in two cases and fluoxetine being dominated by amitriptyline (TCA).

Table C3.7: table showing cost-effectiveness of drug treatments for depression

Study	Treatment	comparator	incremental cost	incremental QALY	ICER (£/QALY)
NICE CG 90 [21]	Combined SSRI and CBT (severe depression)	SSRI alone			£5,558
	Duloxetine (SNRI)	SSRI			£6,300
	Duloxetine (SNRI)	Mirtazapine (TCA)			£2,400
	Duloxetine (SNRI)	Venlafaxine (SNRI)			dominates
	escitalopram (moderate depression) (SSRI)	Sertraline (SSRI)			£32,987
	escitalopram (severe depression) (SSRI)	Sertraline (SSRI)			£27,172
Lenox-Smith et al. 2009[22]	Venlafaxine (major depression) (SNRI)	SSRI			£20 600
	Fluoxetine (SSRI)	Amitriptyline (TCA)			dominated
Kendrick et al. 2006[23]	SSRI	TCA			£2,692
	TCA	Lofepramine (TCA)			dominated
	SSRI	Lofepramine (TCA)			£5,686
Hatziandreu et al. 1994[24]	Sertraline (SSRI)	Dothiepin (TCA)			£2,172
Peveler, 2005[25]	SSRI	Lofepramine (TCA)	0.035	£199	£5,686
	TCA	Lofepramine (TCA)	–0.004	£93	dominated
	SSRI	TCA	0.039	£105	£2,692
Kendrick, 2009[26]	SSRI + SC	SC			14,854

Table C3.8 provides the results of the combination therapies for moderate and severe depression presented in CG90 [21]and Simon et al.[27]. These studies considered the impact of combined SSRI and CBT versus SSRI alone. Both of these studies find combined CBT and antidepressant to have ICERs of less than £8,000 per QALY.

In addition table C3.8 provides results of analyses of computerised CBT (CCBT) compared to treatment as usual or relaxation. The results generally find CBT and CCBT to be highly cost-effective, with the exception of BT Steps[28] all ICERs are found to be under £18,000 per QALY.

Table C3.8: table showing cost-effectiveness of psychological and social intervention for depression

Study	Treatment	comparator	incremental cost	incremental QALY	ICER (£/QALY)
NICE CG 90[21]	Combined SSRI and CBT (moderate depression)	SSRI alone			£7,052
	Combined SSRI and CBT (severe depression)	SSRI alone			£5,558
Simon et al. 2006[27]	CBT + antidepressants in severe dep	fluoxetine			£5,777
	CBT + antidepressants in moderate dep	fluoxetine			£14,540
Kaltenthaler et al., 2002 [29]	Beat the Blues CCBT (BtB)	Treatment as usual (TAU)			£1,209 to £7,692
Kaltenthaler et al., 2006 [28]	BtB	TAU	£147	0.08	£1,801
	Cope CCBT	TAU	£193	0.03	£7,139
	Overcoming Depression CCBT	TAU	£64	0.01	£5,391
	FearFighter CCBT	Relaxation CBT	£138	0.058	£2,380
	Therapist lead CBT	Relaxation	£194	0.011	£17,604
	BT Steps CCBT	Relaxation	£360	-0.01	Dominated
Hollinghurst et al., 2010 [30]	Online CCBT	TAU			£17,173

C.4. Step 4: connection to investment and disinvestment decisions

In this section we discuss the investment and disinvestment decisions made considering the cost-effectiveness information in the previous section. If we believe that decision makers will invest in treatments below their cost-effectiveness threshold and disinvest in treatments above this threshold then by considering the ICERs of treatments subject to investment and disinvestment we can create a range for their cost-effectiveness threshold. This approach and its role in the consideration of a cost-effectiveness threshold has been previously discussed by Appleby et al. [31]. With a view of the cost-effectiveness threshold within PBC 5 we consider how its exclusion from our calculation of the threshold might influence our results.

To identify the broad areas of investment in the disease areas we make use of recent NICE guidance documents. While NICE clinical guidance does not definitively represent observed shifts in practice and are often not well implemented in mental health trusts[32] it can help to inform our evaluation. NICE guidance does not identify areas where disinvestment should occur within a disease; as a result we have consulted experts in the respective fields to gain an understanding of any significant recent disinvestment decisions. For schizophrenia we were provided expert opinion by Professor Tim Kendall (Centre for Psychological Services Research, University of Sheffield) and for depression by Professor Simon Gilbody (Health Sciences, University of York).

For both schizophrenia and depression we will briefly discuss the areas of investment and disinvestment in two care categories: (i) drug treatments, and (ii) psychological and social interventions.

C.4.1. Results of step 4 for schizophrenia

C.4.1.1. Analysis of drug treatment

Antipsychotics used for the treatment of schizophrenia can be broadly identified as first or second generational (typical and atypical antipsychotics). To a certain extent there is still a debate over the relative strengths and weaknesses of each[33], and the significance of the adverse events associated with the second generation may still not be fully understood (such as the impact on new-onset type-2 diabetes [34]). However, our clinical experts indicated that clinicians were largely concerned with the adverse effects associated with the second generational drugs, and the increasing evidence questioning the relative efficacy, such as Rosenheck [33] who suggests that the first generational drugs in many cases are just as effective. Recent NICE guidance leaves the choice of first or second generational drugs to the clinician to decide[16].

When considering the impact on our estimate of the threshold of the possible shift to first generation from second generation antipsychotics we must attempt to generalise about the relative cost-effectiveness of the two. Clearly this is difficult as each generation represents many different drugs. However, from CG82 the costs and benefits of the mainstream antipsychotics are broadly similar Table C3.4. This would suggest that a shift away from the second generation back towards the first would have little impact on the overall threshold as the costs and benefits associated with each are very similar.

Olanzapine came off patent in the third quarter of 2011[33]. Olanzapine and similar second generation antipsychotics are associated with a cost of around £30million a year[34], clearly the introduction of generics to the market would significantly reduce this cost and thus increases the cost-effectiveness of these drugs. While this shift does not fall within the years of our analysis, it will have a significant impact on the future value of the cost-effectiveness threshold.

The other significant area of debate, as identified by our clinician, is the role of the antipsychotic clozapine, which has often been viewed as the most effective antipsychotic drug for schizophrenia however has been connected with some severe adverse events (such as myocarditis[35], agranulocytosis[36] and central nervous system depression). This has lead to the NICE guidelines advising clozapine only if an array of other antipsychotics has failed[16]. While clozapine is highly clinically effective it is associated with a higher overall cost (a significant proportion due to the associated adverse events). As is shown in table C3.4 Davies et al. [18] show clozapine to have an ICER of £33,240 when compared to other second generation antipsychotics. Disinvestment of clozapine suggests that the threshold is lower than £33,240 per QALY. However, current investment in other 1st line antipsychotics suggests that the threshold in mental health is over £23,237 per QALY.

C.4.1.2. Analysis of psychological and social intervention

In this section we discuss all non-drug related interventions for schizophrenia. The NICE guidelines[16] outline the provision of CBT, arts therapy and family interventions to treat schizophrenia, however, efficacy of these interventions is disputed[37] and little is known about their cost-effectiveness. The systematic review only yielded one paper that was relevant to our analysis, as is shown in Table C3.6. The Barton et al.[20] study found that SRCBT had an ICER of less than £20,000 per QALY relative to case management. However, as mentioned previously, this study may not represent all forms of CBT or schizophrenia.

Our clinical advisors informed us that CBT provision varies significantly across PCTs and therefore represents an intervention likely to be subject to investment and disinvestment at the margin. The variation in CBT provision (and indeed other psychosocial/social interventions) is largely a result of the poor support for its efficacy and significant initial cost.

Other interventions of relevance to this investigation include art therapies and family interventions. As with CBT there is a significant variation in the provision of family interventions. No information on its cost-effectiveness was found from our search. Art therapies include: music therapy, art therapy, and body

[33] See: http://www.dispensingdoctor.org/content.php?id=1335 accessed 03/05/2012
[34] Estimate by Tim Kendal

movement or dance therapy. Our clinical advisors have highlighted increasing investigations into arts therapy, including the "Matisse trial"[38], publications around which have shown that art therapy as adjunctive therapy had little benefit over a comparator activity or treatment as usual[37]. No information on its cost-effectiveness was found from our search.

Early interventions in schizophrenia, which aims to identify and treat early symptoms associated with schizophrenia, have been a significant area of investment over recent years in the NHS. While we were unable to identify any relevant cost-effectiveness literature around early interventions in schizophrenia it is generally expected that these represent cost-effective interventions over the long term[35].

While the lack of cost-effectiveness literature clearly limits the potential to directly associate these interventions with the wider cost-effectiveness threshold it is widely accepted that many social interventions for schizophrenia (specifically around CBT and family interventions) are cost-saving for the NHS[36], as they reduce hospitalisation by reducing emergency hospital access and relapse rates that are high in schizophrenia representing the majority of related hospitalisations[39].

