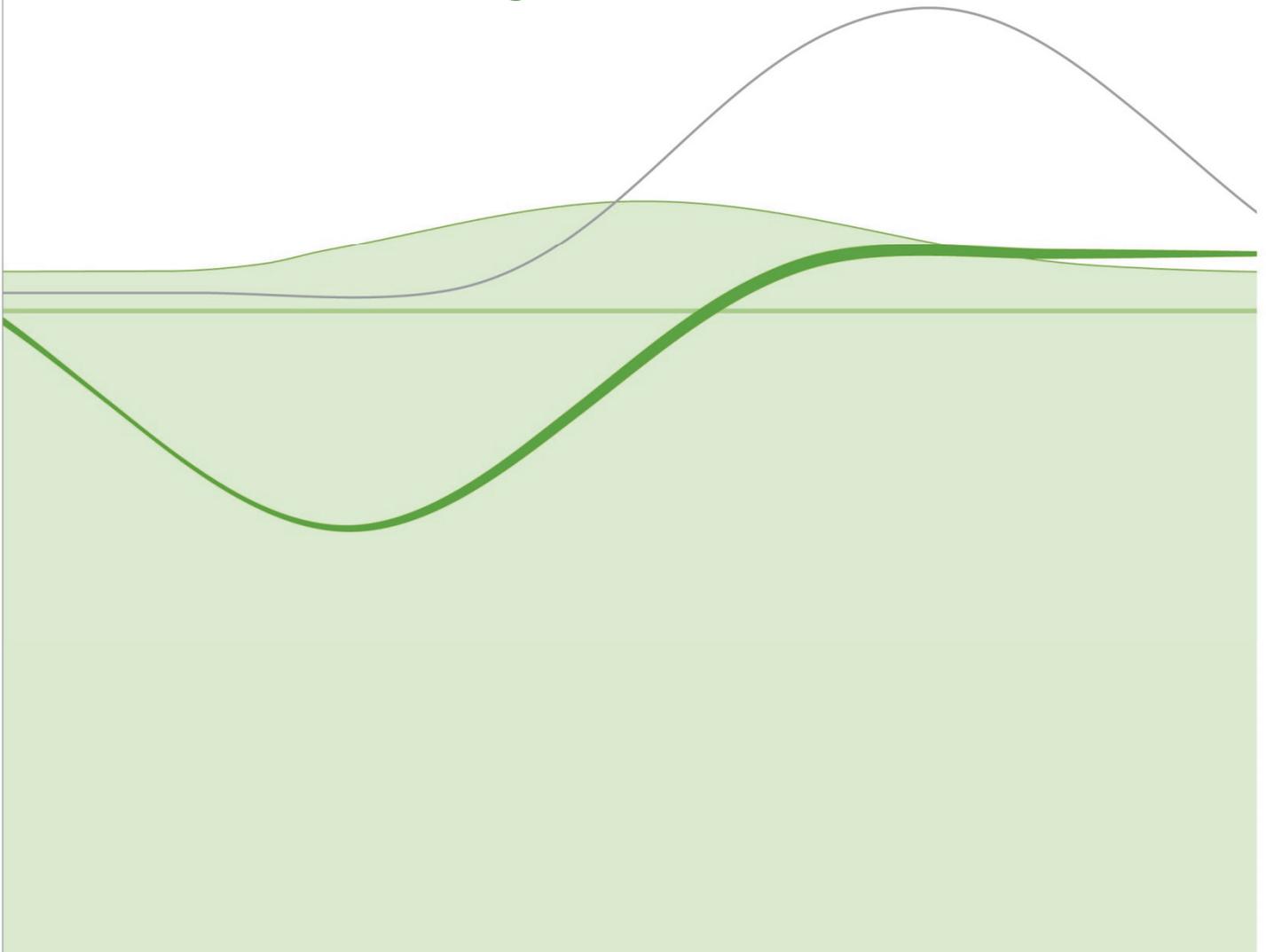


# Critique of CHE Research Paper 81: *Methods for the Estimation of the NICE Cost Effectiveness Threshold*

December 2013

Paul Barnsley, Adrian Towse,  
Sarah Karlsberg Schaffer and Jon Sussex



# Critique of CHE Research Paper 81: *Methods for the Estimation of the NICE Cost Effectiveness Threshold*

Paul Barnsley, Adrian Towse,  
Sarah Karlsberg Schaffer and Jon Sussex

Office of Health Economics

December 2013

Occasional Paper 13/01

For further information please contact:

Adrian Towse

[atowse@ohe.org](mailto:atowse@ohe.org)

Office of Health Economics

Southside, 105 Victoria Street

London SW1E 6QT

United Kingdom

Tel: +44 207 747 1407

©Office of Health Economics

## **About OHE Occasional Papers**

OHE Occasional Papers are intended to provide information on and encourage discussion about a topic.

Any views expressed are those of the authors and do not necessarily reflect the views of, or approval by OHE, its advisory boards or any sponsors of the research.

## **Acknowledgements**

The authors would like to thank the following for comments on an earlier version of this paper: a member of the OHE Editorial Board; an external expert reviewer; several of the authors of CHERP 81, who gave us comprehensive feedback. The authors of this OHE Occasional Paper are solely responsible for any remaining errors.

## **Funding**

The Office of Health Economics (OHE) receives research and consulting funding from the Association of the British Pharmaceutical Industry (ABPI). No separate funding was received for this paper.

## TABLE OF CONTENTS

Table of contents .....	iii
Executive summary .....	iv
Introduction .....	1
Overview of CHERP81 methodology .....	3
PBC data quality .....	8
Estimating opportunity cost based on differences between PCTs .....	10
What are PCTs aiming to achieve? .....	13
Conservative and optimistic assumptions in estimating the value of the threshold .....	19
The impact of changes to expenditure and productivity on the estimated threshold .....	37
How should we address uncertainty about the threshold? .....	45
Optimal policy threshold where there are non-marginal effects .....	49
Other approaches to investigating the threshold .....	52
Ways forward .....	56
References .....	61

## EXECUTIVE SUMMARY

### Context

This report is a critique of the Centre for Health Economics Research Paper No. 81 (CHERP81), which estimates that the marginal cost of a QALY for the NHS in England in 2008/9 was £12,936. There is much to admire in the approach, detailed analysis, and reporting set out in CHERP81 and we recommend detailed reading of it. The focus of this OHE Occasional Paper is, however, on the points where we take issue with or have queries about the CHERP81 work.

The magnitude of the marginal cost of a QALY in the NHS is important as it informs the cost effectiveness threshold used in health technology assessments (HTAs) by the National Institute for Health and Care Excellence (NICE). Indeed this was the motivation for the commissioning of the study by the Medical Research Council (MRC) as part of a programme of methodological research relevant to NICE. Understanding the expected QALYs displaced in the NHS by a NICE recommendation is an important part of helping NICE make optimal decisions in relation to the approval of new technologies. The CHERP81 estimate for 2008/9 is lower than NICE's announced threshold range for approving new products of £20,000 to £30,000 per QALY.

### The CHERP81 analysis

The CHERP81 report is a complex and impressive attempt to arrive at a usable estimate of this displacement effect. Given the incomplete nature of the data available, however, CHERP81 is forced to rely on a series of strong assumptions in extrapolating from what we can observe – changes in expenditure and in *mortality* in *some* Programme Budget Categories (PBCs) *across* different Primary Care Trusts (PCTs) – to what we really want to know, which is what happens when NICE decisions impact the amount of budget available for PCTs to spend on other treatments. This is the impact of changes in expenditure on changes in *mortality and quality of life across all PBCs within* the same PCT *over time*. While we understand the necessity of making assumptions in order to arrive at an estimate of the threshold, the number of assumptions and their importance means that the estimate presented is highly uncertain and sensitive to the adoption of plausible alternative assumptions. It also means that the estimate can be improved by the collection of additional data over time.

### Key assumptions

CHERP81 arrives at its estimate of the threshold by estimating the magnitudes of statistical links between changes in overall budgets and changes in expenditure on particular PBCs, and between changes in PBC expenditure in a particular PCT and changes in observed mortality in the same area. There are a number of challenges in using this approach.

First, because the accounting assumptions used to allocate spending across the different PBCs may differ across PCTs or through time, apparent differences in expenditure used to estimate the value of extra spending could be the result of something as mundane as differences in book keeping practices.

Second, the threshold is a prediction about how spending and health outcomes will change if individual PCTs' budgets change in different ways over time, but the data used by CHERP81 only tells us about the behaviour and outcomes of different PCTs at a single point in time. CHERP81 assumes it is possible to predict a PCT's response to a new budget by looking at what other PCTs that already have a budget of that size spend their money on and the results they achieve. This may not be the case.

Third, and linked to the second point above, CHERP81 does not have a testable model of how PCTs actually behave or what their goals are. This is fine as long as the analysis relies on behaviour which we can actually observe, such as reductions in mortality. But it becomes a problem when estimating how good PCTs are at the unobservable parts of their job, such as quality of life improvement. There are, broadly, three ways PCTs might behave. They may be QALY maximisers, in which case they do not seem to be very efficient in some programme areas where CHERP81 finds the cost-per-QALY to be very high. PCTs may, alternatively, be trying to maximise the achievement of a broader measure of health gain or societal benefit than just the QALY; in which case CHERP81's approach of modelling productivity of non-mortality-reducing PBCs by assuming productivity equal to that of mortality-reducing PBCs estimates the marginal cost of a unit of health gain, not that of a QALY, as activity in these PBCs is likely to be directed at achieving other health benefits as well as QALYs. Or, thirdly, PCTs may focus on reducing mortality (which is readily measurable) and be less efficient at reducing morbidity. In this third case, assuming quality of life gains are achieved at the same rate as quantity of life gains (as CHERP81 does) will lead to an overestimate of likely quality of life gains and an understatement of the value of the threshold.

Fourthly, CHERP81's econometric model of PCT expenditure explains only 72% of changes in PCT budgets due to a change in the overall NHS budget. The remaining 28% is assumed to be allocated across all PBCs in proportion to their estimated responsiveness to changes in budget. CHERP81's central estimate of the threshold is extremely sensitive to this assumption; assuming that the funds not captured by their model accrued only to PBCs which do not reduce mortality led to an earlier estimate of the threshold (in the June 2013 version) of £18,317, which was 42% higher than their current £12,936 figure.

Fifthly, CHERP81 makes a number of assumptions which, when taken together, may in our judgement understate the true value of the threshold:

1. CHERP81 argues that its decision to ignore the fact that current spending can reduce future as well as current mortality tends to overstate the threshold (i.e. spending in 2008 impacts mortality in years beyond 2008). This is true, but, it also ignores the potential for past spending to influence the current mortality figures used to estimate the threshold, which will be important if PCTs show some

consistency in how they spend money through time. We think, therefore, that this assumption has less effect on the threshold than CHERP81 suggests.

2. CHERP81's assumption that patients whose death is averted by extra spending subsequently enjoy the life expectancy of average members of the population overestimates QALYs gained elsewhere in the NHS. CHERP81 recognises this, but argues that its assumption in relation to 1. above offsets this. We are of the view that it does not.
3. CHERP81 does not, in our view, accurately discount the benefits used to estimate the threshold. The rate at which NICE requires health gains to be discounted in HTAs is 3.5% per annum. Where CHERP81 does provide a discounted estimate of the threshold it understates the required level of discounting by assuming that all unobserved quality of life gains from NHS treatments are enjoyed immediately, instead of over a number of years like the observed improvements in mortality.

