Using Decision Models To Overcome Limitations in Data from Randomised Trials: The Case of Glycoprotein 2b/3a Antagonists For Acute Coronary Syndrome

Mark Sculpher[a]
Stephen Palmer[a]
Zoe Philips[a]
Mike Robinson[b]
Laura Ginnelly[a]
Keith Abrams[c]

a. Centre for Health Economics, University of York, York.
b. Nuffield Institute, University of Leeds, Leeds
c. Department of Epidemiology and Public Health, University of Leicester, Leicester

Correspondence: Mark Sculpher, Centre for Health Economics, Heslington, York, YO10 5DD. Email mjs23@york.ac.uk. Tel. 01904 433641

PAPER TO BE PRESENTED AT THE DEVELOPING ECONOMIC EVALUATION METHODS SEMINAR, OXFORD 11-12 APRIL 2002

Acknowledgements

This work was funded by the NHS Health Technology Assessment Programme. The case study summarised here was developed, in part, for a Technology Assessment Report for the National Institute of Clinical Excellence. Other co-authors to that report were: Khaled Alfikah, Ameet Bakhai, Nicola Cooper, Marcus Flather, David Gray and Chris Packham. The views expressed in this paper and any errors are, however, those of the authors alone.

Summary

The randomised clinical trial is considered the gold standard for testing hypotheses about particular clinical parameters. The increasing demand for formal analysis to synthesise available data on costs and outcomes to support system-wide decisions about which health care technologies should be funded from collective resources continues to highlight deficiencies in trials as a basis for economic evaluation. These limitations relate to factors such as partial comparison of options, short-term follow-up and the use of intermediate end-points. Question marks are also raised about the value of trial-based economic evaluation for this form of decision making when evidence exists from a number of trials. The only approach that has emerged formally to synthesise data in costs and outcomes for cost-effectiveness analysis is decision analytic modelling. Although recent papers have considered how 'good decision modelling' might be defined in general terms, the specific approaches that need to be taken in models to overcome limitations in trial data need to be considered further. Using the specific example of the economic evaluation of glycoprotein IIb/IIIa antagonists in the management of acute coronary syndrome, where over 40,000 patients have been randomised in trials, this paper seeks to illustrate the types of limitations of RCTs as a main source of data for economic evaluation. It also aims to show how decision models can provide an analytical framework within which to address these limitations.

1. Introduction

The randomised clinical trial (RCT) has developed an important role the clinical evaluation of health care technologies. The primacy of the RCT in this field is seen in its role in the licensing of new pharmaceuticals, where its high level of internal validity has resulted in it being the 'gold standard' in establishing efficacy. In recent years, a number of collectively-funded health care systems have begun formally to use economic analysis to make decisions regarding which health care interventions should be funded from collective resources.[1] This area of 'reimbursement' decision-making is quite distinct from that relating to licensing for a number of reasons. These include the need for information on effectiveness (rather than efficacy), resource use, costs and impact on health-related quality of life; the requirement for evidence which is more specific to a particular jurisdiction; and the fact that, for some health care systems, formal reimbursement decisions extend to more health care technologies than pharmaceuticals.[2]

Despite the additional evidential requirements of reimbursement decision-making, the RCT is seen as retaining a key role in this area. For example, in its technology appraisal guidance, the National Institute for Clinical Excellence (NICE) argues that 'the ideal source of effectiveness data is a prospective, randomised, controlled trial with a naturalistic design which imposes the minimum restriction on the normal decision-making processes …' (p13).[2] The use of trial-based economic evaluation is one aspect of the contribution of RCTs to this form of decision making. This form of economic study involves the use of patient-level data from a RCT on health-related outcomes and resource use to estimate the cost-effectiveness of the options being compared. This use of RCTs as a vehicle for economic evaluation is increasingly prevalent. For example, a total of 1103 records, published between 1998 and 2001, have been abstracted onto the NHS Economic Evaluation Database. Of these, 302 (27%) were trial-based economic studies (Julie Glanville, personal communication.

www.york.ac.uk/inst/crd). Even if full trial-based economic evidence is not available, the RCT is seen as the main source of evidence on health-related outcomes for reimbursement decision making.

The value of the RCT in providing estimates of treatment effect which are free of the implications of selection bias is clear. However, the emergence of more explicit approaches to reimbursement decision-making has highlighted the limitations of the RCT as the main source of data for economic evaluation. This is apparent in the first few years of the NICE appraisal programme, for example. In part, limitations have manifested themselves in terms of the near absence in useful trial data relating to non-pharmaceutical technologies such as hip prostheses and wisdom tooth extraction (www.nice.org.uk). However, even with pharmaceuticals, factors such as partial comparisons, short time horizon, incomplete measurements and lack of generalisability have limited the value of RCTs for guidance in areas such as cancer, motor neurone disease and multiple sclerosis (www.nice.org.uk).

Given these shortcomings of RCTs, there is a need to develop explicit, rigorous and defensible analytical methods to pool available data from trials and other sources. Although it has its detractors,[3,4] the decision model is the only explicit analytical framework to have emerged to synthesise data on a range of effectiveness, resource use and value parameters to provide guidance on optimal reimbursement decisions under conditions of uncertainty. In the face the increased profile of modelling in health service decision-making, there have been a number of recent general guidelines on what constitutes 'good' decision modelling.[5,6] However, little has been written on the specific approaches to decision modelling required when data from RCTs exists but are characterised by particular shortcomings.