Investment in CBT with an expected ICER of £18,844 per QALY suggests that the threshold for mental health treatments is above this value.

C.4.2. Results of step 4 for depression

As table C3.3 shows depression is associated with a wider range of treatment than schizophrenia, specifically a wider range of drug treatments are available. As with the schizophrenia section of this addendum we will deal with the treatments under the two categories of drug treatments and psychological/ social interventions. Electroconvulsive therapy (ECT), which is included in the treatment options available in the NHS as shown in table C3.3, is excluded from this analysis based on expert opinion on the grounds of it being a very rarely used but extreme treatment that is not likely to be further subject to substantial investment or disinvestment, so is not relevant for our analysis.

Recent NICE clinical guidance[21] highlights a range of key priorities for implementation. As with schizophrenia there is no guarantee that these are the areas of investment in depression care but it represents a suitable outline of the areas of interest. Several areas are highlighted:

- Early identification and diagnosis
- Low intensity psychological interventions (CBT, CCBT and group physical activity) for persistent sub-threshold depressive symptoms or mild to moderate depression
- Reduced routine use of antidepressant for sub-threshold depressive symptoms or mild depression
- Combination therapies (antidepressant and psychological) for moderate or severe depression
- Extension of therapy (antidepressant and psychological) beyond remission to reduce relapse
- SSRIs are presented as the preferred type of antidepressant due to their equivalent efficacy and favourable risk-benefit ratio.

These are the areas of investment that our analysis will focus on.

C.4.2.1. Analysis of drug treatment for depression

Our clinical advisors reported that the current area of activity in antidepressants is the creation of drugs such as escitalopram (an SSRI) and venlafaxine (an SNRI) that are relatively similar to generic treatments currently in the market. As these new drugs are covered by patents they are relatively expensive. Table C3.7 reports the results on the cost effectiveness of these two drugs from NICE CG90[21] as well as Lenox-Smith et al.[22]. In both reports the drugs are compared to alternative SSRIs in moderate and severe depression. In both cases the newer SSRIs were approved by NICE with an ICER for escitalopram of £32,928 per QALY and for venlafaxine of £20,600 per QALY. If mirtazapine and

[35] This view was informed by our clinical advisors
[36] This view was informed by our clinical advisors

sertraline are not suitable then the ICER of escitalopram for moderate depression is £5,357 per QALY compared to citalopram. While evidence was not available on whether clinicians were making use of these newer SSRIs, an investment in them away from alternative SSRIs may represent an increase in the cost-effectiveness threshold due to the relatively high ICERs reported in the two studies. However, the cost-effectiveness of each depends on what they displace and ICERs may be lower if the more cost-effective treatments have failed.

Investment decisions in the NHS for antidepressants are likely to represent changes in the type of antidepressant being prescribed rather than a shift from no treatment to treatment. The majority of trials discovered by systematic review given in table C3.7 show that while the ICER of SSRIs versus TCAs is very low [25] this is largely driven by very small gains in QALYs but for a similarly small increase in cost. As a result any observed investment in SSRIs away from TCAs is likely to lead to a small decrease in an observed threshold for the NHS.

C.4.2.2. Analysis of psychological and social intervention for depression

The NICE guidelines reported in CG90 place a lot of focus on the provision of psychological interventions such as CBT (and CCBT) over antipsychotics wherever possible. Table C3.8 provides the results of the combination therapies for moderate and severe depression presented in CG90 [21]and Simon et al. [27]. These considered the impact of combined SSRI and CBT versus SSRI alone and concluded that combined therapies in both populations had ICERs of less than £15,000 per QALY. According to our clinicians, this is an area that is likely to have had significant investment in recent years.

The two HTAs reported in table C3.8 [28, 29] provide a good analysis of the cost-effectiveness of computer cognitive behavioural therapy (CCBT) versus treatment as usual (TAU). They show that the CCBTs investigated have ICER of less than £8,000 per QALY relative to TAU. Further analyses investigated different kinds of CCBT and found that compared to relaxation CBT ICERs ranged from £2,380 per QALY to dominated. Hollinghurst et al. report that two CBT interventions compared to TAU had ICERs of £17,173.

As NICE guidelines encourage the use of CBT and our clinical experts believe this has been an area of increased investment, this review suggests that the threshold in mental health is over £17,173 per QALY.

D. **Conclusion**

There is very little accessible data on the investment and disinvestment decisions in specific areas of mental health and so we relied on the opinions of clinical experts. The NHS Information Centre has some information on prescriptions of mental health treatments, however it was not clear for which diseases these treatments were being used or for which line of therapy. As a result this data was not included in our analysis as it was decided it may not represent the investment and disinvestment decisions that we were seeking to identify.

Most treatments reviewed had an ICER of less than £24,000 per QALY. Two treatments had higher ICERs. Clozapine for the first line treatment of schizophrenia was found to have an ICER of £33,240 per QALY compared to other 2nd generation antipsychotics. NICE's recommendation to use clozapine only as a last line treatment suggests that the threshold is less than £33,240 per QALY. Escitalopram for moderate depression has been recommended by NICE and was reported to have an ICER of £32,987 per QALY compared to seratraline. Conclusions on the threshold from this finding are unclear. The cost-effectiveness of escitalopram in the NHS will depend on its use. If it is used rather than seratraline then the threshold may be over £32,987, but if it is used as third line therapy than according to CG90 its use is less costly and more effective than the next best options.

How well the actual threshold reflects the ICERs reported above depends on how well clinical practice matches the clinical guidelines i.e. whether the more cost-effective treatments are being used first.

Search Strategies

Search Strategy for schizophrenia

CEA Registry search:

Six keywords associated with the entire Schizophrenia, schizotypal and delusional disorders ICD10 sub-chapter were search for in the CEA Registry, these were: schizophrenia, schizotypal, delusional, psychotic, schizoaffective and psychosis. A search for any of these keywords in the Registry yielded 18 different papers at the time of searching, with four of these being deemed suitable for our investigation (Barton 2009, Davies 2008, Jarbrink 2009, Knapp 2008 and Davies 2007).

NHS EED search:

A single relatively simple search strategy was defined to investigate NHS EED, this was as follows:
((Schizophrenia) AND (cost effectiveness):TI) and Economic evaluation:ZDT and Abstract:ZPS)
This result strategy yielded 28 hits, only one of which was both relevant to our search and not discovered in the CEA Registry search (Rosenheck 2007).

Medline search:

Medline was searched using the strategy:

cost benefit analysis and (schizophrenia or schizotypal personality disorder or delusions) and Great britain(MeSH)

This strategy yielded 13 hits, none of which were both relevant and had not been previously identified through the CEA Registry of NHS EED searches.

NICE Technical Appraisals (TA) and Clinical Guidelines (CG):

NICE's online database of published mental health related TAs and CGs

(http://www.nice.org.uk/guidance/index.jsp?action=byTopic&o=7281) was searched for schizophrenia related publications. Only one was found to fulfil our criteria for schizophrenia: CG82.

NIHR's HTAs:

Finally NIHR's database of published HTAs was searched. This activity discovered one additional relevant publication: HTA 00/20/01 - Bagnall, 2003.

Search Strategy for Depression

CEA Registry:

Two keywords were searched on the CEA Registry, they were: depression and depressive. These keywords yielded 17 papers, 5 of which were deemed relevant for our purposes (Hollinghurst 2010, Lenox-Smith 2009, Kendrick 2006, Simon 2006, and Hatzinandreu 1994).

NHS EED search:

A search similar in structure to the search for schizophrenia papers was conducted in NHS EED:
((depressive OR depression):TI AND (cost-effectiveness):TI) and Economic evaluation:ZDT and Abstract:ZPS) IN NHSEED

This yielded 43 hits, none of which were both relevant and previously undiscovered by the CEA registry search. Due to the complete nature of the CEA Registry and NHS EED searches as well as time constraints on the systematic review, a Medline search was not conducted as it was decided it would not provide sufficient added value.

NICE Technical Appraisals (TA) and Clinical Guidelines (CG):

Searching the NICE database of TAs and CGs yielded one publication deemed relevant to the analysis: CG90- depression in adults.

NIHR's HTAs:

A search of the NIHR's online database of published HTAs yielded four relevant publications:

HTA 01/23/01- Bennett, 2000

HTA04/01/01- Kaltenthaler, 2006

HTA- 96/61/11- Peveler, 2005

HTA- 01/70/05- Kendrick, 2009

Appendix C: Addendum 4

What type of health is forgone by the approval of a new technology?