### Additional issues

Finally, there are two further issues raised by CHERP81's analysis:

1. CHERP81 recognises that the value of the threshold will tend to increase with the NHS budget due to diminishing marginal returns to health spending. In other words, as the budget goes up in real terms the NHS will start to provide services that it could not previously afford (as it did not regard them as providing such good value, i.e. having a higher cost-per-QALY) when the budget was lower. However, it argues that this effect will be offset by rises in NHS productivity, which will reduce the cost of providing existing technologies and services. CHERP81 argues this should reduce the value of the threshold. We disagree. While budget increases will increase the threshold, there is no reason to expect productivity increases to decrease it. Improving productivity increases how much health the NHS can produce with a given amount of money *on average* but that means they also increase *how much* health the NHS is producing, which leads to diminishing marginal returns, i.e. the provision of additional treatments that have a higher cost-per-QALY.
2. CHERP81 argues that uncertainty in the estimate of the threshold means that we ought to use a lower threshold than the central estimate when assessing new technologies. As a £1 overestimate leads to more QALY losses than a £1 underestimate, it concludes that it is better to err on the side of caution and underestimate the threshold. The importance of this effect increases with the potential for the "true" underlying threshold to approach zero. This effect assumes, however, as CHERP81 recognises, that the average *expected size* of errors in each direction over and under the threshold, (i.e. the likelihood of a £1 overestimate or £1 underestimate occurring), is the same. The expected errors around the CHERP81 threshold estimate are not equally sized. There is a positive skew, which means that there is a small probability that the threshold may be substantially underestimated, a larger probability that the threshold has been slightly overestimated but almost no chance that it has been substantially overestimated. We agree that uncertainty makes a case for using a lower

threshold than the central estimate, but the effect is reduced by the positive skew of potential errors.

## Further research

A number of the concerns we have with the CHERP81 approach can be addressed by gathering additional data. As the report notes this should include:

1. More time-series data on PBC expenditure and time-series data on mortality to directly estimate the effect of observed budget changes;
2. Data on changes in morbidity over time. Getting better data on patients' quality of life could, for example, be achieved through the expansion of the programme of routine collection of Patient Reported Outcomes Measures (PROMs). As well as improving estimates of the threshold NICE should use, such data would also improve the ability of the NHS to understand how well it is performing.

Additional data should also be collected on how individual PCTs say they make their budgetary decisions. We detail a potential methodology for collecting and using such data, outline some of the potential difficulties and discuss the CHERP81 view on alternative methods. Such an approach can be used to complement extrapolation from observed differences between different PCTs by combining improved understanding of how decisions are made with data showing how expenditure and outcomes are estimated to have changed over time.

## Implications for policy

CHERP81 describes an impressive and thought-provoking attempt to answer a difficult question. The assumptions on which it is based and the economic principles governing its application as a policy tool require further consideration. CHERP81 works within the limitations of the available data, and we believe that the uncertainty of the estimate provided by CHERP81 is sufficiently large so as to render it a questionable guide to changes in policy until more data are collected. Such additional data would hopefully serve to narrow the range of values for the key elements of the threshold prediction and to justify or refute certain key assumptions CHERP81 has relied upon. The difficulties evident in CHERP81 underline the importance of attempts to directly identify marginal services at a PCT level.

In the interim, we believe that an overall downward bias has been introduced into the CHERP81 "best estimate" of the marginal cost of a QALY in 2008/9 by a number of the assumptions made in its estimation. This could be addressed using existing data in most cases by changing assumptions to give a better sense of the degree of uncertainty of the current estimate. Such an exercise would highlight the importance of collecting additional data to address the degree of uncertainty before an estimate derived using this methodological approach could be confidently used in policy making. Decision makers will be able to have more confidence in the applicability of a revised estimate to policy making. For these reasons, we do not think use of the figure of £12,936 per QALY in NICE decision making is currently justified. A revised estimate may be higher or lower.

We suspect, on the basis of the analysis we present here, that it will be substantially higher but we may be wrong.

## INTRODUCTION

This paper critiques a University of York Centre for Health Economics Research Paper revised in November 2013 entitled "Methods for the Estimation of the NICE Cost Effectiveness Threshold. Final Report" (which, following the authors, we will refer to as "CHERP81"). It is downloadable at

<http://www.york.ac.uk/che/research/teehta/thresholds/>

CHERP81 sets out the findings of a major research project concluding that in 2008/9, the latest year analysed, the "central or 'best'" value of the cost per QALY threshold of the NHS in England "is estimated to be £12,936 per QALY". (paragraph 3.1, page viii)

This replaces an earlier, June 2013, version of the paper, which arrived at a central estimate of the cost per QALY threshold of £18,317 on the basis of a different assumption about the allocation of spending not predicted by the CHERP81 model<sup>1</sup>.

Given the importance of the threshold opportunity cost concept and the sensitivity of the overall estimates to key assumptions made in CHERP81, we have undertaken a critical review of the research it presents. There is much to admire in the approach, detailed analysis, and reporting set out in CHERP81 and we recommend detailed reading of it. The focus of this OHE Occasional Paper is, however, on the points where we take issue with or have queries about the CHERP81 work.

CHERP81 argues that the relevant threshold cost of a QALY is the marginal opportunity cost of a QALY in the NHS: how much has to be spent at the margin to gain an extra QALY. This determines how many QALYs would have to be displaced to find room for increased spending mandated by NICE. We agree that this "opportunity cost" approach is theoretically correct, and should be used, provided we are sure that the health care "displaced" by a spending decision can be predicted in practice, at least on average.

The research underlying CHERP81, which was funded by an MRC Methodology Research Programme grant, represents a thorough development and application of one method for evaluating the size of this opportunity cost. We discuss alternative options for modelling NHS responses to budgetary changes. CHERP81 emphasises the attractiveness of basing estimates of cost per QALY on routinely collected data. We agree, although improving the accuracy of the estimate will, as the authors note, require additional morbidity data to be routinely collected.

CHERP81 is an impressive piece of work and represents an important contribution to the debate surrounding the optimal value of the threshold to be applied by NICE. The authors have skilfully handled an enormous quantity of data and considered a range of practical and theoretical challenges. However, on balance the CHERP81 estimate of the value of the threshold cost per QALY may have been biased downwards by the series of

---

<sup>1</sup> The June 2013 version of CHERP81 was itself a revised version of the original January 2013 version of the paper. The June 2013 version took account of comments from peer-reviewers of the January 2013 version. The central estimate of £18,317 did not change as between these two versions.

assumptions made in the course of its estimation. In addition, while we appreciate that the analysis in CHERP81 has to work within the limitations of the available data, we believe that the uncertainty of the estimate provided by CHERP81 is large enough to render it an unreliable guide for policy making until more data are collected. Additional data would hopefully serve to narrow the potential range of key variables and to justify (or not) certain assumptions relied on for its production. Ideally the analysis of additional data would also be combined with looking at different approaches to modelling the behaviour of PCTs (now CCGs). We set out the basis for our conclusions throughout the remainder of this paper.

**Note: How to read this paper**

A non-technical summary of the material presented in each part of this report is provided in the opening "Summary" section. This covers almost all of the material contained in the report, but is not intended to provide technical detail or extensive justification for the conclusions presented in CHERP81. A more detailed and technical discussion of the issues, including extensive references to the portions of CHERP81 being discussed, is included in each section under the heading "Analysis". This construction is to enable readers to begin with the summaries of each section and only delve into the analysis as they wish.

## OVERVIEW OF CHERP81 METHODOLOGY

### Summary

CHERP81 draws on available data showing differences between Primary Care Trust (PCT) budget allocations across PCTs and differences in mortality for some of the broad Programme Budget Categories (PBCs) which are used to divide spending across different health care areas by the NHS in England. The CHERP81 model controls for differences in input costs and level of need. It draws on these data to estimate:

- How much, on average, expenditure on each PBC will change in response to a change in overall PCT budgets – the elasticity of expenditure for each PBC; and
- How much, on average, mortality in each PBC will change in response to a change in spending on that PBC – the elasticity of mortality for each PBC.

Combining these two sets of values, CHERP81 estimates for some PBCs how many additional (fewer) deaths will be averted as a result of increasing (reducing) a PCT's expenditure on these PBCs by a certain amount. It also then predicts the effects of changes in expenditure on deaths in the remaining PBCs for which there is no mortality data. It then makes assumptions about the future life expectancy of patients whose deaths are averted in order to convert these into life years saved.

This gives an estimate of incremental cost per life year saved by PBC. In order to calculate an incremental cost per QALY, rather than per life year, CHERP81 has firstly to convert life years saved into QALYs saved, which it does by assuming these patients experience improvement in typical quality of life for sufferers of the disease in question proportional to the calculated improvement in reduced life expectancy. Secondly, it has to estimate QALY gains from improvements in morbidity that are independent of those arising from changes in mortality. It does this by assuming that PCT expenditure on any given PBC reduces the disease burden arising from poor quality of life by the same proportion as it reduces the QALY burden caused by premature mortality. The estimated QALY impact will thus depend on the proportion of disease burden in a PBC area that is due to poor quality of life and the proportion that is due to premature mortality.

### Analysis

The method in CHERP81 involves three main steps:

1. Econometric analysis to estimate the relationship between differences in PCT spending across different PBCs and mortality from diseases covered by those PBCs, using instrumental variables to adjust for the confounding effects of the NHS funding formula, which gives more money to PCTs with higher mortality;
2. Translating the resultant (negative) effects of expenditure on mortality (i.e. reduced mortality) into (positive) effects expressed as life years gained;
3. Adjusting the effect of spending on life years gained to account for: (i) quality of life in the extra life years, and (ii) spending which impacts on morbidity not mortality.

This gives a relationship between differences in expenditure and in mortality and morbidity averted (measured in QALYs) by individual PCT.

The process in a) is very similar to earlier work by some of the authors, in particular Martin et al (2008), but CHERP81 goes further in developing the approach:

- It uses data from spending on *all* parts of the NHS whereas the previous work used the “big four” PBCs (cancer, circulatory, respiratory, gastro-intestinal) and diabetes;
- There is a more extensive focus on adjusting the estimates to reflect quality of life effects as well as life years gained, to give QALYs gained;
- It uses mortality averages across three years’ data, not just a single year, in order to give more stable estimates of mortality.

There are four disease areas (cancer, circulatory, respiratory and gastro-intestinal) with good mortality data at PCT level. Together they accounted for 25% of total expenditure in 2008/09. Another seven PBCs can be linked to mortality data, albeit less completely, giving 11 PBCs with some linked mortality data. These 11 PBCs comprise 48% of total expenditure<sup>2</sup>. The health impacts of 10 of the remaining 12 PBCs are extrapolated from the 11 PBCs. For the remaining two PBCs (Social care needs” and “Other) no health impacts are assumed. The latter is primarily General Medical Services, i.e. GPs.

The mortality data are used to arrive at an estimate of years of life lost (YLL) via a series of elegant adjustments to the available data sets in order to initially estimate YLL up to the age of 75 (standardised mortality data goes up to 75) and then to take account of YLL from mortality reductions among populations over 75. This included adjusting for “counterfactual deaths” which would have occurred independently of the impact of expenditure on mortality from the diseases in each PBC. In calculating YLL any death averted by expenditure in one year is assumed to return the individual to the mortality risk of the general population, which gives an average of 4.5 additional years of life gained per PBC death averted. An alternative “upper bound” is calculated assuming only 2 additional years of life gained per PBC death averted.

