Response to the OHE critique of CHE Research paper 81

Karl Claxton and Mark Sculpher

Summary

CHE Research Paper 81, November 2013 (Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith PC, Sculpher M. Methods for the estimation of the NICE cost effectiveness threshold), is the final revised version of the report on research funded by the National Institute of Health Research (NIHR) and Medical Research Council to establish methods that can estimate the NICE cost-effectiveness threshold using routinely available data. This final peer reviewed report was accepted for publication in the NIHR's *Health Technology Assessment* monograph series.

The Office of Health Economics (OHE) published a critique of CHE Research Paper 81 in December 2013 (OHE Occasional paper 13/07). We responded in great detail and considerable length to an earlier draft of this critique in February 2013. The revised OHE critique comments that "There is much to admire in the approach, detailed analysis, and reporting set out in CHERP81 and we recommend detailed reading of it." We welcome this comment and would suggest reading Chapter 5 of CHERP81 and especially Section 5.4, which includes an important summary of key assumptions (Box 5.1 page 82-83), and their qualitative effect (optimistic 1-4, conservative 5-6 or unknown when the effect is an empirical matter that current evidence cannot resolve, 8-10). This Box also provides links to text and footnotes where these assumptions are examined in greater detail elsewhere in the report.

However, the OHE critique does not appear to be a dispassionate examination of what the balance of evidence suggests for the NICE threshold. Throughout there is a tendency to favour higher threshold estimates by overstating the optimism of some assumptions, understating the conservatism of others and omitting a proper discussion of others that are conservative with respect to the health effects of changes in expenditure. This selectivity is also evident in the way the analysis is discussed, including a failure to acknowledge fully the implications of analysis that suggests that the threshold for NICE decisions is likely to be lower than a central estimate. On the contrary, our analysis has erred on the side of conservatism with respect to the health effects of changes in expenditure. A threshold that NICE might apply, when considering new technologies that impose costs on the NHS, is also likely to be lower than any central estimate based on our analysis.

In common with four independent peer reviewers, we remain confident that our analysis has made best use of existing evidence, that our interpretation is careful and balanced, with the implications of the necessary assumptions explored and clearly stated (see Box 5.1 in Section 5.4 for a summary). As such, it provides the best evidence of the scale of the health opportunity costs when NICE guidance imposes costs on the NHS.

In Table 5.1 in Section 5.2 we report a central estimate of £12,936 per QALY based on the analysis of 2008/09 expenditure and 2008/09 to 2010/11 mortality data. This evidence suggests that NICE is currently using cost-effectiveness thresholds which mean that more harm than good is being done to NHS patients' health overall (Claxton et al 2015). For example, NICE has not been rejecting

technologies with incremental cost-effectiveness ratios (ICERs) below its stated upper bound of £30,000 per QALY, so £30,000 has evolved as an effective minimum threshold. Recent evidence suggests that, even when special considerations, such as end of life criteria, do not apply, many technologies with ICERs above £30,000 per QALY are nonetheless recommended (a 0.5 probability of rejection resulted from ICERs of £39,417 to £43,949 per QALY, see Dakin et al. 2014). It is quite clear (empirically) that the upper bound of the NICE threshold range has effectively become the starting point for deliberation.

The implications of holding different views about which assumptions would provide extreme but equally likely upper and lower bounds for the threshold are illustrated in Figure 1 and explored more fully in 'Implications of alternative assumptions and judgements on the central estimate of the cost-effectiveness threshold '. This analysis illustrates that even if very pessimistic assumptions about the health effects of changes in NHS expenditure are regarded as credible the overall policy implication, that the current thresholds used by NICE are almost certainty too high, remains unchanged. This analysis also demonstrates that those that wish to claim otherwise, for whatever reason, must adopt assumptions which are so extreme they appear perverse.