The methods of analysis described in this work can identify not only how many QALYs are likely to be forgone across the NHS as a consequence of approving a technology which imposes incremental costs on the NHS, it can also indicate where those QALYs are likely to be forgone and how they are made up, i.e., the additional deaths, life years lost (unadjusted and adjusted for quality of life) and the quality of life impacts on those with disease. Based on the 2008 central estimate of the cost per QALY threshold, we will exemplify within this Addendum the likely health displaced elsewhere in the NHS as a consequence of approving a new technology.

The example of ranibizumab for diabetic macular oedema

In 2011, NICE considered whether ranibizumab for the treatment of diabetic macular oedema in patients with central retinal thickness ≥400 micrometres should be approved for widespread use in the NHS (TA237[40]). Initially this technology was rejected by NCE on the grounds that, at its current price, it would be unlikely to be cost effective. In 2012, however, a rapid review of TA237 [41] approved Ranibizumab if use was restricted to the most cost effective sub group (those with central retinal thickness ≥400 micrometres) and after a Patient Access Scheme (PAS) for this subgroup of patients was offered (details of the PAS which provides a discounts to the NHS is commercial in confidence). The Committee concluded that the most plausible ICER for the subgroup of people with thicker retinas was likely to be higher than the manufacturer's estimate (of £13,322 per QALY), but would be under £25,000 per QALY gained.[41]

The appraisal and guidance documents (http://guidance.nice.org.uk/TA/Wave23/41) provide the information required to estimate the additional NHS costs of treating this sub group of patients each year. The original manufacturer submission presented an estimate of the numbers of patients in the NHS eligible to receive ranibizumab, based on its licensed indication[42]. These estimates are presented in Table C4.1. In the first year of implementation, up to 44,000 NHS patients would be eligible for treatment with ranibizumab based on its licensed indication. No consideration is made as to the size of the sub-population approved for treatment, however the RESTORE trial (that informs the submission) found approximately half of the participants in the study to be in this sub-population [114 of 217 (52.5%)].[40] The subgroup of patients where ranibizumab was ultimately approved is thus likely to be approximately 23,000 in the first year after approval.

Table C4.1 Estimated size of the NHS population eligible for ranibizumab [42]

	2011	2012	2013	2014	2015
Licensed indication					
Prevalent cases	43,847	0	0	0	0
Incident cases		5,481	5,481	5,481	5,481
Total eligible number of patients	43,847	49,328	54,809	60,290	65,771
Sub-population approved for treatment by NICE					
Prevalent cases	23,020	0	0	0	0
Incident cases	0	2,878	2,878	2,878	2,878
Total eligible number of patients	23,020	25,897	28,775	31,652	34,530

The incremental costs associated with the new treatment (compared to laser monotherapy) in the initial submission (TA237) were £3,506 per patient[42]. Given estimates reported in the rapid review are not available (commercial in confidence), we will use this estimate of incremental costs for the subpopulation of interest. These data suggests that the approval of ranibizumab in this subgroup at the original appraisal in 2011 (i.e., without a PAS) would impose just over £80m of additional NHS costs for treating the eligible population each year.

Table C4.2 Estimated total budget impact of ranibizumab

	2011	2012	2013	2014	2015
Total eligible number of patients	23,020	25,897	28,775	31,652	34,530
Total cost, without PAS (£)	£80,708,120	£90,794,882	£100,885,150	£110,971,912	£121,062,180
Total cost, 30% lower incremental costs(£)	£56,495,684	£63,556,417	£70,619,605	£77,680,338	£84,743,526

With introduction of the PAS, it is likely that a simple discount on the acquisition price of the new technology has been approved by the DH.[41] Given the scale of the discount is not available (commercial in confidence) we assumed that this discount would reduce incremental costs by 30% (to £2,454 per patient). After such a PAS, the approval of ranibizumab in this subgroup would impose just £56m (rather than £80m) of additional NHS costs for treating the eligible population in the first year.

What type of health is forgone by approval of a new technology?
Based on the 2008 central estimate of the cost per QALY threshold (£18,317 in Table 5.1) the approval of ranibizumab without a PAS would have been likely to displace 4,367 QALYs elsewhere in the NHS. However, the analysis which underpins the threshold estimate can also be used to identify where the additional NHS cost of £80m are likely to impact and where and what type of health effects are likely to be forgone. This is illustrated in Table C4.3.

Table C4.3: Heath forgone across PBCs due to the approval of ranibizumab (£80m budget impact)

PBC	PBC description	change in spend (m) [1]	Additional deaths [2]	Life years foregone [3]	QALYs foregone Total QALYs forgone [4]	QALYs foregone Due to premature death [5]	QALYs foregone Quality of life effects [6]
2	Cancer	£2.59	22	217	153	141	11
10	Circulatory problems	£4.40	132	672	625	427	198
11	Respiratory problems	£2.66	78	93	1,330	58	1,272
13	Gastro-intestinal	£1.86	15	143	255	94	161
	Big 4	£12	246	1,126	2,362	721	1,641
1	Infectious diseases	£1.89	4	31	91	21	70
4	Endocrine problems	£1.10	4	29	351	19	332
7	Neurological problems	£3.47	7	38	632	25	608
17	Genito-urinary problems	£2.69	13	19	61	12	49
16	Trauma & injuries	£4.46	0	0	0	0	0
18+19	Maternity & neonates	£3.96	0	2	1	1	0
	11 PBCs	£29	275	1,245	3,500	798	2,701
3	Disorders of Blood	£2.33	1	6	82	4	78
5	Mental Health Disorders	£20.25	12	55	406	35	371
6	Learning Disability	£1.18	1	4	15	3	12
8	Problems of Vision	£2.20	0	2	29	1	28
9	Problems of Hearing	£0.99	0	1	52	0	52
12	Dental problems	£3.27	0	0	59	0	59
14	Skin	£2.23	2	7	13	5	8
15	Musculo skeletal system	£4.11	3	15	203	10	193
20	Poisoning and AE	£1.05	0	2	6	1	5
21	Healthy Individuals	£4.01	0	1	3	0	2
22	Social Care Needs	£3.41	0	0	0	0	0
23	Other	£5.88	0	0	0	0	0
	All (23 PBCs)	**£80**	**295**	**1337**	**4367**	**859**	**3509**

How the additional NHS cost of £80m will tend to affect spending in each of the 23 PBCs (see column 1) will be based on the estimated expenditure elasticities and total PBC expenditure (see Table C.55 in Appendix C).[37] In calculations of the threshold, the inputs above (expenditure elasticities and total expenditure) allow predicting how a 1% change in total spend is distributed by PBC. The same rationale is used here to establish how the additional NHS cost of £80m will affect each PBC. Hence, changes in spend reported here will be proportional to changes in spend across PBCs evaluated in calculations of the threshold (as in Table C.76 in Appendix C).

The estimated outcome elasticities (Table C.55 in Appendix C) allow the absolute changes in spend in each PBC described above to be translated into a change in deaths and life year effects for the 11PBCs where mortality effects

[37] The independently estimated expenditure elasticities for all 23 PBC do not account for all of a change in overall spend. The remaining change in total spend was assigned to the group PBCs where mortality effects could not be estimated. This will tend to overestimate the effect on spend in these PBCs and underestimate the effects on spending in the 11 PBC where mortality effects could be estimated.

could be estimated (see columns 2 and 3 of Table C.43 in Appendix C). Applying the estimated proportional effect on the mortality burden of disease to measures of QALY (including the other PBCs) provides an estimate of the total QALY effect of the change in spend in each PBC (see Column 4).[38] The QALY consequences of changing expenditure by £80m thus reflect PBC estimates of cost per QALY – for example, for the cancer PBC, the predicted total health foregone of 153 QALY was calculated from the change in spend of £2.59m and reflects the PBC specific cost per QALY estimate of £16,997 reported for the threshold estimate in Table C.76. In an analogous way, the comparison of life year and total QALY effects allows the distinction to be made between QALY effects due the life year effects of additional deaths and QALY effects due only to quality of life (see column 5 and 6).

[38] Although there was insufficient mortality available at PCT level to estimate outcome elasticities for the other PBCs, the measure of QALY burden in some of these PBCs does include some mortality (based on ONS data). Therefore, applying a proportionate effect to measures of QALY burden of will include some mortality and life year effects although they represent only a small proportion of the total QALY effects.