Using the specific example of glycoprotein IIb/IIIa antagonists in the management of acute coronary syndrome, where over 40,000 patients have been randomised in trials, this paper seeks to illustrate the types of limitations of RCTs as a main source of data for economic evaluation. It also aims to show how decision models can provide an analytical framework within which to address these limitations. The next section of the paper provides greater detail on the typical weaknesses of RCTs as the main source of information for reimbursement decision making. Section 3 describes the case study. The objective is to summarise the key steps that were taken to overcome the problems in trial evidence rather than to provide full details of the economic model and its results. Section 4 discusses some of the issues that emerge from this sort of modelling.

2. The limitations of trials for economic evaluation

There are some clear advantages of trial-based economic evaluations, for example the opportunity offered by the availability of patient-level data to allow for the correlation between costs and outcomes and to explore cost-effectiveness in sub-groups of patients. The potential weaknesses of (at least explanatory) trials as vehicle for economic evaluation have been described previously.[7] However, the potential weakness of trial-based economic evaluation has been emphasised by the increased need to synthesise all appropriate evidence for the purposes of economic analysis for reimbursement decision making. If several trials have been published on a particular comparison of interventions, trial-based economic analysis alongside one trial will, at best, provide only partial guidance to decision makers. In terms of *clinical measurement*, a meta-analysis of all trials is seen as representing the highest level of evidence.[8] However, unless all of the trials collected broadly the same resource use and outcome data and patient-level data were available from each study, a cost-effectiveness meta-analysis would probably not be feasible.

There are a series of more general limitations of RCTs as the main source of data for reimbursement decision making even when it is recognised that trial-based economic analysis using patient-level data is not desirable for feasible. The first of these is that a decision about the most cost-effective form of management for a patient group requires the comparison of the full range of feasible management options. At best, a pragmatic trial will compare a new technology to standard practice which will only be adequate if that comparator has itself been shown to be cost-effective. At worst, a RCT will compare the new form of management to one of several possible options, or to a placebo as is usually the case with Phase III regulatory trials undertaken for drug licensing.

A second limitation of the trial as the main source of evidence for economic evaluation is that, as well as being partial in comparison, they are often partial in measurement with not all the parameters of interest being measured. A good example of this is the frequent use of intermediate outcome measures (e.g. time until disease progression in trials of cancer therapies; changes in CD4 count or viral load in trials of HIV treatments). The use of intermediate outcomes is understandable given the large number of recruits and extensive periods of follow-up that are often needed to measure ultimate health outcomes with the power required for clinical research. However, for cost-effectiveness analysis, the 'down-steam' implications of interventions for quality-adjusted survival are essential to estimate.

A third limitation is that the RCT is often characterised by a short time horizon. Again, this has much to do with the cost of the trials themselves, particularly when they are being undertaken to support licence applications where the onus is to achieve that objective at the earliest opportunity. However, a mismatch frequently occurs between trials and decision making regarding reimbursement because the latter is concerned with estimates of cost and outcomes over the full period that they

are expected to differ between options being compared. This mismatch is perhaps at its most pronounced in the case of potentially life-saving interventions where cost-effectiveness is greatly dependent on how long the additional survivors identified in a trial are likely live, with what quality of life and at what cost to health the service.

A fourth limitation of trials is that they can exhibit a lack of external validity. This is a result of the tendency of trials to recruit unrepresentative patients from specialist centres staffed by professionals with a research interest. The implications of this for clinical measurement may not be significant, at least in relative terms, if these are considered to be more transportable across settings. However, lack of generalisability may have a profound effect on the reliability of resource use estimates. Perhaps the clearest example of this is the use of multi-national trials to generate more statistical power for clinical estimates, but which may produce estimates of cost (and hence cost-effectiveness) which do not apply directly to any one jurisdiction.

If any of these limitations applies, additional methods are required to synthesise those (typically partial) measurements from RCT(s) with data from other sources (e.g. observational studies), as well as explicit assumptions, to provide a complete picture of cost-effectiveness whilst also reflecting uncertainty in parameter estimates. A case-study where many of these limitations with trials apply is introduced in the next section together with a description of how decision modelling can be used to address the limitations in trial data.

3. A case-study: glycoprotein IIb/IIIa antagonists for non-ST-elevation acute coronary syndrome

3.1 Background

3.1.1 Acute coronary syndrome

The glycoprotein IIb/IIIa antagonists (GPAs) represent a new class of drug for the acute treatment of non-ST-elevation acute coronary syndromes (ACS). Further information about these drugs and ACS can be found elsewhere,[9-11] and the purpose here is only to provide a brief summary. ACS is a term that includes a range of patients with a similar underlying pathology. At one end of the spectrum are those patients with evidence of ST elevation on a resting electrocardiogram (ECG) who are eligible for treatment with thrombolysis and who may subsequently develop Q-wave on their ECG – this is a full myocardial infarction (MI). The potential role for GPAs, however, relates to the remaining ACS patients who are classified as having either unstable angina or non-Q-wave MI. Non-Q-wave MI is the term used when the cardiac enzymes are elevated to the range indicating that MI has occurred, but a Q-wave does not develop on ECG tracings. Unstable angina itself represents a spectrum of clinical states that fall between stable angina and acute MI. It includes new onset angina and angina occurring >24 hours post-MI. Not only is unstable angina an unspecific diagnostic category, but patients present with varying degrees of atherosclerosis (stenosis size, location and plaque fragility), thrombus formation (low or high platelet content) and vasospasm. Each of these contributes to the morbidity and mortality of the disease and, therefore, represents a potential target for intervention with medical therapy. Aspirin and heparin are currently used to reduce thrombus formation, and nitrates are used to help reduce vasospasm and cardiac oxygen requirements. Interventional therapy typically involves percutaneous coronary intervention (PCI), such as angioplasty, or coronary artery bypass surgery.