### What is the econometric analysis used for?

The aims of the econometric analysis are:

1. **To derive an “expenditure elasticity” for each PBC**, i.e. by what percent spending on a PBC changes when there is a 1% change in the total NHS budget.
2. **To derive an “outcome elasticity” for each PBC**, i.e. by what percent the number of life years (not QALYs yet) changes with a 1% change in spending on that PBC.

CHERP81 uses these expenditure elasticities to derive weights for the individual PBCs’ costs per life year (and ultimately their estimated costs per QALY) so as to estimate an overall marginal cost per QALY across the 23 PBCs.

The following equation explains the link between the cost of an additional life year, the expenditure elasticity and the outcome elasticity.

---

<sup>2</sup> CHERP81 notes that these 11 PBCs do account for 78% of the overall health effects estimated in the report.

The cost of an additional life year in a care programme:

$$= \frac{\text{change in expenditure in the care programme}}{\text{change in life years lost in the care programme}} = \frac{\text{change in annual spending} * \text{expenditure elasticity}}{\text{annual mortality} * \text{mortality elasticity}}$$

Taking the PBC "circulatory problems" as an example:

1. Annual expenditure in 2006/7 was £6,161 million (note that the analysis in this section of CHERP81 is illustrated based on 2006/7 data, rather than the 2008/9 data from which the headline £12,936 figure has been derived);
2. Expenditure elasticity for this PBC is 0.54: this coefficient of changes in the total budget means that, on average, based on observed differences in expenditure between PCTs, where a PCT has a budget 1% greater than its neighbour, expenditure on circulatory problems will be 0.54% higher than that of its neighbour;
3. Change in expenditure is 0.54\*£6,161 million\*1%, which is equal to £33.27million. This means that a 1% increase in the total NHS budget is estimated to increase spending on circulatory problems by £33.27million;
4. The total life years lost for patients with circulatory problems under 75 years in 2006, 2007 and 2008 is 1,361,634 years; the average annual life years lost is 1,361,634/3;
5. Outcome elasticity is -1.434: the coefficient of expenditure for the 'Circulatory' PBC in the health outcomes model. This estimate is arrived at by noting that, on average and adjusting appropriately for need and input costs, PCTs which spend 1% more on circulatory conditions experience a 1.434% reduction in mortality from those causes;
6. Change in life years lost is -1.434\*0.54\*1,361,634/3\*1%, which is equal to -3,515 years. This means that a 1% increase in the NHS budget will decrease the total life years lost for this PBC by 3,515 years.
7. The cost of a life year gained under the circulatory problem budget category in 2006/07 is therefore £33.27million/3,515years, equal to £9,466 per life year saved.

The cost of an additional life year gained in circulatory problems in 2006/07 is therefore:

$$\begin{aligned} \text{cost of an additional life year in circulatory programme} &= \frac{£6,161 * 1\%}{0.54 * \left(\frac{1,361,634}{3}\right) * 1\%} \\ &= £9,466 \text{ per life year saved} \end{aligned}$$

The same calculation steps are applied to the other ten PBCs where good, or at least some, mortality data are available on a PCT basis (two of these PBCs, Maternity and Neonates, are combined into one category for the analysis). The raw numbers in the above example are from Table B8.20 of Appendix B, page 79.

Given the absence of mortality data for the remaining 13 PBCs and of quality of life data for any PBC, CHERP81 extrapolates its estimates for cost per life year from the 11 PBCs

for which mortality data are available to a cost per QALY estimate across all 23 PBCs using a proportional extrapolation approach we summarise below.

### **Extrapolating from cost per YLL to cost per QALY**

Cost per life year (based on net years of life lost, YLL) is then adjusted for quality of life in the additional years of life to get to QALYs gained from mortality effects. CHERP81 calculates these using three different assumptions about the quality of life of patients enjoy during additional life years: (i) the average quality of life of the rest of the (age and gender adjusted) population who do not have the disease, (ii) the quality of life reflecting the average patient with the original disease state; and (iii) the average disease state quality of life (i.e. (ii) ) increased to reflect reduced quality of life burden equiproportional to the estimated mortality burden reduction generated. CHERP81 opts for (iii), rejecting (ii) as “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 disease state until death.” (p59, third paragraph).

We now have a weighted average cost/QALY based on mortality effects. This is then adjusted for the relative contribution of changes in QALYs due to premature death and to morbidity in each PBC area. The burden of mortality and morbidity was derived from WHO Global Burden of Disease (GBD) data, adjusted to convert Disability Adjusted Life Year (DALY) ratios into QALY ratios based on EQ5D quality of life decrements estimated from HODaR and MEPS data, and then into QALY burden of disease estimates (also using HODaR and MEPS). QALY gains arising from reduced morbidity were thus assumed to be proportionate to QALY gains based on mortality effects. Thus, if a disease area delivers 1,000 QALYs from mortality improvement and mortality accounts for one third of the disease burden, then equiproportionate improvements in morbidity are assumed to deliver another 2,000 QALYs, giving a total gain of 3,000 QALYs.

Thus estimated mortality and life year effects for the 11 PBCs with mortality data were used, via a series of assumptions, to determine the, assumed proportionate, effects of premature death and of morbidity for all 23 PBCs. As CHERP81 puts it, the authors “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 for QALY burden [as in the WHO’s GBD report] for all PBCs.” (p66, first paragraph).

CHERP81 is not able to estimate all 23 expenditure elasticities simultaneously leading to a shortfall in the additional expenditure accounted for. CHERP81 assumes that the shortfall, which represents 28% of any change in overall budgets, is divided across all 23 PBCs in proportion to their calculated expenditure elasticities.

The result is CHERP81’s “best estimate” or “central estimate” of the threshold cost per QALY in the NHS in England in 2008/9 financial year of £12,936. This, and the other results highlighted by the authors are summarised in Table 1.

**Table 1. Headline results – cost per QALY estimates for 2008/9**

<b>Programme Budgeting Category (PBC)</b>	<b><i>Lower bound</i></b>	<b>Best estimate</b>	<b><i>Upper bound</i></b>
Big 4 with best mortality data	<i>£1,194</i>	£4,827	<i>£11,040</i>
All 11 with mortality data	<i>£1,175</i>	£8,308	<i>£18,827</i>
All 23	<i>£2,018</i>	<b>£12,936</b>	<i>£29,314</i>

Source: Table 5.1 of CHERP81, p74

## PBC DATA QUALITY

### Summary

The CHERP81 model relies upon differences in expenditure between different PBCs across PCTs, but if some of these differences are due to different cost apportionment rules or other accounting practices then the threshold estimate will not be accurate. CHERP81 has to take the expenditure data at face value since there is currently no way to determine which differences in spending are real or to assess the uncertainty involved in using the data. CHERP81 is clear about the assumptions it makes and we agree that they are both reasonable and unavoidable, but they do increase the uncertainty of CHERP81's threshold estimate in a way that cannot be quantified.

### Analysis

Section 3.4.1 and Appendix B section B4.2 of CHERP81 make clear there is a lot of allocation and apportionment involved in creating the PBC numbers:

*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...*

Appendix B, page 12, thus recognises the "simplistic" and "work in progress" nature of the PBC expenditure allocations. We note that since 2008/9 there have continued to be changes to the way PBC expenditure is estimated. Until 2009/10, PCTs calculated programme budgeting data using cost information prepared by provider organisations, in particular NHS Reference Costs. However, for 2010/11 and subsequently this has been fundamentally changed, with PCTs required to calculate programme budgeting data from the prices they paid for activity, rather than their providers' costs (DH, 2013). We do not know what scale or direction of impact this change would have on estimates for 2010/11 and later years using the CHERP81 approach.

Even after adjusting for need and geographical cost variations there are large differences across the PCTs in spend per head on any given PBC. Table 3.2 of CHERP81 shows, for example, that, even after adjusting for local variations in population need and unavoidable geographical cost variations, spend per head on Cancer varies threefold from £55 to £154 across PCTs, and spend on Circulatory varies more than fourfold from £76 to £327 per head (2008/9 figures). These large differences may reflect historical accident and inertia, major differences in explicit decisions made by different PCTs, data errors, or some combination of the three.

CHERP81 recognises this point Appendix B on page 14, citing an NAO (2008) report and another study (Appleby et al, 2011) which question the quality of the programme budget expenditure data:

*The variation in expenditure across PCTs has led some commentators to question the reliability of the programme budgeting data.... The National Audit Office [19] undertook a survey of Trusts, PCTs and SHAs... 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 improvements to the robustness of some of the data (e.g. non-admitted patient care).*

*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... 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 a case study..Appleby et al [20] report the results of West Kent PCT's use of an alternative approach to producing programme budgeting data for cancer and tumours. This alternative approach identified similar levels of expenditure ..at the aggregate level, but there were differences ..at the sub-programme level...*

(Appendix B, page 14 paragraphs 4, 5 and 6).

Appleby et al (2011) state clearly in the summary of their study:

*There is prima facie evidence of problems with the quality of the Programme Budget data. There are a number of possible reasons for this, including a lack of incentive on the part of providers to devote necessary effort and time in producing high quality data, and technical problems with accurately allocating cancer-related resource use to the cancer programme (for example, GP prescribing). (p.7)*

There is no reason to suspect that these unavoidable weaknesses and gaps in the data bias the CHERP81 threshold estimate in a particular direction, but they underline the importance of further research in this area and suggest a degree of uncertainty in the CHERP81's central estimate. If apparent differences between levels of funding for PBCs across PCTs arise from different classification of expenditure then any statistical effect attributed to that difference will be inaccurate.

## ESTIMATING OPPORTUNITY COST BASED ON DIFFERENCES BETWEEN PCTS

### Summary

CHERP81's threshold estimate is derived from predicting how the expenditure of individual PCTs in particular programme categories changes as their budgets change through time and of the impact on outcomes they achieve with it. However, CHERP81 is not able to make use of time-series data to directly estimate how changes to individual PCT budgets caused by NICE decisions will affect mortality. Instead, it has to use snapshots of the differences in spending and mortality between different PCTs at a single point in time,<sup>3</sup> and assume that less well funded PCTs would achieve the same outcomes as their better funded neighbours if given the same level of funding.

If two PCTs have the same priorities and capabilities, apart from their level of funding and a few random differences, then CHERP81 will be able to predict future changes to individual PCTs' health outcomes from future changes in funding by estimating a relationship between current average funding and current average outcomes across PCTs. We need, however, to be aware of possible structural differences between PCTs reflecting in part differences in preferences about health care provision – for example, some PCTs might be attached to their current priorities, so their behaviour (in terms of the health care they buy) does not change to match their neighbours just because funding does. This means that the threshold estimated by CHERP81 does not predict how PCT outcomes will change as a result of the impact of NICE decisions on PCT budgets.

### Analysis

The key units of statistical analysis in CHERP81 are cross-sectional differences in PCT outcomes linked to cross-sectional differences in PCT expenditure, rather than time-series data showing how mortality in fact responds to changes in expenditure, or how programme spending by PCTs actually responds to changes in their budgets. Over time it will become possible to directly compare outcomes and expenditures for each PCT using time-series data, closing a significant gap in the CHERP81 analysis between what is being observed (how different PCTs with different budgets spend and produce health) and what is being predicted (how individual PCTs' spending and production change as their budget changes). Until such data are available, the CHERP81 estimate rests on an assumption that *inter*-PCT comparisons are not only suggestive but actually determinative of *intra*-PCT behaviour through time.