Table C4.4: Heath forgone across specific PBCs and groups of ICDs due to the approval of ranibizumab (£80m budget impact)

Total change in spend analysed = **£80 m**	change in spend (m) [1]	Life years foregone [2]	QALYs foregone: Total QALYs forgone [3]	QALYs foregone: Due to premature death [4]	QALYs foregone: Quality of life effects [5]
Overall		**1337**	**4367**	**859**	**3509**
PBC specific					
PBC2 *(Cancer)*	***£2.59***	***217***	***153***	***141***	***11***
Malignant neoplasms, digestive organs (C15-C26)		59	41	38	2
Malignant neoplasms, respiratory system and intrathoracic organs (C30-C39)		50	33	33	0
Malignant neoplasms, breast and female genital organs (C50-C58)		40	31	25	6
Malignant neoplasms, stated or presumed to be primary, of lymphoid, haematopoietic and related tissue (C81-C96)		19	13	13	0
Malignant neoplasms, human male genital organsmale genital organs (C60-C63)		13	9	8	1
Other ICD codes in this PBC		37	25	24	1
PBC10 *(Circulatory)*	***£4.40***	***672***	***625***	***427***	***198***
Ischemic heart diseases (I20-I25)		379	345	242	103
Cerebrovascular diseases (I60-I69)		137	132	85	47
Other forms of heart disease (I30-I52)		54	46	35	11
Congenital malformations and deformations circulatory system (Q20-Q28)		13	31	10	22
Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified (I80-I89)		26	20	16	4
Other ICD codes in this PBC		63	50	39	11
PBC11 *(Respiratory)*	***£2.66***	***93***	***1330***	***58***	***1272***
Chronic lower respiratory diseases (J40-J47)		41	1137	26	1111
Lung diseases due to external agents (J60-J70)		6	52	4	48
Other diseases of upper respiratory tract (J30-J39)		5	47	3	44
Other respiratory diseases principally affecting the interstitium (J80-J84)		2	19	1	17
Other diseases of pleura (J90-J94)		2	19	1	17
Other ICD codes in this PBC		37	57	23	34
PBC7 *(Neurological)*	***£3.47***	***38***	***632***	***25***	***608***
Episodic and paroxysmal disorders (G40-G47)		7	464	5	459
Extrapyramidal and movement disorders (G20-G26)		7	52	4	48
Other degenerative diseases of the nervous system (G30-G32)		6	32	4	28
Other disorders of the nervous system (G90-G99)		1	16	1	15
Nerve, nerve root and plexus disorders (G50-G59)		1	14	1	13
Other ICD codes in this PBC		14	54	10	45
PBC5 *(Mental Health)*	***£20.25***	***55***	***406***	***35***	***371***
Mental and behavioural disorders due to psychoactive substance use (F10-F19)		21	166	15	151
Mood (affective) disorders (F30-F39)		1	69	1	68
Organic, including symptomatic, mental disorders (F00-F09)		18	48	11	37
Neurotic, stress-related and somatoform disorders (F40-F48)		1	38	0	38
Behavioural syndromes associated with physiological disturbances and physical factors (F50-F59)		1	23	1	22
Other ICD codes in this PBC		13	63	8	55

The results reported in Table C4.3 suggests that approval is likely to results in 295 additional deaths (column 2) and 1,337 life years (column 3) forgone, most of which are likely to occur in Circulatory, Respiratory and Cancer PBCs. However, impact of approval of this technology on QALYs forgone due to premature death (column 5) only accounts for a proportion of the total QALY effects (column 4). Most (3,509) are associated with quality of life forgone during disease (column 6). These quality of life impacts are most likely to occur in Respiratory, Neurological and Mental health PBCs. The PBC level effects in Table C4.3 can also be examined at ICD level (Table C4.4) whilst recognising the caveats discussed in Chapter 4. [39] For example, within the respiratory PBC, it appears to be Influenza and Pneumonia (J09-J18) where most additional deaths, life years and quality of life would be forgone. In the Mental Health PBC the additional deaths appear to be associated with disorders due to psychoactive substance use (F10-F19) and Schizophrenia, schizotypal and delusional disorders (F20-F29).

The impact of a reduction in the price of this technology, either through value based pricing or the PAS that was offered during the rapid review, can also be examined in the same way. The PAS was commercial in confidence, so here we will consider the hypothetical case that a 30% reduction in NHS costs (incremental costs) would make this technology cost-effective for this subgroup of patients. Such a discount would be expected to save 1,310 QALYs including 89 deaths averted, 401 life years (258 when adjusted for quality) and quality of life effects during disease equivalent to 1,053 QALYs, compared to approval of the technology at the original list price (Table C4.5).

Table C4.5: Heath forgone before and after a hypothetical PAS scheme on ranibizumab

PBC description	change in spend (m) [1]	Additional deaths [2]	Life years foregone [3]	QALYs foregone		
				Total QALYs forgone [4]	Due to premature death [5]	Quality of life effects [6]
Before PAS						
big 4	£12	246	1,126	2,362	721	1,641
11PBCs	£29	275	1,245	3,500	798	2,701
all 23	£80	295	1,337	4,367	859	3,509
After PAS						
big 4	£8	173	788	1,654	505	1,149
11PBCs	£20	192	871	2,450	559	1,891
all 23	£56	207	936	3,057	601	2,456
Difference						
big 4	-£3	-74	-338	-709	-216	-492
11PBCs	-£9	-82	-373	-1,050	-239	-810
all 23	-£24	-89	-401	-1,310	-258	-1,053

[39] Information about the age, gender and the incidence of sequelae associated with different diseases within a PBC are only available for u-codes which can be mapped to groups of three digit ICD codes. Also allocating PBC level effects to ICD codes was based on the proportion of the total PBC population within each contributing ICD codes because PBC costs are not available at ICD level across PCTs. Although costs from HES data are available at ICD level they are only a small component of total PBC costs and contribute very little to the variability in PBC costs across PCTs especially in those PBCs where mortality effects could not be estimated (also see Footnote 75 and 81 and Addendum 1 in Appendix A).

References

1. Wailoo, A.D., S. Tosh, J, *The incorporation of health benefits in cost utility analysis using the eq-5d: report by the decision support unit.* School of Health and Related Research, University of Sheffield, 2010.
2. NICE, *Guide to the methods of technological appraisal.* ref: N1618, 2008.
3. Dolan, P., et al., *A social tariff for EuroQol: results from a UK general population survey.* CHE discussion paper 138, University of York, 1995.
4. Currie, C.J., et al., *The routine collation of health outcomes data from hospital treated subjects in the Health Outcomes Data Repository (HODaR): descriptive analysis from the first 20,000 subjects.* Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 2005. 8(5): p. 581-590.
5. Cohen, J.W., et al., *The Medical Expenditure Panel Survey: a national health information resource.* Inquiry, 1996. 33(4): p. 373-89.
6. Murray, C. and A.e. Lopez, *The global burden of disease.* Geneva, World Health Organization, Harvard School of Public Health, World Bank., 1996.
7. Claxton, K., et al., *Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development.* Health Technology Assessment, in press 2012.
8. Claxton, K., *The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies.* J Health Econ, 1999. 18(3): p. 341-64.
9. WHO, *The global burden of disease: 2004 update.* 2008.
10. Mathers, C., et al., *Global burden of disease in 2002: data sources, methods and results. GPE Discussion Paper No. 54.* Geneva, World Health Organization. , 2003.
11. Appleby, J. and N. Devlin, *Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision-making* The King's Fund, London, 2010.
12. Gutacker, N., et al., *Truly inefficient or providing better quality of care? Analysing the relationship between risk-adjusted hospital costs and patients' health outcomes.* Health Economics, 2012. doi: 10.1002/hec.2871.
13. Feng, Y., D. Parkin, and N. Devlin, *Assessing the Performance of the EQ-VAS in the NHS PROMs Programme.* Office of Health Economics: Kings Fund, 2012. Research Paper 12/01.
14. Centre, N.I., *Provisional Monthly Patient Reported Outcome Measures (PROMs) in England - April 2012.* 2012.
15. Appleby, J., Harrison, T., Foot, C., Smith, A. and Gilmour, S. , *Explaining variations in primary care trusts' spending on cancer services.* The King's Fund, London, 2011.
16. NICE, *CG82: Core interventions in the treatment and management of schizophrenia in primary and secondary care.* 2009.
17. Knapp, M., et al., *Cost-utility analysis of treatment with Olanzapine compared with other antipsychotic treatments in patients with schizophrenia in the Pan-European SOHO study.* Pharmacoeconomics, 2008. 26(4): p. 341-358.
18. Davies, L.M., et al., *A randomized controlled trial of the cost-utility of second-generation antipsychotics in people with psychosis and eligible for clozapine.* Value Health, 2008. 11(4): p. 549-62.
19. Bagnall, A.M., et al., *A systematic review of atypical antipsychotic drugs in schizophrenia.* Health Technol Assess, 2003. 7(13): p. 1-193.
20. Barton, G.R., et al., *Cognitive behaviour therapy for improving social recovery in psychosis: cost-effectiveness analysis.* Schizophr Res, 2009. 112(1-3): p. 158-63.
21. NICE, *CG90: Depression the treatment and managementof depression in adults.* 2009.
22. Lenox-Smith, A., et al., *Cost effectiveness of venlafaxine compared with generic fluoxetine or generic amitriptyline in major depressive disorder in the UK.* Clinical Drug Investigation, 2009. 29(3): p. 173-184.
23. Kendrick, T., et al., *Cost-effectiveness and cost-utility of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine: randomised controlled trial.* Br J Psychiatry, 2006. 188: p. 337-45.
24. Hatziandreu, E.J., et al., *Cost utility of maintenance treatment of recurrent depression with sertraline versus episodic treatment with dothiepin.* Pharmacoeconomics, 1994. 5(3): p. 249-68.
25. Peveler, R., et al., *A randomised controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine.* Health Technol Assess, 2005. 9(16): p. 1-134, iii.