3.1.2 Glycoprotein IIb/IIIa antagonists

The formation of the thrombus in ACS patients results from a complex interaction of the coagulation system and platelet homeostasis. GPAs are a class of drugs that

may be more effective in preventing platelet aggregation than existing therapies such as aspirin and heparin. Two broad groups of GPA are licensed in the UK: abciximab (ReoPro®, Eli Lilly) is a monoclonal antibody targeted at the receptor (also known as a 'large molecule' GPA); while eptifibatide (Intergrilin®, Schering Plough) and tirofiban (Aggrastat®, MSD) are more conventional pharmacological receptor antagonists (also known as 'small molecule' GPAs). GPAs are used in two general ways to manage ACS patients. Firstly, as an adjunct to PCIs (e.g. angioplasty) for those patients who undergo such a procedure – abciximab is the GPA which is mainly used for this purpose. Also, GPAs can be used as a form of medical management for non-ST-elevation ACS patients regardless of whether or not they subsequently go on to have a PCI – tirofiban and eptifibatide are mainly used for this indication.

3.1.3 The evidence base

A recent systematic review of GPAs for NICE found no shortage of RCTs assessing GPAs.[11] Trials of medical management have randomised over 30,000 patients, and typically they compare GPAs with standard management. Overall these trials have shown a reduction in the risk of non-fatal MI or death – an odds ratio of 0.91 (95% CI 0.84, 0,98) in a recent meta-analysis.[12] Trials of the use of GPAs alongside PCI have been more heterogeneous in their intake, with only one trial focusing solely on non-ST-elevation ACS patients. Overall, studies recruiting some patients with ACS (10 trials randomising over 15,000 patients) have shown the use of GPAs to generate a relative risk of non-fatal MI of about 0.68 (95% CI 0.57, 0.80) and of death of 0.80 (95% CI 0.60, 1.09) by between 30 days and 6 months.[13]

The scope for GPAs to be cost-effective relies on the extent to which, by reducing mortality and non-fatal MI, they are able to generate enough gain in quality-adjusted survival for the average patient, and/or to reduce 'downstream' health care costs sufficiently to justify their acquisition cost. A fair amount of economic evidence has

been amassed on GPAs,[11] some of which is based on analysis of patient-level data alongside specific RCTs. However, none of these studies has adopted a long-term time horizon, expressed outcomes in terms of generic measures of health gain such as quality-adjusted life-years (QALYs), focused on UK costs and clinical practice and compared a full range of feasible strategies for the use of GPAs in ACS. As such, these studies provide minimal assistance for UK decision makers concerned with the reimbursement of GPAs.

3.1.4 Cost-effectiveness analysis for UK decision making

The aim of a recent technology assessment undertaken for NICE[13] was, using the available evidence base, to provide an analytical framework to evaluate the cost-effectiveness of GPAs in non-ST-elevation ACS. Despite the large number of patients randomised to GPA trials, these studies exhibited major limitations as a source of evidence for decision making about cost-effectiveness in the UK. The principle shortcomings are summarised below.

1. The fact that a large number of different trials have been undertaken on GPAs indicates that using trial-based economic analysis with *any one study* to address cost-effectiveness using patient-level data would be limited. To do so would inevitably be partial because it would be ignoring data on thousands of patients randomised in other studies. In principle, it would be possible to undertake a patient-level meta-analysis of all trials but, even if such data were made available, there is little likelihood that they would have collected data fully and consistently, particularly on resource use.
2. The RCTs were undertaken wholly or largely outside of the UK. This is particularly important with GPAs, for two reasons. The first is that the baseline event rates in patients not having GPAs in the UK may be quite different to those patients randomised to the control groups in the trials. This might reflect

differences in the epidemiology of the disease or, more probably, differences in overall management in the UK. The principle difference in management of coronary heart disease in the UK is that fewer patients are considered for PCI than in most developed countries.[14] Given that it has been argued that the medical management of patients with GPAs is less likely to be effective in those patients who do not subsequently undergo PCI,[15] this feature of the trials cannot be overlooked.