In order to provide a useable estimate of the consequences of removing, say, £100m of expenditure from the general NHS budget to implement new NICE decisions, these cross-sectional observations across PCTs need to serve, on average, as a reliable proxy for hypothetical time-series data showing how *individual* PCTs would respond to different potential changes to their budget. This is equivalent to assuming that different PCTs are

---

<sup>3</sup> Average mortality from the year in question and two subsequent years is used as a proxy for current mortality.

separate instantiations of a single underlying production relationship between budget and mortality, or that the various PCTs can be modelled as lying along a single, common, budget-mortality curve. This constant elasticity relationship between overall budget and programme expenditure, which is assumed to hold on average across all PCTs, is estimated as the parameter  $b$  in equation (6.1) and the constant cross-PCT expenditure-mortality relationship for a given PCT is estimated as the parameter  $f$  in equation (6.2), both on p29 of Appendix B to CHERP81.

Our concern is that, regardless of the robustness of the observed relationship between differences in spending and mortality *between* PCTs, it may not be appropriate to extrapolate this relationship to relationships between budget and mortality *within* individual PCTs through time. CHERP81 has no underlying testable model of how PCTs behave in allocating funds or in using them. It relies on an underlying structural assumption, underpinning all its parameter estimates, that wealthier PCTs, random differences aside, are exact models for how poorer PCTs would behave and perform if only they were better funded. Collection of time-series data and analysis of PCT behaviour over time time-series should enable this assumption to be tested.

The authors of CHERP81 argue, in their document accompanying the final report, “Why is the central estimate of the cost per QALY threshold lower?”,<sup>4</sup> that:

*[W]e have not constructed or estimated a predictive model ... We do not infer changes over time from the cross section but estimate the expected health effects if the NHS had less or more resources (based on variation in resource use and health effects). This provides an estimate of the threshold in that period. 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.*

We are unable to agree with this characterisation of the CHERP81 model. The “expected health effects if the NHS had less or more resources” is properly modelled as a comparison of counterfactual *changes* over time in PCT budgets and outcomes resulting from NICE decisions. The application of a cost-per-QALY threshold in a given period represents a *prediction* as to the change in QALYs associated with a *future* (not contemporaneous) potential change in PCT budgets arising from that decision. The data required necessarily possess time-series characteristics. We agree CHERP81 does not provide a predictive model for deriving the future effects of current NICE decisions.

Even if NICE decisions are viewed as instantaneously bringing about a situation in which “the NHS had less or more resources”, the cross sectional model used by CHERP81 is inappropriate for this purpose. Its core assumption – that the effects of a given PCT counterfactually “having less or more resources” can be modelled based on the

---

<sup>4</sup> Available at the time of writing from [http://www.york.ac.uk/media/che/documents/Why%20is%20the%20central%20estimate%20lower%20\(final\).pdf](http://www.york.ac.uk/media/che/documents/Why%20is%20the%20central%20estimate%20lower%20(final).pdf)

behaviour and outcomes estimated using static cross-sectional comparisons between PCTs – remains unproven and, indeed, unexamined.

## WHAT ARE PCTS AIMING TO ACHIEVE?

### Summary

CHERP81 uses a snapshot of different PCTs, spending different amounts, to derive a positive relationship between spending and reductions in mortality. In order to arrive at an estimate of the cost of a QALY, rather than the cost of a life year, it also needs to determine how much more quality of life improvement higher spending PCTs achieve along with these reductions in mortality.

The key difficulty in arriving at this estimate of quality of life improvement is that there are no data available to show how quality of life varies with PCT spending. CHERP81 needs to arrive at some relationship between what it can measure – differences in mortality – and what it wants to calculate – differences in quality of life.

In addition, CHERP81 needs to predict how changes in NHS spending through time will change a particular PCT's spending and outcomes (both quality and quantity of life). As we explain above, it only has access to a snapshot of data across all PCTs for a single point in time – so there are no time-series data to test CHERP81's assumption of how PCTs behave when their budgets change.

CHERP81 deals with the absence of quality of life data by assuming that PCTs are as good at improving quality of life, which we cannot observe, as they are at reducing mortality, which we can. But how much we might expect PCTs to improve patient quality of life when their funding is increased depends on how we predict PCTs behave. It seems from the CHERP81 analysis that PCTs do not simply try to maximise QALYs. Do they maximise some wider definition of health, which includes QALYs, or do they not systematically try to "maximise" anything at all?

The first possibility: if PCTs maximise some broad definition of "health gain", which includes both QALYs and other valuable outcomes (like increased patient dignity and convenience, for example, as required by NHS England's "Mandate" (DH, 2013)) then, when CHERP81 assumes the existence of an equally sized bundle of non-mortality gains to go along with life extension, it is not estimating the cost of a QALY, it is actually estimating the cost of a unit of "health gain", which includes all these extra factors. So PCTs spend their money efficiently, but they try to buy more with it than just QALYs, which means we need to compare the opportunity cost of funding new technologies not to the amount of QALYs they generate but to the broader definition of benefit being used by PCTs. If this is right then CHERP81, in estimating the threshold cost of a QALY based on a fixed relationship between life years and quality of life, is only picking up part of what PCTs think of as health gain.

The second possibility: if PCTs do not behave like efficient maximisers of anything, even on average, then we have a different problem in calculating quality of life improvements without any data to measure them. CHERP81 assumes that PCTs are proportionately as good at removing the quality of life effects of illness as they are at reducing their effects on mortality. But this assumes that there is a strong link between how PCTs value preventing mortality and how they value improving quality of life – otherwise there is no reason to expect these equiproportionate reductions to be the same.

If PCTs do not all have a single, consistent objective that they all try to meet – and there is no evidence that they do – then one cannot predict one effect of increased PCT spending (e.g. reduced morbidity) by assuming it has a consistent relationship with another predicted effect (reduced mortality). We would hope that increased spending will tend to both decrease mortality and increase quality of life, but we cannot necessarily rely on changes in one to estimate unseen changes in the other. We would also hope, for example, that increased spending would decrease outpatient waiting times, but we would not expect the proportional reduction in mortality to be the same as the proportional reduction in average waiting times, even though both are positively related to expenditure. The same is true of the relationship between mortality and quality of life.

So CHERP81's assumption – that when we see changes in mortality we can assume that PCTs are doing a similarly good job in relation to quality of life – may not be correct, and so may not provide a good foundation for an estimate of the cost of achieving a QALY gain. Indeed the assumption, that PCTs are as good at reducing burden on quality of life as they are at reducing the mortality burden of disease, may be too optimistic for two reasons.

First, we believe that PCTs will be better at targeting mortality simply because data on mortality are collected while data on quality of life are not. What gets measured is likely to determine at least partially what PCTs care most about. It may also be that mortality, and not quality of life, is measured partly *because* PCTs inherently care more about averting death than they do about improving quality of life. So we expect PCTs to be better at the observable, and actually observed, parts of their job than the ones we cannot readily measure, rather than being equally good at both, as CHERP81 assumes.

The second reason is that CHERP81 is intended to produce an estimate of a threshold which can be used to directly compare the health gain from new technologies with the health gain foregone when we in effect reduce PCT budgets to fund them. So, in order for these measures to be comparable, health gain from PCT expenditure should be measured in the same way as health gain from proposed new treatments. The threshold applied by NICE is arguably a cost per *proven* QALY, and provable QALY increases are, by definition, less than or equal to actual gains. The opportunity cost calculated by CHERP81 is a cost per assumed QALY gain, which may lie above or below actual QALY gains.

The CHERP81 method does need some basis for making an assumption about the relationship between the achievement of quality of life improvement and mortality reduction, while recognising that there may not be any systematic relationship at all. While it seems reasonable to assume that additional funding leading to measured mortality reduction also leads to some improved quality of life, and that improving quality of life is also very important to PCTs, we think assuming that these are delivered with equal efficiency is too optimistic and that as a result CHERP81 has underestimated the value of the threshold which should be applied in practice.

## Analysis

It is important to consider exactly what motivations and behaviour CHERP81 is implicitly capturing when it models differences between PCTs' spending and mortality.

In specifying the equations to be estimated on p29 of Appendix B, CHERP81 explains that the dependent variable in the health production function,  $h_i$ , is "the health gain in PCT  $i$  in the selected programme" and suggests that it will be positively correlated with need. On p13 paragraph 7 of CHERP81 the  $h$  term is described as follows:

*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.*

And, also on p13, paragraph 6, that:

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

$$W = W(h_1, h_2, \dots, h_j) \quad \text{Equation 3.1}$$

which is a function only of the health outcomes,  $h_i$ , in each programme of care.

From these descriptions of the model to be estimated, we conclude that CHERP81 treats individual PCTs as engaging in optimising behaviour, but recognises that both the function being optimised and its arguments are undefined and unobservable.

There is the suggestion, above, that the "most obvious" interpretation of the health benefits to be maximised is life years, or perhaps QALYs. If this were the case, PCTs would act as life year or QALY maximisers; unless the PCT social welfare function placed different weights on equal life year or QALY gains depending only on PBC,<sup>5</sup> we would expect to see equal marginal cost per life year or QALY across all PBCs. It is also possible that PBCs do attempt to maximise QALYs, but that they are constrained by "lumpy" (non-linear) expenditure opportunities and indivisibilities, preventing them from achieving (or even approaching) equality of benefit at the margin. Alternatively, PCTs may attempt to achieve QALY maximisation, but fail to do so due to principal-agent or other constraints.

Assuming these constraints apply systematically, both these models are observationally equivalent to PCTs lacking an objective function, the consequences of which we discuss below.

Based on Table 2 below it is clear that we do not observe anything approaching equal marginal productivity of expenditure between PBCs when measured in QALY terms. We conclude that PCTs were doing something other than trying to maximise life years or QALYs.

---

<sup>5</sup> Since the social welfare function depends only on the  $h_i$  values, if those values represent QALYs then social welfare is maximised by maximising QALYs, potentially weighted, *but only by the PBC they happen to fall in*. A similar result holds in the case where the  $h_i$  represent life years. So this model is not compatible with individual or group based equity weights, for instance, or any other system of weighting which depends on characteristics other than patient-PBC.

**Table 2. Estimated PBC marginal cost per QALY in 2008/9, from Table 5.2 of CHERP81 (page 74)**

<b>Programme Budget Category</b>	<b>Cost per QALY</b>
Cancer	£16,997
Circulatory	£7,038
Respiratory	£1,998
Gastro-intestinal	£7,293
Infectious Diseases	£20,829
Endocrine	£3,124
Neurological	£5,480
Genito-urinary	£43,813
Trauma & Injuries	NA
Maternity & Neonates	£2,969,208
Disorders of Blood	£28,305
Mental Health	£49,835
Learning Disability	£78,854
Problems of Vision	£76,850
Problems of Hearing	£19,070
Dental Problems	£55,916
Skin	£174,775
Musculoskeletal	£20,254
Poisoning and AE	£163,766
Healthy Individuals	£1,483,012
Social Care Needs	NA
Other	NA

So, PCTs attempt to maximise an unknown function of “health outcomes”,  $h$ , but we do not know what these  $h$ -values mean, only that they reflect something other than straightforward QALY maximisation or mortality minimisation.

There is evidence to suggest that health spending decisions may reflect a variety of non-QALY sources of value (Hansen, 2012; Shah et al, 2012). The Department of Health’s stated intention is that “value based pricing” of medicines will do so in future (DH, 2011), and its Mandate to NHS England includes much besides QALYs (DH, 2013). So it may be reasonable to assume that the unobservable health outcome values reflect the overall value placed by PCTs on *all* relevant features of a health outcome brought about

by their spending decisions. Among other things, this represents a key reason why marginal productivity *particularly when measured in QALYs* which are a subset of the PCT's objective function (if it exists) cannot simply be extrapolated from PBCs for which good data exist to those for which no data are available: the mix of mortality, quality of life and non-QALY outputs may well differ systematically by ailment and by patient population.