For example, the OHE suggests that adopting much more pessimistic assumptions would result in an upper and lower bound for the threshold of £29,314 to £12,936 per QALY (see page 26). Such assumptions are neither credible nor balanced. Nevertheless, these bounds would actually result in a central estimate of threshold of £17,969 (see Figure 1 below). Later they propose an even more extreme upper bound of £67,664 and regard our discounted central estimate (£13,724) as an equally extreme lower bound (see page 27). The assumptions required are so extreme they appear perverse. Nonetheless, they would result in a central estimate of £22,820 per QALY, substantially lower than the thresholds actually being used by NICE (see Figure 1 below).

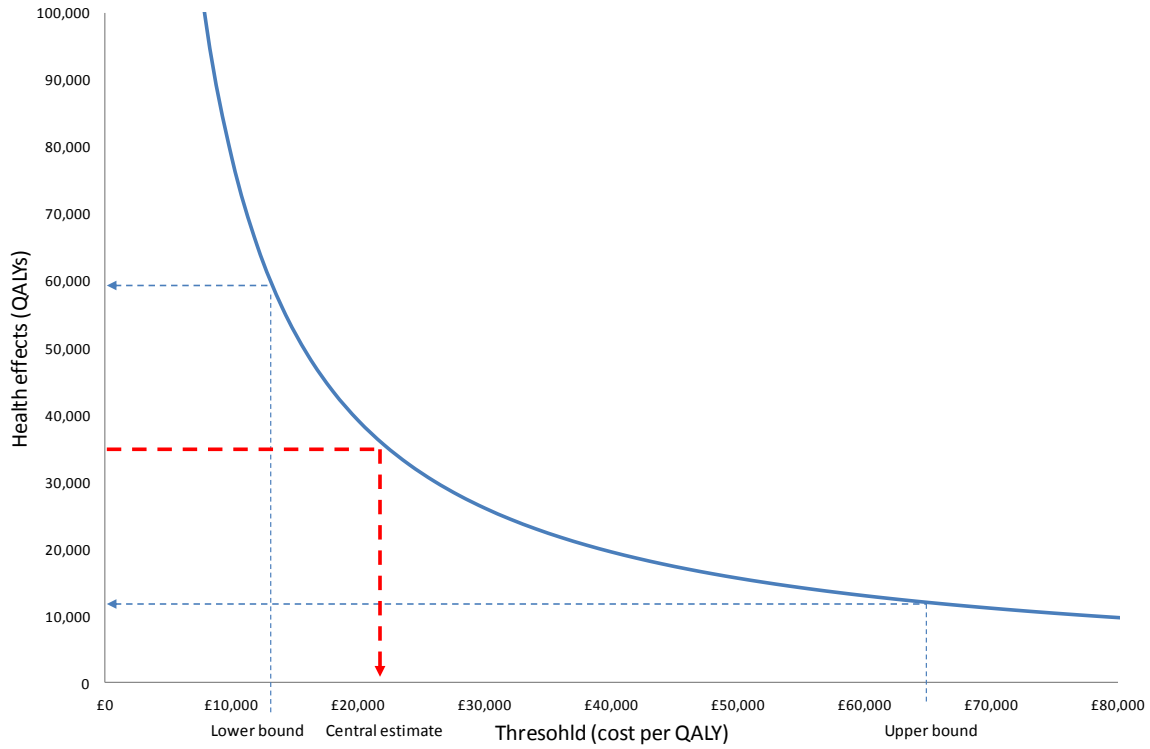
Therefore, this examination of the OHE critique and the analysis reported in 'Implications of alternative assumptions and judgements on the central estimate of the cost-effectiveness threshold', suggest that the overall policy implications of this research - that the thresholds currently used by NICE are too high and there are certainly no grounds to increase them - are robust to even extreme assumptions that lack credibility. There is little to recommend continuing to apply thresholds that are higher than the balance of evidence suggests. It will mean that more harm than good will continue to be done to NHS patients' health overall, because technologies will be approved even though existing evidence suggests that the health opportunity costs associated with their additional costs will exceed the health benefits they offer.