3. The trials also have short follow-up, typically no more than six months and often as little as 30 days. However, the use of GPAs to reduce the risk of mortality and non-fatal MI in non-ST-elevation ACS will have important long-term implications for quality-adjusted survival and health service costs, and these 'downstream' consequences are not directly informed by the trials, although they need to be considered as part of the decision making process. This lack of long-term mortality data, together with the fact that none of the trials collected preference-based quality of life data, precludes patient-level estimates of QALYs. This represents a major limitation for NHS-wide decisions about reimbursement.
4. None of the trials directly compared the various ways in which GPAs could be used in ACS patients in the UK. The medical management trials typically compared GPAs with standard practice, and the PCI trials evaluated the adjunctive use of GPAs alongside the procedure with PCI alone. However, no trials directly compared the use of GPAs as a medical management in all (or a sub-group of high-risk) ACS patients with its use alongside PCI in that proportion of patients who undergo interventional therapy and with no use of GPAs.

Given the limitations of the trial evidence as a single basis for decision making, a decision model was developed to assist NICE decision making.[13] In developing the model, a range of approaches was necessary to overcome the features of the trial evidence.

3.2 Methods

3.2.1 Model overview

Full details of the methods and results of the modelling exercise are available elsewhere,[13] and it is not the purpose of this paper to provide full details. Rather, a brief overview is provided focusing on the methods that were adopted to overcome the constraints imposed by the available trial data. A full report of the model, providing information on the structure, data inputs and results is available on request from the authors.

The model was developed to estimate, over a 50-year time horizon, costs from the perspective of the UK NHS, and health outcomes in terms of quality-adjusted life-years (QALYs). The model was made up of two parts: a short-term element, which related to a period of six months after a patient presents with non-ST-elevation ACS; and a long-term element which extrapolated a patient's lifetime costs and outcomes conditional on surviving the first six months after the acute episode. The model was probabilistic, and Monte Carlo simulation was used to propagate 2nd order uncertainty in input parameters through the model to be reflected in decision uncertainty. A 2000-2001 price base was used, and annual discount rates of 6% for costs and 2% for benefits were adopted based on UK guidance.[16]

3.2.2 Identifying relevant treatment strategies for the UK

Based on a review of clinical literature and advice from clinical collaborators, four treatment strategies were identified as being relevant options for the use of GPAs in ACS patients in a UK setting:

- *Strategy 1: GPA as part of initial medical management*. This involves patients with ACS receiving an infusion of GPA as soon as their "high risk" nature has been established.
- *Strategy 2: GPA in patients with planned percutaneous coronary interventions (PCIs)***.** GPA is started once a decision to undertake PCI (or angiography with a view to proceeding to PCI) has been made.
- *Strategy 3: GPA as adjunct to PCI*. GPA is used at the time of PCI or is started up to 1 hour before the procedure in those patients undergoing such a procedure.
- *Strategy 4: No use of GPA*. With this strategy, patients are assumed to receive standard therapies (e.g. heparin, aspirin, nitrates and analgesia), without the use of GPA.

None of the trials of GPAs compared directly these strategies, although trials were available to consider, individually, Strategies 1 to 3, in each case relative to Strategy 4. Therefore, it was necessary to use indirect trial evidence in the model.

3.2.3 Allowing for different case mix and clinical practice in the UK

As indicated above, RCTs undertaken to evaluate the clinical effectiveness of the GPAs were mainly or wholly undertaken outside the UK.[11] In many respects, treatment patterns and resource use in the UK can be expected to differ from those in centres involved in the trials. For example, the rate of PCI in patients with ACS, and in inshaemic heart disease generally, is lower than in most developed countries.[14] One implication of these differences in UK practice is that the baseline event rates observed in the trials (i.e. in the control groups) are unlikely to provide reliable estimates for UK practice. These baseline event rates include death, non-fatal MI, PCI, coronary artery bypass grafting and bleeding rates (a complication of GPAs).

For this reason we constructed baseline event rates, specific for UK practice, from an alternative data source – the Prospective Registry of Acute Ischaemic Syndromes in the UK (PRAIS-UK).[17] This is an observational cohort registry of 1046 patients admitted to 56 UK hospitals with ACS in 1999. Patients were followed-up for 6 months after their index hospital admission and the hospitals included in PRAIS-UK served 24% of the UK population. For the purposes of the study, patients who received GPA in PRAIS-UK (n=13; 1%) were excluded from the analysis.

The parameter estimates from PRAIS-UK relating to patients who received a PCI during the acute phase of their ACS were based on a relatively small number of patients (n=53). For this reason, an audit of unstable angina patients undergoing acute PCI at a large UK cardiac centre (Leeds) was undertaken to supplement data from PRAIS-UK (n=231).

Within the short-term model, baseline resource use data (i.e. relating to Strategy 4) were also taken from PRAIS-UK. In part, resource use related directly to the clinical events in the model (e.g. PCIs, MIs, coronary artery bypass grafting). In addition, mean length of in-patient hospital stay was taken from PRAIS-UK. Three other areas of resource use were modelled explicitly within the baseline model: MI, complications associated with the use of GPAs and costs associated with death. Each of these, together with all unit costs, were based on UK data sources.

3.2.4 Incorporating the effectiveness of GPAs

By incorporating UK-specific data on baseline event rates, the position taken in developing the model was that the event rates in the control groups of the trials were unrepresentative of what would be reflected in UK clinical practice. It was necessary to address the question of whether the relative risks associated with GPAs, which

have been estimated in the trials, should also be adjusted to reflect differences in UK practice.