26. Kendrick, T., et al., *Randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of selective serotonin reuptake inhibitors plus supportive care, versus supportive care alone, for mild to moderate depression with somatic symptoms in primary care: the THREAD (THREshold for AntiDepressant response) study.* Health Technol Assess, 2009. 13(22): p. iii-iv, ix-xi, 1-159.
27. Simon, J., et al., *Treatment options in moderate and severe depression: decision analysis supporting a clinical guideline.* Br J Psychiatry, 2006. 189: p. 494-501.
28. Kaltenthaler, E., et al., *Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation.* Health Technol Assess, 2006. 10(33).
29. Kaltenthaler, E., et al., *A systematic review and economic evaluation of computerised cognitive behaviour therapy for depression and anxiety.* Health Technol Assess, 2002. 6(22): p. 1-89.
30. Hollinghurst, S., et al., *Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial.* Br J Psychiatry, 2010. 197(4): p. 297-304.
31. Appleby, J., et al., *Searching for cost effectiveness thresholds in the NHS.* Health Policy, 2009. 91(3): p. 239-245.
32. Mears, A., et al., *Progress on NICE guideline implementation in mental health trusts: meta-analyses.* The Psychiatrist, 32:383-387., 2008.
33. Rosenheck, R.A., et al., *Cost-effectiveness of second-generation antipsychotics and perphenazine in a randomized trial of treatment for chronic schizophrenia.* Am J Psychiatry, 2006. 163(12): p. 2080-9.
34. Lambert, M.T., et al., *New-onset type-2 diabetes associated with atypical antipsychotic medications.* Progress in Neuro-Psychopharmacology and Biological Psychiatry, 2006. 30(5): p. 919-923.
35. Haas, S.J., et al., *Clozapine-associated myocarditis: a review of 116 cases of suspected myocarditis associated with the use of clozapine in Australia during 1993-2003.* Drug Saf, 2007. 30(1): p. 47-57.
36. Alvir, J.M., et al., *Clozapine-induced agranulocytosis. Incidence and risk factors in the United States.* N Engl J Med, 1993. 329(3): p. 162-7.
37. Crawford, M.J., et al., *Group art therapy as an adjunctive treatment for people with schizophrenia: multicentre pragmatic randomised trial.* BMJ, 2012. 344.
38. Kendall, T., *Treating negative symptoms of schizophrenia.* BMJ, 2012. 344.
39. Tarrier, N., et al., *The Salford Family Intervention Project: relapse rates of schizophrenia at five and eight years.* The British Journal of Psychiatry, 1994. 165(6): p. 829-32.
40. NICE, *TA237: Ranibizumab for the treatment of diabetic macular oedema.* 2011.
41. NICE, *Macular oedema (diabetic) - ranibizumab (rapid review of TA237): appraisal consultation document.* 2012.
42. Novartis, *Single technology appraisal (STA) manufacturer submission: Lucentis ® (ranibizumab) for the treatment of visual impairment due to diabetic macular oedema (DMO).* 2010.

Appendix D: Project Protocol

1. Title

Methods for estimation of the NICE cost-effectiveness threshold.

2. Importance

A comparison of the incremental cost effectiveness ratio (ICER) of a new technology with a cost-effectiveness threshold is not the only consideration when the National Institute for Health and Clinical Excellence (NICE) and its advisory committees issues guidance. But is an important one: it allows an assessment of whether the health expected to be gained from the use of a technology exceeds the health expected to be forgone elsewhere as other NHS activities are displaced. For this reason a comparison of the ICER of a technology to a threshold range is a critical part of the reference case in the NICE Guide to Methods of Appraisal[1] and is often taken to be the starting point for deliberations about other consideration including judgements of social value. Therefore, the value of the threshold or the range of values used is critical to the assessment of whether technologies can be regarded as cost-effective with implication for NHS patients, local NHS decision makers, the Department of Health, HM Treasury, manufacturers (pharmaceuticals and devices) as well as NICE itself.

i) *What is the cost-effectiveness threshold?* – In principle, the cost effectiveness threshold is an estimate of health forgone as other NHS activities are displaced to accommodate the additional costs of those technologies recommended in NICE guidance.[2] A national decision-making body like NICE needs an estimate of what is likely to be forgone on average across the NHS as we currently find it. Of course, this will change as circumstances and the NHS changes; tending to rise with increases in budget and health care costs but tending to fall with increases in the productivity of health technologies and the efficiency of the NHS in general - including better local commissioning decisions.[3] A body like NICE cannot and does not necessarily need to know what *specific* services and treatments will be displaced in particular localities or who will actually forgo health. What is required, however, is an accountable and empirically-based assessment of the health that is likely to be forgone on average across the NHS.

ii) *What are current estimates based on?* – Currently NICE uses a threshold range of £20,000 to £30,000 per quality adjusted life year (QALY) gained, where additional considerations are required towards the upper bound. The empirical basis of this range of values is very limited. At best it represents an informal assessment of the health gained by some of the least productive (in health outcome terms) of the activities currently undertaken by the NHS. Unsurprisingly, it does, to a certain extent, represent the implied values from past NICE decisions. It is widely recognised, by NICE[4] and the House of Commons (HoC) Health Committee[5] among many others,[6] that the current range ought to be more firmly based on empirical analysis. The HoC Health Committee highlighted particular concerns that the additional costs of NICE guidance imposed on local NHS commissioners might be causing the displacement of more valuable health care.[5] On the other hand, manufacturers and others have argued that the threshold range is too low, restricting market access, prices and revenue and ought to be based on how much individuals are willing to pay for improvements in health.[7] In 2009 NICE convened a workshop to discuss what the threshold ought to represent and how it might be more securely estimated.[8] Most of the applicants contributed to that workshop and it has informed the plans for research set out below.

iii) *What is needed?* – Explicit scientific methods for estimation are required which will provide accountability so that estimates can be scrutinised by a range of stakeholders. Since estimates of the threshold will need to be periodically revised, methods which make best use of routinely available NHS data are needed. As well as accountability, this will also provide more predictability in likely changes to the threshold for the investment decisions of technology manufacturers. Providing more secure estimates of the threshold will necessarily require application of specialist expertise and the development and application of sometimes sophisticated methods. Nevertheless these must be communicated effectively to stakeholders to ensure transparency. Suitable methods should provide estimates relevant to NICE - that is, relevant across the NHS; they should capture the effects of health care on both length and quality of life, offer the opportunity to estimate changes in the threshold over time and indicate the impact of 'non marginal' changes which have a significant budget impact on the NHS.

3. Scientific Potential

3.1 People and Track Record

The principal investigator and co-applicants form a substantial and complementary team with long experience

of the type of multi-disciplinary research required for the development of methods and their application in health policy. Importantly, they have complementary expertise in key areas of: analysis of system performance and the application of econometric methods; familiarity with a range of relevant data sets; methods of outcome measurement in health and relevant sources of evidence; and how estimates of cost-effectiveness can be used to inform health care decisions. They also have complementary and intimate knowledge of the NICE process from a number of perspectives; knowledge of the wider policy environment and an understanding of the needs and difficulties faced by local commissioners and providers. Together they can draw on the resources of a large, well established and experienced team of researchers. In addition, a multidisciplinary advisory group for the project will include representation of key stakeholders.