References

- Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith PC, Sculpher M. Methods for the estimation of the NICE cost effectiveness threshold. CHE Research Paper 81, November 2013. http://www.york.ac.uk/che/research/teehta/thresholds/
- Barnsley P, Towse A, Karlsberg Schaffer S and Sussex J. Critique of CHE Research Paper 81: Methods for the Estimation of the NICE Cost Effectiveness Threshold. Office of Health Economics, Occasional Paper 13/01 (December 2013).
- Claxton K, Sculpher M, Palmer S, Culyer AJ. Causes for concern: is NICE failing to uphold its responsibilities to all NHS patients? Health Economics 2015;24(1):1–7.

- Dakin, H., Devlin, N., Feng, Y., et al. (2014), 'The Influence of Cost-Effectiveness and Other Factors on NICE Decisions. DOI: 10.1002/hec.3086', Health Economics.
- Implications of alternative assumptions and judgements on the central estimate of the cost-effectiveness threshold. http://www.york.ac.uk/che/research/teehta/thresholds/
- Why is the central estimate of the cost per QALY threshold lower? http://www.york.ac.uk/che/research/teehta/thresholds/

Figure 1. Impact of alternative upper and lower bounds on the central estimate of the threshold



Legend for Figure 1

Figure 1 illustrates the health effects of a 1% change in NHS expenditure (£784m in 2008/09) for a range of values of a cost-effectiveness threshold.

A lower bound for the threshold of £12,936 per QALY suggests health effects of 60,605 QALYs
An upper bound of £29,414 per QALY suggests health effects of 26,653 QALYs
If these lower and upper bounds are regarded as equally likely the average of these health effects is 43,629 QALYs
Therefore the central estimate of the threshold would be £17,969 per QALY (£784m/43,629), not, £21,175 per QALY (the average of the upper and lower threshold values).

A lower bound for the threshold of £13,724 per QALY suggests health effects of 57,125 QALYs
An upper bound of £67,664 per QALY suggests health effects of 11,586 QALYs
If these lower and upper bounds are regarded as equally likely the average of these health effects is 34,356 QALYs
Therefore the central estimate of the threshold would be £22,820 per QALY (£784m/34,356), not, £40,694 per QALY, (the average of the upper and lower threshold values).

Response to the OHE critique

Our response to the issues raised follows the key claims made in the executive summary of the OHE critique.

Key assumptions

Quality of programme budget data reported by PCTs
The quality of programme budgeting category (PBC) data is important. We report a review of the evidence that was available about its quality for our purposes (at the aggregate level PBCs) and concluded that, as well as being the best data currently available, they were also sufficiently reliable to be meaningful for our purposes (see Section 3.4). The statistical models where developed using 2006/07 expenditure data and provided consistent and credible estimates when applied to subsequent waves of expenditure data. We carefully examine any changes in national (Department of Health) guidance on allocation of expenditure to PBCs in 2006/07, 2007/08 and 2008/09 (see Section 3.6.4). The change that took place after 2006/07 means that we believe that the 2007/08 and 2008/09 data reflect a better allocation of expenditure to PBCs. However, changes in national guidance on allocation of spend to PBCs over time does not undermine the cross sectional reliability PBC data.

Predicting the threshold
We have not constructed or estimated a predictive model, nor do we infer changes over time from the cross section. We discuss the possibility of developing a predictive model in the case for support (available on CHE web) and make clear the considerable challenges, which include access to sufficient panel data with sufficient observations in the cross section and time variant instruments, none of which was available to us (see Section 5.8).

Our analysis estimates the expected health effects if the NHS had less or more resources. This provides an estimate of the threshold in the period when the expenditure took place. A different question is how this might inform a judgement about an appropriate threshold in the next and subsequent periods. This requires consideration of other things that might change over time and what we know about their effects. These issues are examined in Section 5.6. We make no argument or assumptions about whether or not the threshold will rise or fall over time when budget, prices, productivity change.