To inform this decision, meta-regression analysis was undertaken to establish whether, across published trials and taking each strategy separately, the relative risk in a trial was related to the absolute baseline risk in that study.[18] No statistically significant association was found, which may reflect the small number of trials in the analysis. For this reason, the relative risks from the trials were incorporated into the model without adjustment, which is equivalent to assuming that relative risks are transportable across health care systems whilst the baseline risks in those studies are not.

Pooled relative risks for non-fatal MI, death, revascularisation and gastrointestinal bleeding were incorporated into the model based on pooling using a random effects meta-analysis of which two were undertaken. The first related to the medical management trials which compared GPAs to standard management in ACS patients early in the acute presentation, with only a proportion of patients subsequently undergoing PCI (in effect a comparison of Strategies 1 and 4). The second meta-analysis was of the use of GPAs alongside the PCI procedure (in effect a comparison of Strategies 3 and 4). Only one trial evaluated GPAs as in Strategy 2, so no meta-analysis was necessary. Therefore, the relative risks of the three GPA strategies were estimated separately compared to standard care (because no trial directly compared them), and these were then brought together within the model.

A further complexity of the trial evidence was that the three GPAs licensed in the UK had been used in more than one way in the trials. This was the case despite the fact that the drugs are licensed for more specific purposes: abciximab is licensed for use with PCI (Strategy 3); and eptifibatide and tirofiban are licensed for medical management

(Strategies 1 and 2). Furthermore, clinical advice and some available data[14] indicated that the licence generally reflects routine clinical use. Despite this mismatch between how the clinical evidence had accumulated and the routine use of the drugs, it was decided to include all trials in the meta-analyses. For example, the trial GUSTO IV was included in the pooled estimate of Strategy 1 trials despite the fact that the relevant GPA was abciximab which is not licensed in the UK for that indication. Indeed, three medical management (Strategy 1) trials were used in the meta-analysis despite the fact that they used a GPA which is unlicensed in the UK. The principle behind this was that the most reliable estimate of overall treatment effects of GPAs would be generated by using as much experimental evidence as was available in ACS patients. Sensitivity analysis was used, however, to examine the effects of alternative inclusion criteria for the trials in the meta-analysis.

3.2.4 Incorporating the costs of GPAs

The acquisition costs of the three licensed GPAs were based on undiscounted prices from the British National Formulary.[19] For Strategy 1, the total drug costs per patient were based on the average cost of eptifibatide and tirofiban. This was based on clinical collaborators' indication that these were the GPAs typically used for this strategy – indeed they are the only ones licensed for this indication in the UK. This decision was taken despite the fact that abcixmab was used in some of the Strategy 1 trials as noted above. A similar decision was taken for costing GPAs in Strategy 2: the average cost of eptifibatide and tirofiban was used due to the licence and clinical advice, despite the fact that the only trial evaluating that strategy actually used abciximab. In Strategy 3 the drug costs are calculated on the basis of an infusion of abciximab because this is the drug licensed and routinely used for this purpose.

3.2.6 Quantifying the long-term effects of GPAs

Most of the trials set up clinically to evaluate GPAs were characterised by very short-term follow-up – typcially no longer than 6 months. Any assessment of the cost-effectiveness of GPAs, however, has to allow for the long-term cost and outcome implications of the short-term effects of the drug. This 'extrapolation' was needed for two reasons. Firstly, many patients who are treated for ACS will continue to consume health service resources for their heart disease for the remainder of their life, and the effect of GPAs in the first 6 months may influence these costs. Secondly, in order to compare the cost-effectiveness of GPAs with other uses of health service resources (inside and outside cardiology), it is necessary to express the benefits of the drug in terms of a generic measure of health gain which can be compared across treatment areas such as the QALY. In order to provide a realistic estimate of the impact of GPAs on QALYs, the long-term implications for survival and health-related quality of life of the short-term (within 6 months) effects of the drugs need to be modelled.

Therefore, a long-term (extrapolation) model was developed to estimate a future prognosis for patients who finish the short-term (six month) model in one of two disease states: those having experienced a non-fatal MI and those who have not but remain alive. That prognosis will include the possibility of patients experiencing further non-fatal MIs as well as dying for any reason. Hence, the extent to which the use of GPAs reduces the risk of death and non-fatal MI, relative to baseline, during the initial 6-month period was translated into differences in long-term costs and QALYs on the basis of the long-term model.

The long-term model took the form of a 4-state Markov process with states of ischaemic heart disease (that is, patients who had not experienced a non-fatal MI), non-fatal MI (where patients spend a single cycle of one year), post-MI (which surviving patients entered after one year following an MI) and death. Transition probability and cost data for the long-term model were taken from a UK-specific observational study - the

Nottingham Heart Attack Register (NHAR). Two cohorts of patients (total n = 1,279) from the NHAR were used with a diagnoses indicative of ACS which had follow-up data for up to 5 years. Survival analysis indicated that fixed hazard rates were consistent with the follow-up data, and event rates were extrapolated beyond 5 years on this basis. Quality-adjustment was based on data in the literature,[20] with a constant decrement applied to all living patients in the model.

3.3 Results

3.3.1 A brief summary of the base-case results

Full results are presented elsewhere,[13] and it is not intended to report them in detail here. Base-case results are presented in Table 1 based on mean costs and QALYs. Figure 1 presents the cost-effectiveness acceptability curves for the base-case analysis showing uncertainty in the decision based on the precision of the input parameters. The base-case analysis indicated that Strategy 1, the use of GPAs as a medical management early in presentation regardless of whether the patient subsequently went on to have a PCI, was the most cost-effective strategy assuming that the UK health service is willing to pay at least £5,667 per additional QALY.