The assumption that PCTs maximise an unobservable variable is, by definition, unfalsifiable. The more parsimonious view, given the only evidence we have suggests that PCTs are not maximising anything we can observe, is that PCTs do not systematically attempt to maximise any consistent measure of health output. On this view, CHERP81 is in practice modelling PCTs as health producing technologies rather than as agents.

If there is a clear statistical relationship between expenditure and mortality it does not matter *why* PCTs arrive at their observed mortality outcomes: reducing mortality might be a by-product of PCTs trying to optimise a "health gain" function which mortality is part of, or it might be the result of increasing inputs to production technology which systematically transforms those inputs into reductions in mortality. We do not need to know or make assumptions about what is in the PCT "black box." However, the final estimate of the threshold preferred by CHERP81 relies not only on observable mortality reductions, but on *assuming* that the observable proportional reduction in the mortality portion of disease burden is a reliable proxy for the unobservable proportional reduction in the quality of life portion of disease burden. NHS expenditure is likely to influence quality of life as well as mortality, but, if PCTs are maximising something other than a QALY-based objective function, or are not optimising at all, then *there may not be any consistent relationship between the degree of proportional mortality reduction and the proportional reduction in quality of life burden*. Even if PCTs were maximising QALYs, we would not necessarily expect it to be the case that they were seeking to reduce the quality of life burden in proportion with the mortality burden. This is even less likely to be the case if PCTs are trying to do a variety of things, or cannot properly be modelled as "trying" to do anything at all.

### **If PCTs do not maximise QALYs, what does the estimated threshold represent?**

To the extent that PCTs consider non-QALY health outcomes to be socially valuable, they may be directing their spending to non-mortality-reducing PBCs in order to reap these non-QALY benefits: societal concern for the non-health wellbeing of mothers and infants for instance might drive expenditure on maternity and neonates. On this interpretation, the total social value (as interpreted by PCTs) of expenditure on each PBC might be equal at the margin, but the opportunity cost of displacing expenditure allocated on this basis would not be denominated in cost per QALY terms, and would have no obvious interpretation as a cost per QALY threshold in other budget decisions.

If PCTs do optimise this broader, non-QALY measure of health gain, then CHERP81's threshold, estimated on the basis of equiproportional reductions in mortality and quality of life burdens, is implicitly capturing a portion of the non-QALY value associated with

the expenditure, particularly in non-mortality-reducing PBCs. So the threshold estimate arrived at by CHERP81 would not be a cost per QALY figure, but rather a “cost per socially valuable health gain”. Under this model, it is the full social value of changes to health outcomes brought about by expenditure on new technologies which should be compared to the CHERP81 threshold estimate, not QALYs alone.

We are not in a position to extrapolate from the directly calculated QALY gains associated with new technologies considered by NICE to the full social value of those gains calculated according to the methodology implicit in CHERP81. At a minimum, the additional social value embedded in an observed, rather than assumed, QALY must be sufficient to explain the non-random part of observed differences in marginal cost per QALY across different PBCs at the margin, assuming PCTs do attempt to maximise some measure of health gain as CHERP81 suggests.

The alternative to assuming that PCTs maximise some non-QALY health measure is that there is no objective function which PCTs are consistently able to maximise.<sup>6</sup> If this is accurate, the assumption of equiproportional reductions in mortality and quality of life burden adopted by CHERP81 is likely to be “optimistic”<sup>7</sup>. It requires that PCTs, which are able to observe mortality and which seek to maximise, if anything, some non-QALY measure of health, nonetheless end up reducing a quality of life burden they neither observe nor exclusively focus on, in the same proportion as observable mortality. There is no reason a priori to expect such a relationship between the different arguments of the social welfare function. There is, of course, the possibility that PCTs attach *more* importance to securing QALYs from quality of life gains than from mortality reduction, but we neglect this possibility below, as we see no rationale or evidence to support it.

How, then, should we model unobservable reduction in quality of life burden? While such reduction is presumably non-zero, a conservative approach would be to only include quality of life improvements from general NHS expenditure for which there was supporting data, which would be consistent with NICE’s approach to calculating QALY gains from new technologies.

An intermediate approach would suggest that, if there *is* a systematic relationship between the different outputs of PCT expenditure, PCTs may well be better at targeting more readily observable mortality measures than hard to measure quality of life gains, implying that the proportional reduction in quality of life burden lies below the proportional reduction in mortality with an upper limit of equiproportionality and that, by only considering the case of equiproportional reductions, CHERP81 is likely to understate the optimal value of the threshold. By assuming that PCTs are as good at proportionally reducing QALY burden as they are at reducing mortality, CHERP81 predicts more marginal quality of life improvement than may well occur in practice and overstates the marginal health gains from general NHS spending.

---

<sup>6</sup> As noted, we consider the equivalent possibilities that PCTs “attempt” to maximise QALYs but are unable to do so, or that indivisibilities or non-linear marginal costs prevent optimisation at the margin.

<sup>7</sup> Throughout this paper we follow CHERP81’s convention of referring to assumptions which are likely to be generous as to the productivity of NHS expenditure and therefore to underestimate the threshold as “optimistic” and those which are likely to overestimate the threshold as “conservative”.

## CONSERVATIVE AND OPTIMISTIC ASSUMPTIONS IN ESTIMATING THE VALUE OF THE THRESHOLD

### Summary

CHERP81 has to make a number of assumptions in arriving at its estimate of the threshold. That inevitably renders the estimate uncertain. Some of these assumptions are “optimistic” (by which the CHERP81 authors mean tending to understate the threshold) and others “conservative” (tending to overstate the threshold). Our view is that taking all of the assumptions together, on balance, CHERP81 may be “optimistic” in its assumptions and understate the threshold as a result.

We have already argued that CHERP81 is likely to overestimate the quality of life improvements in a PBC because of the equiproportional link to the reduction in mortality, and in doing so would be understating the threshold. Other than this, the most important assumptions CHERP81 makes are in relation to: (i) the long term effects of increased spending, and (ii) the future health of patients whose deaths are prevented by increased spending. CHERP81 notes that it makes “optimistic” assumptions on (ii), in relation to the future mortality of patients whose deaths are prevented, but argues that it more than offsets this “optimistic” assumption in the way that (i) is dealt with, by only counting the immediate benefits of spending more money today, and neglecting the longer term benefits of this spending. CHERP81 also makes assumptions about (iii) discounting, (iv) the future quality of life of patients whose deaths are prevented, and (v) how unallocated funding is spread across the PBCs.

### *The future effects of current spending*

Starting with (i), the future effects of current spending: we agree with CHERP81’s argument that if a PCT spends more today, then not only will mortality today decrease, but there will also be some future reduction in mortality as a result of today’s spending, even if we stop spending the additional money in the future. But, if we look at the differences in mortality between two PCTs today, and measure the differences in their spending today, we will be ignoring the fact that some of the differences in mortality we observe today are caused by differences in how much each of them spent *in the past*. If PCTs that spend relatively more on, say, respiratory care now also tended to spend more on respiratory care in the past – and CHERP81 presents strong evidence that this is the case – then some of the reduced mortality from respiratory causes we measure today is actually the delayed benefit of past spending differences.

We expect that there will be quite a close relationship between high spending by a PCT in a given PBC today and high spending in the past. CHERP81 implies just such a relationship, because the variables used to predict that spending, like number of lone pensioner households, stay fairly constant through time, and also because PCTs may have a tendency to stick to past spending allocation between PBCs. Past studies have assumed that there is a perfect relationship between past and current spending differences, and CHERP81 seems to adopt a similar assumption when looking at future mortality data, and shows that current spending differences are nearly as good at estimating reduced mortality in the past as in the future, which can only be the case if

spending differences persist through time. If differences in spending between PCTs are relatively constant through time, then attributing average mortality effects over a three year period to spending differences in one year, as CHERP81 does, will ignore gains further in the future, but also give credit to today's spending for gains properly attributable to past differences in spending.

So, whilst CHERP81 is right to say that its estimate under-counts some of the future benefits of spending today, it also over-counts the benefits of today's spending, as some of those benefits will be a result of past spending. How these two effects balance out depends on how consistent PCTs are in their need-based spending from year to year. If they are quite consistent - and CHERP81's data and other assumptions show that they believe them to be almost constant through time - then the overall effect of ignoring how current spending influences future health will be relatively small and so the "conservative" (overestimate) effect will be small.

If so, it would not be reasonable for CHERP81 to use this assumption to offset its second, key "optimistic" assumption to which we now turn.

### ***The future mortality of patients whose deaths are prevented***

CHERP81 estimates the number of life years saved from each instance of averted mortality by assuming that patients whose deaths are prevented will go on to live for a number of years equal to an average, healthy, member of the population of the patient's age and gender. CHERP81 accepts that this is an "optimistic" assumption (tending to understate the threshold), since patients who would have died but for a PCT's additional spending are likely to be less healthy, and therefore experience shorter lifespans than population average. CHERP81 argues that this "optimistic" assumption is offset by its "conservative" assumptions about the future effects of current spending but, as we argue above, we do not agree that this offsetting assumption is particularly "conservative". So we believe that any threshold calculated on the basis that patients whose lives are saved will return to normal mortality will tend to be too low.

### ***Combining these two key assumptions in three scenarios***

CHERP81 presents three scenarios for its key assumptions: One where it is "optimistic" about both its assumption (i) in relation to future health benefits and its assumption (ii) in relation to the future mortality of patients whose deaths are averted (leading to a low threshold); one where it is "conservative" about both factors (leading to a high threshold) and one where it is "optimistic" about (ii) patient future mortality but "conservative" about (i) the future benefits of current spending (which it uses as its central or best estimate). CHERP81 does not report the fourth scenario, which is "optimistic" about (i) future health gains but "conservative" about (ii) patient mortality following averted deaths. We can tell from looking at the high and low threshold estimates that this combination of assumptions would yield a higher threshold than the "best" CHERP81 estimate of £12,936. Similarly, if we assumed that the true value of the threshold lay in the middle of the most "optimistic" estimate and the most "conservative" estimate, we would arrive at a significantly higher estimate of the threshold. Even if CHERP81 is exactly right about the degree of "optimism" and "conservatism" in its assumptions, we think that it has still been "optimistic" overall in its choice of which ones

to use for its “central” estimate: it uses an “optimistic” assumption which does more to lower the threshold than its “conservative” assumption does to raise it.

Additionally, we argue above the overall effect of assumption (i), ignoring future effects and so not lowering the threshold, is probably smaller than CHERP81 implies. This means that the true central estimate of the threshold could be somewhere between the most “conservative” value (assuming no net future effects and poor mortality rates for patients who avoid death) and the CHERP81 “central” estimate (assuming no net future effects and “optimistic” future mortality rates for patients who avoid death). This would imply that the threshold lies somewhere between £12,936 and £29,314.

### ***Assumptions about discounting***

The next key assumptions are either: (a) that the future health gains should not be discounted, which is how what is described in CHERP81 as the “best” estimate £12,936 figure is calculated, or (b) that discounting only needs to be applied to mortality effects, because all quality of life gains are enjoyed immediately, which is the source of the £13,141 estimate listed as a discounted estimate of the threshold.