What we report are estimates for 2007/08 and 2008/09 waves of spending which show no evidence of any increase at a time when real budget and prices rose. We discuss at length in Section 5.6 why this might be so, including examining evidence of improvements in allocative efficiency (changes in allocation between PBCs) as well as technical efficiency (increased heath effects in some PBCs). In the absence of data that would allow estimation of a predictive model our interpretation is careful and measured. This, combined with our subgroup analysis of PCTs that are under or over their target allocation in Section 5.5, gives some support to a view that the threshold is more likely to fall than rise when real budget growth is flat or falling and PCTs find themselves under increasing financial pressure. This analysis also suggests that the threshold for NICE decisions which impose costs on the NHS should be lower than the central estimate.

Model of PCT behaviour

Our intention was not to develop and test a theoretical model of PCT behaviour. Nor is such a theoretical model required to estimate the effect of changes in NHS expenditure on health outcomes. We do not need (nor do we) to assume QALY maximisation by PCTs. The interpretation of differences in the implied PBC thresholds is covered at length in Section 5.3. The danger of misinterpreting these differences as evidence of misallocation of resources is carefully discussed. Importantly, we were not able to estimate outcome equations simultaneously, due to insufficient observations (PCTs) in the cross section to examine possible interactions. This also means that total health effects are being underestimated and the threshold overestimated (see item 6 in Box 5.1 and references to text and foot notes). Not assigning health effects to changes in expenditure in PBC23 (general medical services), which accounts for more than 10% of a change in expenditure, is especially conservative in this respect.

Using the observed effect of changes in expenditure on the mortality burden of disease as a surrogate for a more a compete measure of health effects (effect on QALY burden) is summarised in Items 8-10 in Box 5.1. This provides the references to the other sections of the report where this is discussed in more detail. We do explore how the same statistical methods can be applied to the quality of life data that are being collected. This demonstrates that there is an effect of changes in expenditure on quality of life outcomes (see Section 5.8 and Table B8.26 in Appendix B). The implied PBC threshold for the key PBC where mortally effects could not be estimated (Mental Health) is consistent with what is known about the cost-effectiveness of the investment and disinvestments that were available (see Section 5.3 and Addendum C3). Although this is not direct evidence, it does give some indication that the implications of this assumption may not be unduly optimistic. Simply assuming that NHS expenditure has no impact on quality of life outcomes simply because they are not currently routinely collected and reported across all PBCs at PCT level does not seem at all credible. Nonetheless, even if it was assumed that there were no health effects of changes in expenditure in the 11 PBCs where outcomes equations could not be estimated, the threshold would be £16,569, i.e., substantially lower than the current thresholds used by NICE even when making this extreme assumption.

Alternatively, of course one could focus only on those PBCs where it is possible directly to estimate outcome elasticities. This would result in a much lower threshold of £8,308 per QALY (see Table 5.1 in Section 5.2). Our efforts to use what can be observed to inform what currently cannot by using as much information as is available (e.g., using information on QALY burden of disease), increased the central estimate of the threshold from £8,308 to £12,936 per QALY.

*Allocation of changes in expenditur*e

Expenditure elasticities are estimated for all 23 PBCs for the first time in Section 4.4.2 (see column 2 of Table C8 in Appendix C). However, it is not possible to estimate expenditure equations for all 23 PBCs simultaneously because there are not enough observations in the cross section. Consequently the 23 independently estimated expenditure elasticities do not account for all of a change in overall spend. Previously, the remaining change in total spend was not allocated to PBC 23 or the 11 PBCs where outcome elasticities could be estimated, but assigned to the other 11 PBCs. As a consequence, proportionally more of the share of a change in total spend was allocated to these

other 11 PBCs which tend to have a lower QALY burden and, therefore, higher implied PBC cost per QALY ratios.