3.3.2 Sensitivity analysis

Extensive sensitivity analyses were undertaken relating to uncertainties in the model such as particular assumptions and data inputs. Full details are not provided here but, from the viewpoint of the general methods used for the model, these sensitivity analyses included:

(i) The use of baseline event rates from the trials rather than from PRAIS-UK. This had the effect of changing the case mix from that observed in the UK observational study, and of increasing the rate of PCI.

(ii) Using separate relative risk reductions for Strategy 1 according to whether patients subsequently underwent PCI. This was undertaken because of the clinical view that most of the benefit associated with medical management of ACS using GPAs is confined to those subsequently undergoing PCI.[15] These data were taken from sub-group results presented in a recently published patient-level meta-analysis of medical management trials.[12]

(iii) Sensitivity analyses around the trials included in the meta-analysis based on whether the drug used in those trials was used in routine clinical practice in the UK. Similar adjustments to the costing of GPAs.

(iv) Use of sub-group data from PRAIS-UK relating to high-risk patients – diabetics, those with ST-depression on ECG and those aged over 70 years.

The base-case results were found to be generally robust to the variation in the sensitivity analyses.

5. Discussion

The case-study presented here is by no means unique in terms of the characteristics of trial evidence and how it limits cost-effectiveness analysis for reimbursement decision making. The NICE appraisal process has found limitations in trial evidence including short follow-up and the use of intermediate end-points with technologies such as implantable cardioverter defribrillators, coronary stents and pharmaceutical therapies for advanced ovarian disease and obesity (www.nice.org.uk). In principle, it is possible to describe a RCT with the characteristics to provide a vehicle for economic evaluation and to provide adequate information for reimbursement decision making. Such a trial would randomise appropriate patients between each feasible management option and would follow-up patients over a sufficiently long period to measure differential costs and outcomes – this latter feature is perhaps more likely with interventions for acute events

than with those for chronic diseases. In addition, for such a trial to be adequate as a sole basis for decision making, there would need to be no additional sources of evidence other than that trial. It is difficult to list examples of health technologies for which trials with these characteristics have existed.

It should be emphasised that the features of trials which have been highlighted in this paper do not indicate poor quality research. Rather, the trials may well be adequate for the purposes for which they were designed – to test hypotheses relating to particular clinical parameters. For clinical decision making, showing a difference between interventions in terms of the *direction* of a clinical measurement is sufficient. For decision making relating to cost-effectiveness, however, the extent of that difference and its implications of health gain in generic terms is particularly important.

The imperfections of trials as a basis for decision-making about reimbursement poses some important questions for cost-effectiveness methods. The case-study presented here faced some particular characteristics of the trial evidence which needed to be addressed in modelling the optimal use of GPAs. The first of these was the fact that none of the trials formally compared all the management strategies within which GPAs could feasibly be used. Hence is was necessary to use indirect evidence of treatment effects from those partial comparisons undertaken in the trials. The process of expressing treatment effects in terms of *relative* risks of events which were then applied to a common set of baseline risks probably makes this indirect use of trial evidence more reliable than if *absolute* changes in risk had been indirectly compared. However, it remains possible that, due to differences in the populations in the different trials, the relative risks are not completely comparable.

The second limitation of the trial data for the GPAs was that those studies were undertaken largely or wholly outside the UK. This is likely to have implications for the

case-mix of patients presenting with non-ST elevation ACS and their baseline event risks. Furthermore, in the context of the management of coronary heart disease, clinical practice patterns in the UK are different to those in other developed countries, particularly in terms the smaller proportion of patients who undergo PCI. This limitation of the trial evidence for UK decision making is likely to be mirrored in many other clinical areas, particularly relating to drug technologies where the trials are part of an international programme of regulatory studies. In the GPA model, this feature of the trial evidence was handled by using a UK observational study of ACS patients to estimate baseline risks of clinical events such as non-fatal MI and death.

The extent to which the observational study presented up-to-date estimates of baseline risks thus became an important issue with the model. The recent increase in PCI rates in the UK[14] suggested that the observational study (undertaken in 1999) may have been somewhat out of date. This was dealt with using sensitivity analysis to explore the implications of changing baseline risks to more closely reflect those in the trial, and by separately applying relative risks according to whether a patient subsequently underwent a PCI (available from a sub-group analysis in recent patient level meta-analysis[12]).

The assumption behind the modelling approach was that the relative risks from the trials were transportable across health care systems, but the baseline risks were not. Meta-regression was undertaken to assess further the justification of this assumption. This sought to estimate a relationship between the risks of clinical events in the control groups of the trials and the relative risks associated with GPAs.[18] The relatively small number of trials available limited the precision of this analysis, and the lack of a 'statistically significant' relationship was used to justify no further adjustment to the relative risks to reflect UK practice. However, this remains a potentially useful approach to transport the results of a trial from one setting to another.