Karl Claxton is a Professor in the Department of Economics and the Centre for Health Economics at the University of York. His research interests encompass the economic evaluation of health care technologies and how such evidence should be used inform decisions in health care.[2,3,5] He has made a particular contribution to the development of methods to more fully reflect uncertainty in estimates of cost-effectiveness and establish the value of information associated with health care decisions. He has served as a member of the NICE Appraisal Committee since 1999. He is a member of the NICE Decision Support Unit and contributed to the review of the NICE Guide to the Methods of Technology Appraisal.[1]

Nancy Devlin is Director of Research at the Office of Health Economics and Senior Associate at the King's Fund, London. Nancy's principal research interests are the measurement and valuation of health-related quality of life; and the cost effectiveness thresholds used in health technology appraisals. Nancy's 2004 econometric analysis of NICE decision making, together with David Parkin, continues to be highly cited in this field.[9] More recently, she was involved in research (funded by NICE R&D) to investigate the feasibility of identifying the threshold implicit in local NHS decision making.[10] She is a member of the Executive of the EuroQol Group and leads the EuroQol Group's research on the development of time trade-off methods of valuing the EQ5D, and is researching new methods for analysing EQ-5D data collected via PROMs.[11]

Steve Martin is a research fellow in the Department of Economics at the University of York. He has published in various areas of health economics including resource allocation and the impact of waiting times on the demand for and supply of NHS health care. He has participated in a series of studies that have examined the relationship between health care expenditure and health outcomes at the PCT level. These studies have used English programme budgeting expenditure and mortality data to estimate the cost of a life year saved in several disease areas.[12-14]

Nigel Rice is Professor of Health Economics at the Centre for Health Economics, University of York, and is Director of the Health, Econometrics and Data Group (HEDG). HEDG is a research group focused on the use of applied quantitative techniques capable of informing health and health care policy. The Group's research aims to inform health-related policy in areas such as health inequality, health policy evaluation and the performance of health care systems.[12-14] Nigel is also member of the Department of Health sponsored Public Health Research Consortium. He has researched and written extensively on methods for resource allocation and has published on the relationship between health care expenditure and health outcomes. He is a member of the Advisory Committee on Resource Allocation (ACRA) and its technical advisory group.

Mark Sculpher is lead applicant, Professor of Health Economics at the Centre for Health Economics, University of York, and Director of the Programme on Economic Evaluation and Health Technology Assessment. Mark has worked in economic evaluations of range of technologies and contributed to methods in the field, particularly decision analytic modelling and handling uncertainty. He is a member of the UK National Institute of Health Research College of Senior Investigators. He has been a member of the NICE Technology Appraisal Committee and currently sits on the NICE Public Health Interventions Advisory Committee. He chaired NICE's 2004 Task Group on methods guidance for economic evaluation and advised the Methods Working Party for the 2008 update of this guidance.[1,2] He is also a member of the Commissioning Board for the UK NHS Health Technology Assessment programme and the UK Medical Research Council's Methodology Research Panel. He recently led a research team which assessed NICE's methodological research needs, a study funded by the NIHR/MRC.

Peter C. Smith is currently Director of the Centre for Health Economics and will, from October 2009, be Professor of Economics at Imperial College Business School. He is a mathematics graduate from the University of Oxford, and started his academic career in the public health department at the University of Cambridge. He has published widely on the financing and performance of health systems, and has a special interest on the links

between research evidence and policy.[12-14] In the UK, Professor Smith has served on numerous Health Department advisory committees[5] on finance and productivity issues, and was a board member of the Audit Commission. He has also acted as consultant to many overseas ministries and international agencies.

Collaborators: David Parkin from the South Eastern Strategic Health Authority (SHA) will provide an important NHS perspective and specialist expertise[9-11] in identifying possible sources of evidence of disinvestment at a local level to complement the analysis of national programme budget data.

Advisory Group: A multidisciplinary advisory group will provide complementary expertise and experience to inform the development of each stage of this research. We will agree the make up of the advisory group with NICE. However, we anticipate representation from key stakeholders; e.g., NHS commissioners, senior members of NICE and its advisory bodies; Department of Health, HM Treasury; and manufacturers.

3.2 Environment

This research will be collaboration between the Centre for Health Economics (CHE) at the University of York, Imperial College Business School and the Office of Health Economics.

The University of York is internationally renowned for its contribution to health services research in general, health economics in particular and the timely application of methods to inform health policy. The University of York jointly headed the RAE2008 table for quality of health services research (researchresearch.com/RAE2008), with CHE making a major contribution to the submission. The University of York pioneered the development of health economics in the UK, and in 1983 created CHE, which is now one of the largest and best known health economics research centres in the world. CHE operates across all areas of the discipline, with a particular emphasis on methodological thinking and high policy impact. It is known especially for its work in health technology assessment, health status measurement, performance measurement and productivity, health care financing and econometric methodology.

The evaluation team in CHE has long experience of providing high quality technology assessments for NICE and developing methods to support the functions of the Institute. It has a wealth of experience in methods, application and the demands of the NICE process. In addition the team has contributed to a broader understanding of how evidence of cost-effectiveness can be used to inform social decisions about health care, pricing and research. The health policy team and HEDG undertake applied and methodological research to inform health and health care policy. Recent work includes: improving the methods of measuring the productivity informing the changes introduced by the Office of National Statistics; comparative performance and efficiency analyses of health and other public sector organisations; and analysis of the link between expenditure and outcomes in the NHS.[12-14]

The Office of Health Economics (OHE) undertakes research, advisory and consultancy services on economic and policy issues within the health care, pharmaceutical and biotechnology sectors. OHE aims to stimulate and inform discussion and debate about health economic and policy issues among academics, policy makers and industry executives. In addition to OHE's research programme, its activities include seminar series; annual lectures; and an in-house publication series. The quality and independence of OHE's publications are safeguarded through its Policy Board and Editorial Board. Funding for the research undertaken at OHE comes from a range of sources including the Gates Foundation; SDO; NIHR; DH R&D; EuroQol Foundation; and the King's Fund; as well as from industry.

The group has well-established collaborations on a range of methodological issues. It has bid collaboratively for several research grants, and members of each group have published extensively together on a range of topics. They are experienced at communicating methods and results to non-specialist policy audiences. All the applicants pursue active dissemination strategies which, in additional to the traditional routes of publication and presentation, include direct involvement in a plethora of committees and engagement in a variety of advisory roles, as well as contributing to a range of training for relevant stakeholders. In summary, the group has the necessary expertise and experience of the types of methods and their application as well as familiarity with sources of data which are required for this research. They also have a wealth of experience of NICE, wider health policy and the needs of local NHS decision makers which is further strengthened though collaboration with South East Coast SHA. In addition, they have a long history of successful collaboration with well-established and close working relationships.

3.3 Research Plans

The experience of attempts to look in some detail at the decision making processes and outcomes at a local level have

demonstrated the complexity, variability and difficultly of estimating the cost effectiveness thresholds implicit in local NHS decision making.[10,15] Although such studies provide valuable insights into the nature of local decision making, they are unlikely to meet the needs set out in section 2 because: i) they do not readily provide the national picture NICE requires; ii) it is not clear how they could be used to estimate changes over time, the effects of non-marginal changes or add to predictability; and iii) such detailed studies are very much bespoke and specific to time and place, so the more granular view they provide cannot feasibly be routinely collected. In addition, it may not be necessary to know precisely which health technologies are displaced, for which patients or why. What is required is an estimate of the health that is likely to be forgone *on average* across the NHS, ideally based on routinely available data. Therefore, this research will focus on complementary methods which can make best use of those data that are already available, where there are already plans to make data available or where additional data could feasibly be made available at reasonable cost. The research plans fall into the following 4 complementary areas of activity, all of which will be evaluated at a user impact workshop.

3.3.1 Review of principles, methods and estimates of the threshold

The literature which considers the cost-effectiveness threshold has grown over recent years.[3] However, most is rather discursive,[7,16] some more analytic,[2,9] but only a few attempts to offer methods for empirical estimation.[10,12-14] This diverse literature (including policy documents) needs to be thoroughly reviewed and pulled together in a clear, comprehensive and structured way. This will: i) clarify the terms of the debate and establish a common understanding of concepts and the types of estimates that are required; ii) review how different approaches to estimation seek to meet the needs outlined in 2); and iii) outline what is missing from current estimates and how methods might be most usefully developed. This review will be written for a wide policy audience, but will also be comprehensive; covering all the issues raised in recent debates and explain alternative methods of estimation in an accessible way. The review may inform the basis of an agreed framework with all relevant bodies (NICE, DoH, HoC Health Committee, HM Treasury etc). It might also inform the remit of the type of independent body suggested by HoC Health Committee[5] or the periodic reviews suggested by NICE.[8]

3.3.2 Analysis of programme budget data

Since 2003 data on expenditure on health care across 23 programmes of care have been prepared by each Primary Care Trust (PCT) in the English NHS. These programme budgeting (PB) data seek to allocate exhaustively to disease categories (via ICD10 codes) all items of NHS expenditure, including expenditure on inpatient care, outpatient care, community care, primary care and pharmaceuticals and devices. It serves a number of purposes, notably to assist in the local planning of healthcare. But its crucial merit for this study is that it opens up the possibility of examining the relationship between local spending and associated disease-specific outcomes.