However, there is no evidence that the estimates of expenditure elasticities are less reliable for those PBCs where outcome equations could not be estimated. Therefore, a more credible and less extreme assumption was to allocate any residual expenditure between all 23 PBCs reflecting their relative share of changes in expenditure based on their estimated expenditure elasticities (see column 4 of Table C8 in Appendix C). This more credible allocation of residual expenditure is also consistent with the observed stability in mental health spend as a proportion of total expenditure.

The previous extreme assumption that all residual spend should be allocated to the 11PBCs where outcomes equations could not be estimated results in a higher overall threshold of £14,755 (rather than £18,317 once the previous coding error is corrected, see 'Why is the central estimate of the cost per QALY threshold lower?'). An alternative, equally extreme, assumption would be to allocate all residual expenditure to PBC 23 and the 11 PBCs where outcome elasticities could be estimated. This would result in an overall threshold that would be lower than our central estimate of £12,936. Therefore, we believe that we have taken the most reasonable balanced approach to allocating changes in expenditure across all 23 PBCs (See 'Why is the central estimate of the cost per QALY threshold lower?')

Other assumptions

1. The net effect of restricting health effects of changes in expenditure to one year is conservative because no credit at all is given for the future health benefits of current spending (see Sections 4.1, 4.2.5, 4.4.3, 5.4 and 5.8). Some health effects of past spend may be wrongly picked up if the variation in these health effects happens to be correlated with variation in current spend (once it has been controlled by other covariates). This is not the same as correlation in expenditure over time. Since we are excluding all the future effects that should be included but only by chance wrongly picking up some of the effects of past spend the net effects of are conservative with respect to health effects. This is discussed at some length (see 5 in Box 5.1 for summary and references to text and footnotes).

2. We make very clear throughout that assuming that a death averted returns the individual to the mortality risk of the general population, matched by age and gender, is likely to be optimistic with respect to health effects (see 1 in Box 5.1. and the references to text and footnotes). Unlike other approaches to calculating years of life lost due to disease, which implicitly make this same assumption (e.g., ONS and WHO), we do take account of counterfactual deaths (see Section 4.2), which reduces the effect on life years of the estimated changes in mortality. We explain the profound empirical challenges of estimating the full survival effects of changes in expenditure in Section 5.8. When considering how optimistic this assumption might be (estimates imply 4.5 life years per death averted on average) it should be noted that we are using 3 years of mortality data which is first averaged before being used in the econometrics. This means our estimate of mortality effects does not include deaths averted in patients who survive for less than one year (in the 3rd year of the mortality data), less than 2 years (in the 2nd year) and less than 3 yrs (in the

1st year). Such averted deaths in these patients are not captured in our estimate of mortality effects (see Section 4.1 and 4.2.5 as well as 1 in Box 5.1 and references to text and footnotes). This also explains why 2 life years for each death averted is less than the minimum possible life year effect associated with the type of deaths estimated to be averted based on these data and why the estimate of £29,314 in Table 5.1 Section 5.2 represents an extreme upper bound.

3. The central estimate of £12,936 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 £13,141 (see Section 5.2 and footnote 96). The effects of discounting are modest because the health effects of changes in expenditure are restricted to one year. A large proportion of this health effect is quality of life (which occurs in that year so is not subject to discounting, e.g., see Table 5.4 in Chapter 5). The change in mortality due to a change in spend that occurs in that year do have life year effects (adjusted for quality) in subsequent years which are be subject to discounting. Some changes in mortality will have life year effects over many years and other mortality effects will not. On average 4.5 life years is associated with each death averted (see Table 5.1 in Section 5.2,) so on average the effect of discounting is modest. Of course, if health effects in future periods could be estimated using an appropriate lag structure (see Section 5.8) then they should be discounted. However, overall this is likely to capture more total discounted health effects of changes in expenditure, reducing rather than increasing the estimate of the threshold.