The third limitation of the GPA trials for reimbursement decision making was their short-follow-up. Most trials followed patients for no more than 6 months and frequently for only 30 days. This was a weakness because, although the main purpose of the drugs is to reduce the risk of death and non-fatal MI during the short-term acute phase, a cost-effectiveness analysis for system-wide reimbursement decision-making needs to estimate the value of this change for long-term prognosis. The approach adopted with the model was to use data from a second UK observational study to model the long-term implications for health service costs and QALYs for patients who survived 6 months with and without experiencing a non-fatal MI.

The extrapolation in the model was relatively straightforward because it is reasonable to assume that there is no continued treatment effect from GPAs beyond the short-term acute period given the relatively short period over which the drugs are taken. Hence extrapolation amounts to adding a fixed (although uncertain) cost and number of QALYs to survivors from the short-term model conditional on whether or not they had experienced a non-fatal MI during that period. When modelling the long-term effects of pharmaceuticals which are taken over many years (e.g. cholesterol lowering medications), extrapolation methods need to allow for the uncertainty regarding whether the treatment effect will continue over the long-term.

An important general issue needs to be considered in assessing the appropriateness of the methods adopted in the model to overcome limitations in the trial data, and this relates to whether all the uncertainty these methods introduce is fully reflected. The model was probabilistic in that 2nd order uncertainty in the model parameters was reflected in model inputs and propagated through the model using Monte Carlo simulation. However, the process of synthesising data from a range of sources and extrapolating over time introduces additional uncertainty into models over and above that

relating to precision in input parameters. At the moment, it is unclear how these uncertainties should be reflected in decision models.

In conclusion, there is likely to be increasing use of formal economic analysis to inform system-wide decisions about the reimbursement of health care technologies. This process will continue to highlight the limitations of RCTs as the basis for this form of decision making, both as a vehicle for patient-level economic analysis and as source of effectiveness data for decision models. Further consideration is needed regarding the most appropriate methods of synthesis to overcome limitations in trial data.

References

1. Hjelmgren, J, Berggren, F, and Andersson, F. Health economic guidelines - similarities, differences and some implications. *Value in Health* 2001;4:225-250

2. National Institute for Clinical Excellence. *Technical Guidance for Manufacturers and Sponsors on making a Submission to a Technology Appraisal (http://www.nice.org.uk)*. London: National Institute for Clinical Excellence, 2001.

3. Sheldon, TA. Problems of using modelling in the economic evaluation of health care. *Health Economics* 1996;5:1-11

4. Freemantle, N and Mason, J. Not playing with a full DEC: why development and evaluation committee methods for appraising new drugs may be inadequate. *British Medical Journal* 1999;318:1480-2

5. Sonnenberg FA, Roberts MS, Tsevat J. Toward a peer review process for medical decision analysis models. *Medical Care* 1994;32 (supplement):JS52-JS64.

6. Sculpher, M, Fenwick, E, and Claxton, K. Assessing quality in decision analytic cost-effectivness models. A suggested framework and example of application. *Pharmacoeconomics* 2000;17(5):461-477

7. Drummond, MF and Davies, L. Economic analysis alongside clinical trials. *International Journal of Technology Assessment in Health Care* 1991;7:561-573

8. Egger M, Davey Smith G, Altman DG. *Systematic Reviews in the Health Care: Meta-Analysis in Context*. London: BMJ Publishing, 2001.

9. McDonagh, MS, Bachmann, LM, Golder, S *et al*. A rapid and systematic review of the clinical effectiveness of glycoprotein IIb/IIa antagonists in the medical management of unstable angina. *Health Technology Assessment* 2000;4 (30).

10. Fischer A, Frankish R, Taylor R. *Clinical and cost-effectiveness of Glycoprotein IIB/IIIA inhibitors in association with percutaneous coronary intervention*. London: National Institute for Clinical Excellence, 2000.

11. Robinson M, Ginnelly L, Jones L. *et al*. *A Systematic Review Update of the Clinical Effectiveness and Cost Effectiveness of Glycoprotein IIb/IIIa Antagonists*. York: NHS Centre for Reviews and Dissemination and Centre for Health Economics, University of York, 2001.

12. Boersma, R, Harrington, RA, Moliterno, DJ *et al*. Platelet glycoprotein IIb/IIIa inhibitors in acute coronary syndromes: a meta-analysis of all major randomised clinical trials . *Lancet* 2002;359:189-98

13. Palmer S, Sculpher M, Philips Z, Robinson M, Ginnelly L. *et al. A cost-effectiveness model comparing alternative management strategies for the use of glycoprotein IIB/IIIA antagonists in non-ST-elevation acute coronary syndrome*. York: Centre for Health Economics, University of York, 2002.

14. British Cardiovascular Intervention Society . *BCIS Audit Returns of Interventional Procedures 2000 (http://www.bcis.org.uk/audit/oct01.html)*. London: BCIS, 2001.

15. Knight, C. NICE guidance. Too nice to glycoprotein 2b 3a antagonists? *Heart* 2000;85:481-3

16. Department of Health. *Policy Appraisal and Health*. London: Department of Health, 1995.

17. Collinson, J, Flather, MD, Fox, KAA *et al*. Clinical outcomes, risk stratification and practice patterns of unstable angina and myocardial infarction without ST elevation: Prospective Registry of Acute Ischaemic Syndromes in the UK (PRAIS-UK). *European Heat Journal* 2001;21:1450-7

18. Sharp, SJ and Thompson, SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of appraoches. *Statistics in Medicine* 2000;19:3251-3274

19. *British National Formulary (www.BNF.org, accessed October 2001)*. London: British Medical Association and Royal Pharmaceutical Society of Great Britain, 2001.