The opportunity cost of health displaced through a change in general NHS expenditure needs to be calculated in a consistent fashion with the cost per QALY of proposed technology investments, which are required to discount future QALY gains. So it does not seem appropriate to report an undiscounted £12,936 figure. In (b), CHERP81 does assume that delayed mortality brings benefits over the average four and a half extra years patients live when their deaths are prevented by extra spending and that these need discounting. However, (b) assumes that all the quality of life gains from extra spending today are enjoyed immediately, rather than being spread across a number of years, even if all health gain occurs today. There will be a mix. Hip and other joint replacements have long term quality of life benefits, for example, even if their effects on a patients’ health are immediate, whilst many therapies for chronic disease may have no continuing benefit should the patient stop treatment.

We think it is more plausible to assume that quality of life gains occur over the same period as mortality gains and preferable that both are discounted.

### ***The future quality of life of patients whose deaths are prevented***

CHERP81 assumes patients who would have died, but for increased spending, live their additional life years at a quality of life above that of typical sufferers of their disease (proportional to their reduction in YLL). While we understand that in certain circumstances the cause of the prevented death will be acute, so patients may return to normal health after treatment, in general patients who come very close to dying will experience worse, rather than better, health than average sufferers of their disease over the years for which their deaths are delayed. A reasonable compromise would be to assume that patients’ life is typically extended at average quality of life associated with their disease.

### ***Dividing unallocated funding across PBCs***

CHERP81’s econometric modelling only allocates 72% of any change in spending across the 23 PBCs. The allocation of the remainder needs to be assumed, and CHERP81’s

estimate of the threshold is extremely sensitive to the assumption that is made: CHERP81's earlier assumption (in the June 2013 version of the paper) that these funds affected only PBCs which do not prevent mortality resulted in a threshold estimate of £18,317 which is 42% higher than its current figure. We are concerned about how important this assumption is to CHERP81's conclusions and believe that this new assumption may overstate predicted increases in funding for at least one, highly productive area of spending, and so understate the threshold.

### **A revised range for the threshold**

Using revised assumptions on points (i) to (v) above, the estimate of the threshold based on CHERP81's data is above £32,114 (assuming return to normal population mortality for averted deaths) but significantly lower than £67,664 (assuming two year lifespans for averted deaths), with the discounted value of £13,724 (the equivalent of CHERP81's undiscounted central estimate of £12,936) representing a lower bound<sup>8</sup>.

All these estimates continue to accept CHERP81's assumption that PCTs are as good at reducing quality of life burden as they are at reducing mortality burden. Different assumptions about this relationship would produce higher threshold estimates.

## **Analysis**

### **CHERP81's summary of assumptions**

CHERP81 provides, in Box 5.1 on p84, a summary of the assumptions it identifies as underlying its estimate of health effects and characterises them as either tending to overestimate or underestimate health effects or the threshold itself relative to the underlying "correct" values. We provide a brief survey of these summary conclusions and indicate whether we agree or disagree with CHERP81 in each case.

*Assumption 1: "Deaths averted by a change in expenditure returns an individual to the mortality risk of the general population (matched for age and gender)."*

CHERP81 notes that this assumption leads it to overestimate health effects. We agree, but argue below that CHERP81 understates the extent of the overestimate.

*Assumption 2: "Expenditure and outcome elasticities are uncorrelated."*

CHERP81 notes that this assumption ignores some evidence of correlation and so is likely to lead to a slight underestimation of the threshold. We agree, but argue below that evidence of systematic heterogeneity in PCT productivity needs to be formally accounted for in CHERP81's model of PCT behaviour. CHERP81 argues on p90 of Appendix B that:

*...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" (emphasis added).*

---

<sup>8</sup> The values £32,114 and £67,664 (£30,270 and £63,844 undiscounted) have not been revised from the June 2013 version of the report and may therefore be an overestimate. However, for consistency, we use the figures in the November 2013 version of CHERP81...

If productivity causes high spending rather than the reverse it will be difficult to disentangle causal effects of spending across a pooled sample of PCTs with levels of productivity which vary systematically with spend.

Assumption 3: *"Mortality effects of changes in expenditure (reported at PCT level) can be applied to all mortality recorded in a PBC."*

CHERP81 suggests that this has the potential to overstate health effects relative to other, in CHERP81's view less-plausible, assumptions. We agree.

Assumption 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."*

CHERP81 suggests that the result of this assumption is either neutral or tends to overstate health effects, depending on the preferred alternative model for changes in investment. We agree.

Assumption 5: *"Health effects of changes in expenditure are restricted to the population at risk during one year."*

CHERP81 argues that this assumption tends to understate health effects. We agree, but argue below that CHERP81 greatly overstates the degree to which health effects will be underestimated. This is because differences in expenditure will tend to be strongly correlated through time.

Assumption 6: *"Health effects restricted to the PBC in which expenditure changes. No health effects associated with changes in GMS expenditure (or PBC22, Social Care)."*

CHERP81 argues that this assumption tends to understate health effects as it will fail to detect spillover benefits of spending in a given PBC to patients falling under other PBCs "unless they happen to be correlated with changes in expenditure in these PBCs".<sup>9</sup>

In fact, positive spillover effects will be included in CHERP81's estimates where differences in expenditure between PBCs are positively correlated across PCTs and not be estimated where differences are negatively correlated. Our preliminary understanding is that relatively similar overall PCT budgets imply some level of negative correlation, so we accept that some hypothetical positive spillovers – where spending accounted for under respiratory, say, leads to early diagnosis and therefore delayed mortality for a lung cancer patient, for instance – will be unaccounted for.

There may, however, be possible negative spillovers, whereby a PCT's greater allocation of costed resources to one PBC systematically predicts a greater allocation of uncoded resources – if, for example, a PCT with high spending on oncology relative to respiratory illness also tends to divert more talented staff, or expend more unmonitorable effort on cancer patients than those suffering from respiratory illness. We accept that any negative spillovers are likely to be smaller than positive spillover effects, and that the net effect of the assumption is likely to be an understatement of health effects.

---

<sup>9</sup> This description is not strictly correct, since CHERP81 measures *differences* between PCTs' expenditures not *changes* in PCT expenditures over time.

Assumption 7: Has been removed in this version of CHERP81, but previously read (in the June 2013 version of the paper): *"Remaining change in total spend is assigned to the group PBCs where mortality effects could not be estimated."*

The assumption was assessed by CHERP81 as being likely to understate health effects relative to the alternative assumption of assigning the shortfall in expenditure allocated based on the estimated expenditure elasticities across all PBCs.

This assumption has been altered in the final version of CHERP81, leading to a significant reduction in the threshold. The updated assumption – that unestimated total spend is allocated across all PBCs proportional to their estimated expenditure elasticities, is not captured in Box 5.1.

While it is instructive to compare the approach to the shortfall in calculated elasticities with that adopted to non-PCT budgets in "additional assumption 1", below, we generally accept the reasons given by the authors for relaxing this assumption and are concerned mostly with the high degree of sensitivity displayed by the threshold estimate to plausible alternative assumptions as to the allocation of unestimated expenditure. We also argue, below, that the same face validity tests which caused the authors to abandon their original assumption suggest, when applied more generally, that the new assumption is likely to understate the threshold.

Assumption 8: *"Same proportional effect on QALY burden of disease as the estimated proportional effect on the life year burden of disease."*

CHERP81 argues, and we agree, that this assumption may either over- or understate health effects depending on the true underlying proportional relationship. But, as we have already argued, the true relationship is likely to be less than proportional, meaning that health effects have been overestimated.

Assumption 9: *"Life year effects are lived at a quality of life that reflects a proportionate improvement to the quality of life with disease."*

CHERP81 concludes that this assumption is neither obviously conservative nor obviously optimistic, since it lies between the conservative assumption that life years are lived at a quality of life consistent with the patient's disease and the optimistic assumption that they are lived at normal population health. We believe that patients who would have died that year but for the medical interventions associated with marginal additional spending in that year are likely to experience poor health over the period closely following the life-saving intervention relative the average sufferer of the disease or condition in question. We believe this assumption, on balance, overstates health gains.

Assumption 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."*

As with assumption 8, CHERP81 argues that this assumption could either over- or underestimate health effects, depending on the true relationship between life extension and quality of life improvement. As with assumption 8, we agree but argue above that

the true relationship is likely to be less than equiproportional, and that this assumption therefore overestimates health effects.

In addition to the deleted assumption 7, we identify two relatively significant assumptions made by CHERP81 which are not listed in Box 5.1. These are:

Additional assumption 1: *"non-PCT budget is wholly unresponsive to exogenous shock"* (p86 of Appendix B)

This assumption is acknowledged by CHERP81 to lead to the cost of a life year being lower by 17.7% relative to the plausible alternative assumption that non-PCT budgets are equally responsive to NICE-induced budget changes as PCT budgets. We agree that this assumption may understate the cost of a life year.

Additional assumption 2: *"Only quality adjusted net YLL were discounted, and thus QALYs associated with gains in QoL during disease were not."* (see p65 of Appendix C).

CHERP81 argues that this assumption arises directly from the earlier assumption (in row 5 of Box 5.1) that health effects are restricted to one year. We argue below that this belief arises from a conflation of health effects (which are not the proper subject of discounting) with enjoyed quality of life effects (which are). As such we believe that this assumption is "optimistic" and tends to understate the threshold.

The key offsetting assumptions 1 and 5 are used by CHERP81 to sketch out the upper and lower bounds for its threshold estimate and to settle the central estimate. We argue below that assumption 1 is more "optimistic" and assumption 5 is less "conservative" than CHERP81 recognises. As a result we argue that the true value of the threshold is likely to sit closer to CHERP81's upper-bound than its central estimate, even setting aside the arguments we make in relation to the remaining assumptions.

### **Considering the key assumptions**

Given the methodological challenges associated with estimating the marginal cost per QALY across the NHS, CHERP81 made, of necessity, a number of simplifying assumptions in arriving at its central estimate. While each of these assumptions increases the potential for variance between the estimated value of the threshold and its "true", underlying value, they will not necessarily bias the central estimate unless they are systematically "conservative" (tending to overstate the value of the threshold) or "optimistic" (tending to understate the value of the threshold).

The assumptions made in CHERP81 are a reasonable attempt to deal with unavoidable uncertainty, given the limitations of the available data. However, we disagree with the statement made on CHERP81 p86 (and elsewhere subsequently in the text) that these assumptions are, in aggregate, "conservative" in nature. Contrary to the CHERP81 view of a bias towards "conservatism", we find an overall tendency towards adopting the more "optimistic" assumption from the range of available options, a tendency which we believe results in the best estimate of the threshold being understated.

The first two paragraphs of CHERP81 p86 outline two key assumptions, matching assumptions 5 and 1 in Box 5.1:

(i) *"The health effects of a change in expenditure are restricted to the population at risk during one year"*

which is identified as being "conservative" in three regards: (a) it neglects quality of life effects accruing after the first year; (b) it ignores potential future reductions in patient mortality; and (c) it fails to capture the benefits of future disease prevention,

and

(ii) *"...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)"*

which CHERP81 recognises to be "optimistic" in nature.

We consider these two assumptions in turn and then consider two other issues in the report concerning respectively:

- extrapolating from unobservable QALY gains in mental health
- discounting (which we have labelled "additional assumption 2").

Finally, we look at the implications of our discussion for the threshold estimates.

### **Lagged effects of expenditure on health**

In a discussion of possible future research CHERP81 highlights the desirability of conducting analyses with longer and more complex lag structures when more years of data are available. This is on the basis that NHS spending today may affect health gains further in the future.<sup>10</sup> Given the obvious potential for these kinds of long term effects, we agree that it would indeed be desirable to investigate these lagged effects when sufficient years of data become available.