Additional issues

We do not argue that the effect on the threshold of increases in the NHS budget will necessarily be offset by changes in productivity. We make no argument or assumptions about whether or not the threshold will rise or fall over time when budget, prices, productivity change. What we report are estimates for 2007/08 and 2008/09 waves of expenditure, which shows no evidence of any increase at a time when real budget and prices rose. We discuss at length in Section 5.6 why this might be so, including examining evidence of improvements in allocative efficiency (changes in allocation between PBCs) as well as technical efficiency (increased heath effects in some PBCs). In the absence of data that would allow estimation of a predictive model, our interpretation is careful and measured. This, combined with our subgroup analysis of PCTs that are under or over their target allocation in Section 5.5, gives some support to a view that the threshold is more likely to fall than rise when real budget growth is flat or falling and PCTs find themselves under increasing financial pressure.

The implications of uncertainty in the mean estimate of the threshold (£12,936 per QALY) are as a consequence of its non linear relationship with net benefit (see Section 5.4). The only circumstances when uncertainty in the estimate of the threshold would have no implications for a policy threshold would be if a technology had zero incremental costs. The only circumstances in which the skewness of the distribution might tend to offset the effects of this non linearity, for a technology with additional costs, is a negative skew (i.e., when values of the threshold greater than the mean are more likely) which was not found (see Section 5.4). The general implications of this research - that

the thresholds currently used by NICE are too high and there are certainly no grounds to increase them - does not rest on this discussion. Indeed, we did not quantitatively account for the effect of uncertainty or the implications of the analysis detailed Section 5.5 in our central estimate of £12,936, despite these suggesting that an appropriate threshold for NICE is likely to be less than our central estimate. This demonstrates that our analysis, rather than having "an overall downward bias", has instead erred on the side of conservatism with respect to both the health effects of changes in expenditure and the implications for a policy threshold.

Further research

Our views and reasoning about the need for further research and data collection are set out in Section 5.8. We are not convinced that collecting data on how PCTs (now CCGs) make decisions is a priority, primarily because it is not required to estimate the health effects of changes in NHS expenditure and has been shown to be unsuccessful approach to estimating the threshold in the past (see Section 5.8, 2.3.3 and Addendum C2). We believe it is important periodically to re-estimate the threshold and to improve estimates as additional data become available; notably additional waves of expenditure and mortality data as well as patient reported outcome measures. We also set out why assumptions and judgements will still be necessary even as more data become available. An appropriate threshold that represents the expected health opportunity costs of NICE decisions will inevitably be a scientific judgment based on the balance of existing evidence.

Implications for policy

In common with four independent peer reviewers, we remain confident that our analysis has made best use of existing evidence, that our interpretation is careful and balanced, with the implications of the necessary assumptions explored and clearly stated. As such, it provides the best evidence of the scale of the health opportunity costs incurred when NICE guidance imposes costs on the NHS. Our analysis, rather than having "an overall downward bias" has instead erred on the side of conservatism with respect to the health effects of changes in expenditure and the implications for a policy threshold for NICE.

We have set out the implications of alternative judgements about which assumptions would provide extreme but equally likely upper and lower bounds for the central estimate of the threshold (see Figure 1 and 'Implications of alternative assumptions and judgements on the central estimate of the cost-effectiveness threshold'). This illustrates that, even if very pessimistic assumptions about the health effects of changes in NHS expenditure are regarded as credible, the overall policy implication - that the current thresholds being used by NICE are too high - remains unchanged. There is little to recommend continuing to apply thresholds that are higher than the balance of evidence suggests they should be. It will mean that more harm than good will continue to be done to the health of NHS patients overall, because technologies will be approved even though existing evidence suggests that the health opportunity costs associated with their additional costs will exceed the health benefits they offer.