20. Fryback, DG, Chinnis, JO, and Ulviva, JW. Bayesian cost-effectiveness analysis: an example using the GUSTO trial . *International Journal of Technology Assessment in Health Care* 2001;17:83-97

Table 1. Base-case estimates of mean lifetime costs and QALYs for the four strategies, together with incremental analysis

Strategy	Cost	QALY	ICER	Probability cost effective for maximum WTP[a]:		
				£10,000	£30,000	£50,000
1	£12,331	7.7928	£5,667	81.9	94.17	95.24
2	£11,849	7.6893	D	0.73	0.74	0.65
3	£11,847	7.6961	ED (£33,478)[b]	0.43	2.09	2.58
4	£11,770	7.6938		16.94	3.00	1.53

a. The probability that each strategy is more cost-effective than the others conditional on different maximum willingness to pay for an additional QALY

b. ICER Strategy 3 versus Strategy 4

D = Dominated, ED = Option ruled out by extended dominance

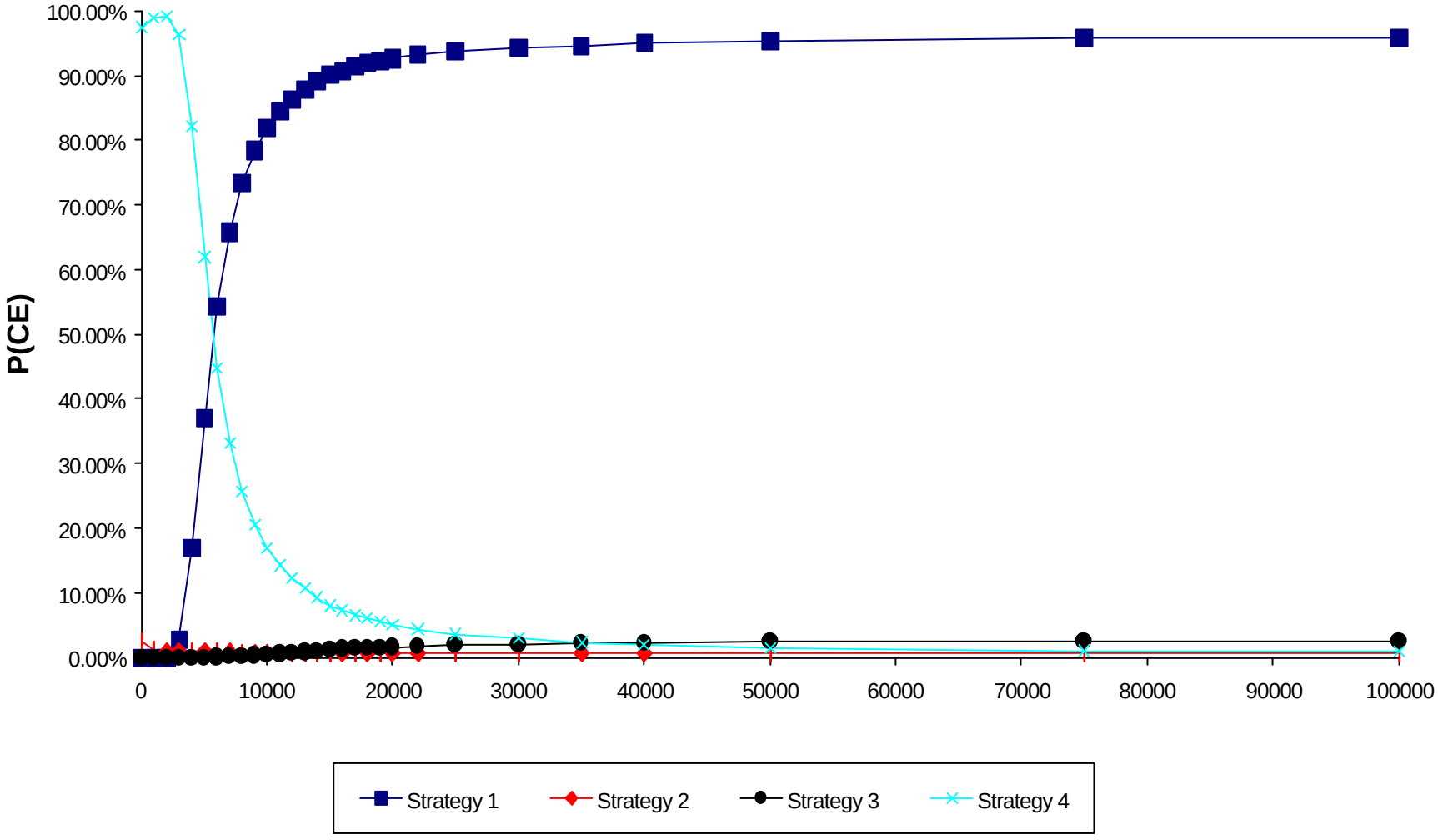

Figure 1. Base-case results in the form of cost-effectiveness acceptability curves.