Previous work by some of the applicants has demonstrated the potential value of programme budgeting data in estimating the link between expenditure on a programme of care and the health outcomes achieved, in the form of disease-specific mortality routinely available from the National Centre for Health outcomes Development. In each programme changes in mortality associated with changes in expenditure are transformed into life years, providing estimates of the marginal cost per life-year gained on average across the NHS. This work has focused largely on spending and outcomes in two of the largest programmes of healthcare: circulatory disease and cancer[12], but has also informed the link across other programme categories[13,14]. Estimates of the cost per life year gained for 2006/07 are £15,387 for cancer; £9,974 for circulation problems; £5,425 for respiratory problems; £21,538 for gastro-intestinal problems; and £26,428 for diabetes. These estimates are based on a straightforward, though carefully constructed, theoretical model of health production which informs the specification and estimation of a system of equations (issues of endogeneity are dealt with by identifying and testing suitable instruments). In doing so, they account for variation in the clinical needs of the local population relevant to the programme of care and broader local environmental factors relevant to the costs of care and outcomes. In principle, this approach, based on routine data, estimates the type of cost-effectiveness threshold required by NICE: the 'average' marginal elasticity of spending with respect to income amongst the PCTs. However, the methods of analysis need to be developed in a number of important respects; not least to express outcome in terms of quality as well as length of life (see 3.3.3). These are outlined below.

i) What is the overall threshold for the NHS? - The overall threshold for the NHS will depend on the programmes of care where disinvestment takes place. Hitherto, each programme of care has been estimated separately so it is not clear how expenditure on particular programmes changes with the overall budget, e.g., does disinvestment tend to fall on respiratory care or diabetes? In principle, spending on programmes is linked by a system of equations, and we will seek to model the programme expenditure equations as a set of linked simultaneous equations, reflecting the potential for interactions between spending on the different programmes, brought about by the need for PCTs to operate within a fixed overall budget. This will offer an opportunity to

study the 'budget elasticity of expenditure' in each programme of care. It is then feasible to derive estimates of the impact of marginal increases (or decreases) in overall PCT budgets on spending in each of the programme categories. As well as indicating budgetary influences on programme spending these can then be linked to changes outcomes by programme. This can provide an estimate of the cost per life year gained on average across the NHS, for marginal changes in budget. This type of estimate of the threshold will take account of how such budgetary changes translate through local decisions into changes in expenditure on programmes of care and then to health outcomes. In addition, knowledge of budget elasticities of expenditure across the different programmes of care, coupled with the estimated relationship between expenditure and health outcome, will indicate how expenditure changes in one programme impacts on health outcomes in other programmes of care, providing more secure estimates of the relationship between overall expenditure and overall health outcome.

ii) *What is the likely impact of non-marginal changes?* - To the extent that data permit, we shall seek to study year-on-year changes in spending, as well as a cross section of spending decisions. Changes in budgets are in practice incremental, and it may be the case that the elasticities of programme expenditure in times of budgetary increase (when new initiatives are introduced) are not the same as in times of budgetary decrease (when the focus is on disinvestment). They may also vary depending on the current level of expenditure (relative to need) on a specific programme of care. In general, elasticity might be expected to increase as spending increases, but this can be tested. This offers the opportunity to explore the possible effect of non-marginal changes on programme expenditure and possibly outcome, providing estimates of the threshold for a range of budget impacts. This type of analysis would provide some guidance to NICE on when a decision might regarding a new technology have such a significant impact on the NHS budget that there will be significant reallocations of expenditure between programmes, with more valuable health care forgone so that a lower estimate of the threshold might be appropriate. It might also suggest when a series of apparently marginal changes (mandatory NICE guidance) will start to have non-marginal effects on the NHS.

iii) *How do estimates change over time?* - We will investigate whether successive years of PB data can be used to form a panel dataset. This might be limited by changes to PCT boundaries but, should it prove feasible, a robust panel will allow an assessment of the stability of our estimates. It will also allow: an investigation of how elasticities and estimates of the threshold change over time; an assessment of the feasibility of periodically re-estimating the threshold based on these types of data; and an exploration of the possibility of making predictions based on overall budget forecasts. Previous analysis of PB data assumed a quasi long-run equilibrium so that health outcomes could be contemporaneously linked to expenditure. This is likely to be more tenable in some programmes of care than others. However, should it prove feasible to construct a robust panel, we will investigate empirically the appropriateness of the assumption of equilibrium and whether the relationship between programme expenditures and health outcomes is better represented by data lags and how a suitable lag structure might differ between programmes. We will also investigate the potential of using other data sources to complement the programme budget data. For example, Hospital Episode Statistics, while restricted to secondary care, may allow analyses at a more granular level than the PB data alone. This may include small area data on heath care needs and supply which, together with practice Quality and Outcomes Framework (QOF) data, may provide a better means of adjusting for variation across practices within and between PCTs.

Overall the analysis will yield a set of budget elasticities of expenditure, disaggregated by programme of care and (if feasible) by level of spending with appropriate links between programme expenditure and outcomes over time. An important policy question is, then, the impact of a policy 'shock' on spending patterns, over time and between programmes. In principle, the study should yield information needed to examine the dynamic impact of such a shock, using simulation methods such as system dynamics.[17] The feasibility and specification of this approach will depend on the information secured in the earlier stages, but the study will seek to develop such a model, to the extent that data permit.

3.3.3 Evidence of quality of life

The link between variations in budget, changes in expenditure on programmes of care and health outcomes, in the form of disease-specific mortality described in 3.3.2, needs to be extended so that outcome can be expressed in terms of quality of life. The previous analysis of PB data made this link by assigning quality of life estimates (by ICD10 from the Health Outcomes Data Repository (HODaR)) to the change in life years estimated from disease specific mortality within each programme. However, the reported costs per QALY gained for each programme did not capture improvements in quality of life independent of effects on mortality, tending to overestimate the programme-specific thresholds, particularly in those programmes where expenditure tends to be associated with improvements in quality rather than mortality. A cost-effectiveness threshold for NICE needs to be expressed as the cost per QALY gained. Therefore a more complete picture of the quality of life outputs from the NHS spending in

3.3.2 is needed. Additional work is required to evaluate complementary sources of evidence and methods of analysis which will allow them to be combined with the results from the econometric analysis outlined in 3.3.2. There are two areas in particular where additional work is required:

i) Weighting improvements in length of life - The estimated gains in life expectancy associated with reductions in disease-related mortality need to be weighted to reflect the health-related quality of life (HR-QoL) of the additional years of life. As each year of life gained is not experienced in full health, costs per life-year gained will tend to underestimate the cost per QALY gained. Also, the HR-QoL of additional years is likely to differ between programmes. Therefore it is important incorporate evidence of HR-QoL to obtain a more complete picture of the relationship between expenditure and health outcome. This will be addressed two ways. Firstly, by reviewing, in each programme and associated ICD10 chapters, published quality of life evidence, exploiting existing databases of quality of life studies, e.g., the TUFTS Cost-Effectiveness Analysis Registry, EuroQol group. It is anticipated that this will reveal considerable heterogeneity between, but also within, programmes. These variations might suggest the need to complement aggregate data at programme and ICD chapter level (see 3.3.4) with other evidence. Any systematic differences and variations will provide a basis for a sensitivity analysis around the quality of life weightings applied to programmes. Secondly, other sources of data currently available (see HODaR below) or likely to become available in the future (see PROMs below) will be reviewed for suitability and quality. Where possible, the impact of using alternative sources of HR-QoL evidence on estimates of the threshold will be demonstrated through sensitivity analysis.

ii) *Capturing improvements in quality of life* - NHS resources are often spent on services which do not have the aim of reducing mortality, but rather of improving quality of life. Failing to account for this type of output will underestimate health outcomes and overestimate the cost per QALY gained, with the scale of overestimation differing across programmes. This will be addressed in two ways. Firstly, the review of published HR-QoL evidence described above will distinguish QALY gains arising from reductions in mortality and improvements in quality alone. We will examine whether there are any systematic differences between programmes in the proportion of health gain arising from improved length or quality of life. Again, we expect to find heterogeneity within and between programmes which might suggest that complementing aggregate data with other evidence of displacement could be useful (see 3.3.4).

Secondly, we will investigate the use of disease specific data sets, which report HR-QoL and may supplement the type of analysis in 3.3.2. For example, the Patient Reported Outcome Measures (PROMs) initiative was introduced in April 2009. The pilot for PROMs provided condition-specific and generic health outcomes data before and after surgery for four elective procedures, which allows analyses of the relationship between variation in spending and improvements in QALYs[11,18] by procedure. The PROMs project is to be rolled out over a range of chronic conditions in the near future. However, it is unlikely that these data will be available within the timeframe for this project. Nevertheless, establishing the method of analysis that will be needed to make best use of these data will be useful, so they can complement the analysis in 3.3.2 when they become routinely available. We will also investigate the use of HODaR which collects single observations of the EQ-5D profile from patients 6 weeks following a range of procedures and services delivered in secondary care. These data record ICD10 chapter and may provide a means of estimating the HR-QoL of patients for a range of interventions within and between programmes. Unlike PROMS there are no repeated measures with HODaR, so estimating *gains* in HR-QoL will not be possible. Nevertheless, it may provide a means of establishing post-treatment QALYs within programmes and, together with evidence from the review, indicate the variation within and between programmes. We are aware of, and will investigate further, the availability and potential usefulness of other NHS datasets collecting only disease-specific measures. For example, some data are routinely collected in mental health, using instruments such as HoNOS and COR-OM. However, there are concerns about data quality and the difficulty of translating these measures into QALYs.[19]

The analysis of elasticities described in 3.3.2 will be used to identify those programmes which are a particular priority for review and analysis of quality of life estimates. For example, it may be that only a few ICD chapters account for most of the changes in programme expenditure due to marginal changes in budget. By prioritising in this way we will be able to focus the search and analysis of quality of life estimates on those programmes which are most critical to the cost-effectiveness threshold. Based on estimates over time, this type of approach will also be used to identify priorities for future routine data collection in those programmes which are most critical to estimates of the threshold.