CHERP81 states that: "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" (page 94, fourth paragraph) and also that "the health effects of a change in expenditure are restricted to the population at risk during one year" (page 86, second paragraph). However, many of the health effects used by CHERP81 in estimating the threshold are extrapolated, not observed, which means that the time at which they occur is unknown. They cannot be said to have been "restricted" to a particular period. More importantly, the potential for out-of-period effects of health spending means not only that spending in 2008 may have had (and still be having) future health impacts, currently unmeasured, but also that differences between PCTs in this year's health gains may in part be due to differences in spending that took place earlier than the single year spending differences used to calculate each of CHERP81's threshold estimates.

Prior differences in spending may be strongly correlated with current spending differences. Such a correlation between current and past differences in PCT expenditure might arise for example if high spending (relative to need and input costs) by a PCT on a particular PBC in a particular period is correlated with higher spending in other years, as

---

<sup>10</sup> As we explain below, the data for subsequent years available to CHERP81 does not permit it to distinguish future health effects from the consequences of persistent differences in spending.

a result of a degree of consistency in the distribution of lone pensioners and other expenditure-correlated variables, or simple institutional inertia, or unobservable idiosyncratic features of the PCT biasing it towards a particular programme.

Consider, by way of example, two PCTs, one of which has a higher proportion of lone pensioner households and therefore higher expenditure, for example, on respiratory health services. To the extent that the distribution of lone pensioner households is relatively stable through time, and holding all else constant, we would anticipate that the higher-spending PCT would have enjoyed higher spending on respiratory care not only in one year, but in previous years as well. If respiratory spending tends to generate future as well as current health gains, any observed difference in mortality this period would be partially attributable to past differences in spending. If the distribution of lone pensioner households and other statistical determinants of respiratory spending were constant, the expected future benefits of present-day spending would exactly cancel out the current benefits of past spending attributed to current spending differentials. In such a steady state, once future health gains are discounted at 3.5% consistent with NHS policy, current differences in health outcomes will overestimate the total value of health gains associated with current differences in expenditure.

CHERP81 explicitly recognises the likelihood that expenditure differences are not only correlated between period but actually remain largely constant through time:

*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)" (see p9 of Appendix B)*

Further, CHERP81 states (in footnote 33 on p44) that:

*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.*

Since no expenditure difference data for periods  $t+1$  and  $t+2$  are used, this is equivalent to assuming that the health gain in period  $t$  is equal in expectation to the average of the health gains observed in periods  $t$ ,  $t+1$  and  $t+2$ . This assumption will hold only if expenditure differences between PCTs are stable through time, consistent with the "equilibrium position" described above.

To put it another way: the only reason to smooth out random variation in mortality measures by using the unweighted average of observed mortality differences over the next three years as an estimate of the health gains from today's spending is if one believes that the difference in mortality next year will be the same as the difference in

mortality this year. This belief, in turn, can only be accurate if the system one is observing is at a steady state, so that observed differences in mortality in any one year fully reflect the lagged effects of all past and current differences in spending and that those differences are constant across the relevant period.

Given the evidence for, and assumption by CHERP81 of, near-perfect correlations between current and past spending differentials, it cannot be determined conclusively whether the additional health gain attributed to differences in current year spending but actually arising from differences in past spending is offset by the (discounted) future health gains of current year spending.

It is also not possible, given the available data, to reject another hypothesis. This is that PCTs which have a higher instrumental variable-predicted spend on a given PBC (for example as a result of containing more lone-pensioner households) engage in both higher spending *and* spending skewed towards present rather than future health benefits. That is: preventative medicine and treatments intended to delay mortality in high need populations are substitutes, rather than complements, meaning that a portion of the observed current difference in mortality between PCTs with high predicted spending and those with lower predicted spending may be explained by a systematic tendency for high need, high spending PCTs to respond to present need by reallocating expenditures away from preventative or other delayed health benefit services. If this hypothesis were accurate, not only would current health gains have the potential to be negatively correlated with future health benefits, but one period mortality estimates would tend to overstate the value of the additional spending undertaken and therefore understate the marginal cost per QALY – though this effect would be accounted for if outcome elasticities were estimated at a steady state, as we describe above.

Considering the complexities associated with estimating the aggregate out-of-period effect of health expenditure, which depend on the relative sizes of the correlations between expenditure differences through time and health gains with lagged expenditure, we believe that their expected size is very small and agree with the CHERP81 approach that the estimate of the threshold ought to assume no such effects until the proposed revised analysis using the model with longer lags has been estimated. If lagged effects are properly excluded from the current model, however, their exclusion can no longer be relied on as a “conservative” assumption intended to offset other, more “optimistic” assumptions, a result which we will consider in more detail below.

### **Assumptions as to future mortality and quality of life where mortality is averted**

We have outlined above the “optimistic” assumptions made by CHERP81 in rows 1 and 9 of Box 5.1 with respect to the future mortality risk and future quality of life of patients whose death is averted as a result of increased expenditure and the extent to which they are neutral or offset by “conservative” assumptions regarding duration of health effects.

For any age and gender defined population group which includes patients who would have died but for increased PBC expenditure, however it is defined or identified, we believe that those particular patients will tend to experience on average worse future

health outcomes than the average for the members of that group as a whole, notwithstanding that their counterfactual deaths have been averted by health care expenditure. That is: other than in acute cases, which are likely to be the exception when restricting our analysis to mortality averted by *marginal* differences in PCT expenditure,<sup>11</sup> being a person who would have died but for a marginal additional health care expenditure is likely to be a reliable predictor of increased future ill-health and future mortality relative to any similar group of patients.

As such, predicting that patients whose death has been averted both return to population mortality norms (assumption 1, above) *and* experience an improvement in the quality of life burden of their disease proportionate to the (overestimated by the first assumption) reduction in life year burden (assumption 9) is likely to be “optimistic”, i.e. it introduces a downward bias to the estimate of the threshold. As noted above, CHERP81 accepts that the assumption as to mortality is optimistic (see summary in row 1 of Box 5.1 on p84), but does not recognise (in row 9 of Box 5.1) that estimating quality of life for counterfactual victims at above the level typical for comparable patients is also optimistic. We would argue that quality of life following a narrowly averted mortality event may well lie below, rather than above, levels experienced by typical patients and that, in the absence of further data to resolve this question, assuming quality of life following averted mortality is at disease sufferer norms is a more reasonable basis on which to arrive at a central estimate of the threshold.

### **Offsetting assumptions on averted mortality and health effects**

Table 5.1 on p74 demonstrates the relative effects of these two countervailing assumptions (i) and (ii) (rows 5 and 1 in Box 5.1): if we adopt the preferred CHERP81 approach and are “optimistic” with regard to future mortality risk for averted deaths – assumption (ii) – and “conservative” with respect to future health effects – assumption (i) – we arrive at the “best estimate” of £12,936. Changing the “conservative” assumption (i) on health effect duration to an “optimistic” one gives the “lower bound” estimate of £2,018, while changing the “optimistic” assumption (ii) on future mortality to a “conservative” one provides the “upper bound” estimate of £29,314.

If the hypothesis we put forward above – that the expected values for the instrumental variables used to model expenditure changes tend to stay constant through time implying a steady state whereby past additional spending generates as much extra health today as present additional spending does tomorrow – is correct, then the true value of the threshold must lie above £12,936.<sup>12</sup>

---

<sup>11</sup> CHERP81 assumes that PCTs prioritise socially valuable health gain, meaning that socially valuable “rule of rescue” situations where immediate or imminent death has the potential to be averted are unlikely to be at the margin and therefore unlikely to differ in their delivery between PCTs. We believe these types of cases are the most likely to be acute.

<sup>12</sup> The upper bound estimate of £63,844 is based on the assumption that life years gained for each prevented death cannot be less than 2, since mortality figures are averaged over three years (see Appendix C p 24). Since CHERP81 is calculating the reduction in mortality *caused* by additional spending today, it is not sufficient to note that this reduction persists into future periods, since this persistence could be explained by future differences in expenditure, correlated with those in the present. Since CHERP81 does not examine future

The caveat to this conclusion is that none of the scenarios considered relaxes the “conservative” assumption that no health benefits will accrue to future sufferers of a disease as a result of current expenditure. We have argued above that future preventative benefits must be offset against current preventative benefits enjoyed as a result of past spending, which will be negative overall when properly discounted in the kind of steady state equilibrium modelled by CHERP81 or, as an alternative, because spending in response to need may prevent rather than promote preventative expenditure. It is likely that relatively constant values of instrumental variables through time imply that most preventative benefits are already captured in the estimates presented and, in any case, believe that the marginal changes in prevention arising from *marginal* expenditure changes are likely to be small relative to direct effects on mortality and quality of life, even if the *average* productivity of preventative spending is high. We therefore conclude that the net effect of these assumptions is “optimistic”, whether or not out-of-period spending effects are correctly treated as being strictly positive.

If the future effects of health care are treated as being offset by the present effects of past spending, then neglecting out-of-period effects will be a neutral rather than “conservative” assumption. The optimal value of the threshold will lie between the (undiscounted) best estimate of £12,936 and the disease-average-quality-of-life-“upper-bound” estimate of £63,844.<sup>13</sup>

### Assumptions in extrapolating unobservable QALY gains

We now consider the assumptions in rows , and 10 of Box 5.1, dealing with the extrapolation from observed mortality to unobserved quality of life effects, based on equiproportional reduction in respective portions of the burden of illness. The arguments made here also apply to assumption 9, concerning quality of life improvement for patients whose deaths are averted. CHERP81 states, on p86, that:

*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.*

While we agree that assuming that unobservable PBCs deliver QALYs at the same rate as observable PBCs would imply a lower threshold estimate, such an assumption would be

---

mortality as a function of future differences in expenditure it is not reasonable to exclude the possibility that the extension in life *caused* by today’s spending may be less than 2 years per averted death meaning that the correct upper bound estimate should lie above £63,844.

<sup>13</sup> This figure is taken from column 2 of Table 5.1 on p74 of CHERP81. We note that, unlike the other values in Table 5.1, the values in column 2 are unchanged from those in the June 2013 version of the report. Since we would expect these estimates to have changed in response to the altered treatment of unallocated funding, which reduced CHERP81’s central estimate from £18,316 to £12,936, we believe this figure may need to be revised downwards to make it consistent with the basis on which other figures have been calculated. If this is the case, then the revised figure, not its current value, should be substituted for our upper bound estimate, above. This also applies to values calculated based on column 2 of Table 5.1 elsewhere in this paper.

inconsistent with the fact that observed PBCs do not deliver life years (their observable output) at anywhere near the same marginal cost as each other and that the evidence of CHERP81 suggests, correspondingly, that PCTs do not appear to have optimised expenditure across PBCs based on QALYs. Contrary to the suggestion that assuming equiproportional burden reductions across all PBCs is a “conservative” approach, we argue above that there is no reason to expect there to be a consistent relationship between mortality and quality of life burden reductions and that PCTs are likely to be more effective at targeting observable mortality than unobservable quality of life. Given this, we conclude that the net effect of this assumption is “optimistic”.