3.3.4 Evidence of investment and disinvestment

Although the analysis of PB data will provide estimates of how programme expenditure responds to marginal changes in overall spending it cannot, by itself, provide details of how changes in expenditure on a particular

programme are allocated within the programme, e.g., the services, treatments and procedures invested in or disinvested. As discussed above, diseases-specific mortality is only one aspect of outcome, but there is likely to be significant heterogeneity within and between ICD chapters in quality of life. Therefore, it would be valuable to have more detailed evidence about the types of investments and disinvestments made within programmes and ICD chapters. Other sources of evidence would enable use of estimates quality of life which more closely matched the types of investment and disinvestment which lie behind the more aggregate changes in programme expenditure and complement the econometric analysis of PB data in 3.3.2. A complete and detailed picture of all investment and disinvestments across the NHS is not feasible on a routine basis nor would it be necessary. Our focus will be on what can be gained from other routinely collected data at a local level, and what additional evidence would be most useful and could be gathered at reasonable cost. We will also explore how and whether such evidence might contribute directly to the quantitative estimates from 3.3.2 and 3.3.3, or provide useful contextual information and help with a qualitative assessment of the considerations that ought to be applied when interpreting estimates of the threshold.

We will be working with the support and guidance of Professor David Parkin, Chief Economist at South East Coast SHA to identify and evaluate potentially useful sources of evidence that are already, or planned to be, collected at a local level (PCT and Trusts). For example, from March 2009, PCTs have been required to provide documents that detail their policies on services that they will not normally fund. Working with SHAs, we will collate these lists across the NHS. Initial communications from one SHA indicate their PCT lists contain nearly 100 such items. We will also examine the 'pledges' for new services or service delivery routinely reported in the Operating Frameworks of SHAs and PCTs. We will link all these items to programmes in 3.3.2 and to the review of quality of life in 3.3.3, whilst anticipating that some items may require bespoke review of published literature. We will explore what impact this more granular information on investment and disinvestment within programmes has on estimates of programme-specific expenditure per QALY gained and estimates of the overall threshold. However, we recognise that these data are unlikely to report all forms of investment and disinvestment (e.g., scale of services and eligibility for treatment). For this reason, we will recruit and work with NHS organisations within the South East Coast SHA to design and pilot additional data collection to supplement that available from the sources described above. We will prioritise any additional data collection based on the analysis in 3.3.2 and 3.3.3. For example, we will focus the pilot of additional data collection on those programmes which account for most of the changes in expenditure (high elasticities) due to changes in overall budget. It is estimating the QALY impact of these changes which contributes most to the cost-effectiveness threshold. While evidence from this pilot data collection may not be representative of the NHS as a whole, it will show whether additional data would be a useful complement to national PB data and routinely collected quality of life measures. It will also identify whether additional data collection is likely to be feasible on a routine and representative basis. The analysis of the PB data will help identify what might constitute a representative sample of PCTs for these purposes.

3.3.5 User impact assessment

We plan to present and assess the impact of this research through a workshop which will involve a range of key stakeholders. We hope that the workshop will be under the auspices of NICE and would be a full day, adopting a similar format to the recent methods workshops which informed the revision of the NICE Methods guide. All material will be pre-circulated, and will clearly pose the questions to be addressed. The presentations of the three aspects of the work described above will be followed by group discussion with facilitators and note takers with feedback on the day. The results will be circulated to participants for comment before a final report is produced. We will consult our advisory group and the Institute about possible workshop participants. However, we anticipate that they will include representation from key stakeholders including: NICE; PCTs and other local NHS decision makers; Department of Health (including those with responsibility for commissioning and pharmaceutical pricing); HoC Health Committee; HM Treasury; National Audit Office; industry (pharmaceutical and medical devices); and academics from the relevant research communities. The primary output will be a series of recommended options that NICE may choose to take forward for public consultation.

4. Ethics and Research Governance

This proposal does not involve patients in the collection of primary data, and there are no ethical implications which would require approval from the University of York ethics committee.

5. Data Preservation for Sharing

There are no issues as our primary purpose is to develop methods which can make best use of data that are, or maybe, routinely available within the NHS.

6. Public Engagement with Science

The recommendations made to the Institute following the workshop may well be taken forward for public consultation by NICE.

7. Exploitation and dissemination

Dissemination to a wide range of stakeholders and audiences will take place in a number of ways. The initial results will be disseminated through the user impact workshop. The final project report and a report from the workshop will, we hope, be disseminated through the NICE website. We intend to present this work at the NICE conference and, internationally, at iHEA and HTAi. We plan to write a number of papers addressing different aspects of this work and different audiences. We anticipate the following publications: *British Medical Journal*, communicating the key principles of methods, estimates and their implications; *Journal of Health Economics* presenting the development of methods and, where possible, their application to UK data. We will also incorporate this knowledge into taught courses both as part of MSc programmes in York and in short course programmes delivered by both centres nationally and around the world.

References

1. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. NICE: London, 2008.
2. Culyer AJ, McCabe C, Briggs A, Claxton K, Buxton M, Akehurst R, Sculpher M, Brazier J. Searching for a threshold, not setting one: the role of the National Institute of Health and Clinical Excellence. Journal of Health Services Research and Policy, 2007; 12(1): 56-58.
3. McCabe C., Claxton K. and Culyer AJ. The NICE cost-effectiveness threshold: what it is and what it means. Pharmacoeconomics, 2008; 26(9): 733-744.
4. House of Commons Health Committee. National Institute for Health and Clinical Excellence: NICE response to the first report of session 2007-2008. HC550. London: Stationery Office, 2008.
5. House of Commons Health Committee. National Institute for Health and Clinical Excellence. First report of the Health Committee 2007-2008. HC27-I. London: Stationery Office, 2008.
6. The government's response to the Health Select Committee's first report of session 2007-08 on the National Institute for Health and Clinical Excellence. Cm7331. London: Stationery Office, 2008.
7. Towse A. Should NICE's threshold for cost per QALY be raised? Yes. BMJ, 2009; 338: 268-269
8. National Institute for Health and Clinical Excellence. Threshold workshop: report of a technical meeting organised by NICE. NICE: London, 2008.
9. Devlin N, Parkin D. (2004) Does NICE have a cost effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Economics 13(5): 437-52.
10. Appleby J, Devlin N, Parkin D, Chalkidou K, Buxton M. Searching for cost effectiveness thresholds in the NHS. Health Policy (forthcoming 2009).
11. Devlin N, Parkin D, Browne J. (2009) Using the EQ-5D as a performance measurement tool in the NHS. Economics Discussion Paper 09/03, Economics Department, City University.
12. Martin, S., Rice, N. and Smith, P. (2008a), "Does health care spending improve health outcomes? Evidence from English programme budgeting data", Journal of Health Economics 27, 826-842.
13. Martin, S., Rice, N., and Smith, P. (2008b), "Further evidence on the link between healthcare spending and health outcomes in England", The Health Foundation, London, May, pp61.
14. Martin, S., Rice, N., and Smith, P. (2009), "The link between health spending and health outcomes for the new English primary care trusts", The Health Foundation, London, June, pp64
15. Williams I, McIver S, Moore D, Bryan S. (2008) The use of economic evaluations in NHS decision-making: a review and empirical investigation. Health Technol Assess. 2008 Apr;12(7):iii, ix-x, 1-175.
16. Raftery J. Should NICE's threshold for cost per QALY be raised? No. BMJ, 2009; 338: 268-269.
17. Smith, P. and van Ackere, A. (2002), "A note on the integration of system dynamics and economic models", Journal of Economic Dynamics and Control, 26(1), 1-10.
18. Browne, J., Jamieson, L., Lawsey, J., van der Meulen, J., Black, N., Cairns, J., Lamping, D., Smith, S., Copley, L., Horrockes, J. Patient Reported Outcome measures (PROMs) in elective surgery. Report to the Department of Health, 2007.
19. Jacobs R. Investigating patient outcome measures in mental health . Background paper, OHE Commission on NHS outcomes, performance and productivity. OHE London: OHE, 2007