### **Estimating unobservable QALY gains from non-marginal decisions in mental health**

On p86, CHERP81 states:

*...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 is likely to have been underestimated and the central estimate of the threshold overestimated.*

This analysis rests on a consideration of the calculated cost per QALY of a range of non-marginal accepted and rejected expenditure programmes in mental health. The key observation is that at least one rejected technology displays cost per QALY below the estimated marginal cost per QALY in PBC 7, which, it is suggested, would imply a lower marginal cost per QALY, at or below the per-QALY cost of the rejected technology. Recall, however, that CHERP81 does not consistently suggest in its underlying structural model that PCTs seek to maximise QALYs, and its results imply behaviour at the margin that is not QALY maximising. Hence the observation of a rejected technology with a cost per QALY lower than marginal cost per QALY estimated for the Mental Health PBC, or the costs-per-QALY of accepted but non-marginal technologies tells us nothing about the marginal cost per QALY.

The CHERP81 estimate of the marginal cost per QALY in mental health lies above the average estimated value for the NHS as a whole. This suggests that the decisions which have determined mental health spending are either not the product of a health-maximising decision maker, in which case previous decisions cannot tell us anything useful about expected health gains from additional future mental health spending, or that they are based on purchasing something more than a QALY at the margin. Since we do not have any way of measuring the mix between QALY and non-QALY benefits for past spending decisions, we cannot draw conclusions about the threshold from what was approved or rejected in mental health spending on the basis of costs per QALY alone.

The examination of mental health in CHERP81 does not tell us about value at the margin. It is not evidence that the central threshold estimate may be “conservative”.

## Assumptions as to discounting

The central threshold estimate of £12,936 reflects undiscounted QALYs, preventing accurate policy comparisons between it and ICERs calculated on the basis of QALYs discounted at 3.5%. It is clearly preferable to estimate a threshold based on marginal cost per appropriately discounted QALY.

CHERP81 does report a discounted estimate for the threshold of £13,141, and this figure should be preferred for policy purposes to the undiscounted figure of £12,936. However, it is not clear that this discounted figure has been calculated in such a way that it can be compared to ICERs calculated on the basis of QALYs discounted at 3.5%.

The relevant passages, on p64 of Appendix C to CHERP81 state:

*Although this 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 (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.*

And on p65 of Appendix C:

*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 discounting model applied here results in an overall average discount rate of 1.6% across all QALYs, which would imply (given the 3.5% annual discount rate) that the mean QALY (whether as a result of reduced mortality or morbidity) delivered by the mean health intervention is enjoyed less than six months from the date of the expenditure in question. The implied time distribution seems short.

It is possible that this very low level of discounting has been arrived at by considering the time distribution of “health effects”. We would argue that this is not the relevant focus of the discounting process, which should be the quality of life benefits *arising* from a given health effect. That is: the outcome, and therefore the time-path, which is considered when discounting the QALY gains from a morbidity-reducing intervention is not the time at which the treatment is *applied* – replacing a hip, for example – but the period of time over which the quality of life gains associated with the new hip are *enjoyed*. So even if CHERP81 is correct to assume that all unobservable health effects should be assumed to occur during the year in question, it is still not correct to assume

that the quality of life effects are also immediate and therefore do not require any discounting.

The assumption that the QALYs arising from quality of life gains associated with health interventions can be regarded as arising exclusively in the year in which the intervention takes place is less plausible given that such QALY gains are extrapolated rather than observed. The equiproportional burden reduction method of inferring QALY changes tells us nothing about their distribution through time. It seems "optimistic" therefore to assume that these QALY changes are enjoyed immediately, given that health interventions delivering morbidity reductions often produce their QALY gains over a lengthy period.

Our preferred assumption would be that QALY gains follow the same approximate time path as life years gained<sup>14</sup>. As well as being more intuitively appealing than either assuming no discounting (as is the case with the central £12,936 estimate) or assuming that all QALY gains accrue immediately (which is the basis of the 1.6% discount rate implied by the £13,141 estimate), this enables us to discount the unit-QALY denominator of the cost per QALY estimate directly.

We are not able to offer a firm estimate of the effect of assuming a more "conservative" distribution of QALY gains through time but, in general terms, taking the 4.5 year average<sup>15</sup> discounting period for gains to be discounted cited above and applying it to the undiscounted 1 QALY denominator of the £12,936 threshold at the NICE-standard rate of 3.5% estimate would imply a discounted threshold estimate of £13,724. This crude estimate would benefit from more detailed understanding of the implied time path of life years gained, and the assumed time path of QALYs underpinning the undiscounted £12,936 figure. We are confident, however, that it provides a better estimate than assuming that the QALY benefits of health gains (rather than the health gains themselves) all occur immediately.

We conclude that the net effects of the assumptions as to discounting implied by both the £12,936 headline estimate and the £13,141 partially-discounted estimate appear to be "optimistic".

### ***Assumptions as to distribution of unallocated expenditure increases***

CHERP81 has concluded, in relation to the now-deleted assumption 7 in Box 5.1, that the 28% of spending not allocated by their elasticity estimates should be divided equally across all PBCs in proportion to their estimated elasticities. While we accept that CHERP81's new assumption is plausible, there exist a wide range of alternative, equally plausible assumptions, which would produce very different estimates of the threshold.

The authors of CHERP81 defend their change in assumption, in the accompanying document "Why is the central estimate of the cost per QALY lower?", arguing that:

---

<sup>14</sup> This assumption is, if anything, optimistic, given that QoL improving treatments are likely to affect a healthier population than mortality-delaying ones, so the true time-path for QoL gains may well be longer than that calculated for mortality reduction. This would lead to further discounting and a higher threshold.

<sup>15</sup> This is roughly consistent with the estimate of 4.56 life years gained per averted death in Table 5.1 on p. 74.

*The original analysis suggested that approximately 25% of a change in overall expenditure would be allocated to the mental health PBC ... if this was applied to changes in total expenditure in subsequent years it would predict a rise in the proportion of total expenditure allocated to the mental health PBC. However, the proportionate share of total spend on this PBC has been almost constant over this period.*

We note that this is a somewhat arbitrary test of the validity of the overall elasticities including the unallocated expenditure. If the assumed allocation of funds is to be modelled using the actual behaviour of PBC budgets after 2008 then it would be helpful to use more than one data point to fine-tune the distributional assumptions used, since any potential allocation of the 28% of expenditure changes not accounted for by CHERP81's model is theoretically permissible.

By way of contrasting the observed result for mental health spending, which lead CHERP81 to alter its assumed distribution of unallocated spending: a brief analysis of spending changes across the "big four" PCBs (cancer, circulatory, respiratory and gastro-intestinal, which are responsible for most observed mortality reduction) between 2008/9 and 2011/12 shows, for example, that the expenditure elasticities used to estimate CHERP81's £12,936 figure greatly overstate the responsiveness of spending on circulatory problems (PCT 10) relative to what has been observed in practice. CHERP81 assumes an elasticity of 0.894 for circulatory expenditure after the additional allocation, while our analysis of actual spending decisions by PCTs between 2008/9 and 2011/12 show an elasticity of only 0.22. This is significant since circulatory spending represents the lowest cost means of preventing mortality, so any reduction in its predicted spending will tend to substantially increase the threshold.

Extending this analysis across the remaining "big four" programmes, we find that if the observed changes in spending between 2008/9 and 2011/12 are used in place of CHERP81 assumed figures, overall change in spend drops by 30%, to a figure slightly below that used by CHERP81 in arriving at its earlier £18,317 threshold estimate. As such, the assumption adopted by CHERP81 in the final version of its paper, predicting increased responsiveness in spending on the "big four" relative to the calculated expenditure elasticities appears to lack face validity.

An alternative way to assessing face validity of CHERP81 new allocation of funds is that its initial assumption, which did not allocate any unexplained funding increase to the "big four," explains observed elasticities in overall big four spending almost perfectly (0.57 million predicted vs 0.56 million actual) while its new assumption predicts weighted average elasticities of 0.80. Given the sensitivity of the threshold to changes in big four spending, particular circulatory spending, it is not obvious that the new assumption is preferable to the previous one, which accurately predicts big four spending.

We accept the authors' caveat that this kind of observed "inconsistency might arise if the cross sectional analysis was a poor approximation to a fully specified predictive model"<sup>16</sup> but, given the time lags between NICE decision making and funding impacts on PCTs, we

---

<sup>16</sup> On p1 of their accompanying explanatory note "Why is the central estimate lower?"

do not believe that decision makers can safely place reliance on a model which is highly sensitive to the allocation of unmeasured funds across PBCs and which is potentially not able to predict the results of a change in NHS funding on PCT spending over time.<sup>17</sup>

### Overall effect of assumptions

CHERP81 sets out the assumptions on which its “best estimate” of £12,936 depends.

We argue above that out-of-period health effects of expenditure are properly assumed by CHERP81 as having a net value of zero in expectation, but that this assumption is neither “conservative” nor “optimistic” in relation to the estimate of the threshold.

Even if the assumption in relation to future health benefits is regarded as “conservative” with respect to the threshold, it is more than offset by the “optimistic” assumptions in relation to the patient health gains associated with averted mortality.

We have argued in the previous section that the assumptions made in relation to extrapolating unobservable QALY gains are likewise “optimistic” (tending to understate the optimal threshold), given that no causal model is presented tying together PCT behaviour in relation to mortality and QALY burden.

This is true notwithstanding the observation that some non-marginal mental health decisions reflect a lower cost per QALY than the estimate at the margin – if PCTs are not optimising across QALYs there is no reason to expect their decisions to follow predictable cost per QALY curves.

The assumptions as to discounting, both the undiscounted headline best estimate and the 1.6% discounted figure rely on “optimistic” assumptions as to the time distribution of unobservable QALY gains. We suggest such gains be modelled as following the same time path as life years gained, which implies a higher discount rate and therefore a higher threshold.

The cumulative effect of the CHERP81 assumptions is “optimistic” in our view, and, if this is the case, means it is likely to understate the true value of the threshold, perhaps significantly.

Using revised assumptions on points (i) to (v) above, the estimate of the threshold based on CHERP81’s data is above £32,114 (assuming return to normal population mortality for averted deaths) but significantly lower than £67,664 (assuming two year lifespans for averted deaths), with the discounted value of £13,724 (the equivalent of CHERP81’s undiscounted central estimate of £12,936) representing a lower bound<sup>18</sup>.

---

<sup>17</sup> We quoted earlier from the additional note the authors had posted which included “[W]e have not constructed or estimated a predictive model ... We do not infer changes over time from the cross section”. The source is referenced in footnote 4 above.

<sup>18</sup> The values £32,114 and £67,664 (£30,270 and £63,844 undiscounted) have not been revised from the June 2013 version of the report and may therefore be an overestimate. However, for consistency, we use the figures in the November 2013 version of CHERP81.

All these estimates continue to accept CHERP81's assumption that PCTs are as good at reducing quality of life burden as they are at reducing mortality burden. Different assumptions about this relationship would produce higher threshold estimates.

## THE IMPACT OF CHANGES TO EXPENDITURE AND PRODUCTIVITY ON THE ESTIMATED THRESHOLD

### Summary

General NHS expenditure grows every year in nominal terms. To the extent that this is due to input price inflation it can be expected to be reflected in increases in the nominal opportunity cost per QALY over time.

When general NHS expenditure goes up in real terms, the effect of additional increases in spending on health can be expected to go down due to diminishing marginal productivity: the best opportunities for improving health may be expected to be adopted first. CHERP81 provides evidence that this relationship holds in practice.

Since the productivity of general NHS spending falls as the overall budget goes up in real terms, we would expect the opportunity cost of a QALY, and the threshold, to rise when the NHS budget increases in real terms. CHERP81 accepts this is true in theory, but argues that this is not likely to happen in practice because offsetting increases in NHS productivity will tend have the opposite effect, reducing the opportunity cost of QALYs.

There is good reason to expect extra spending to increase marginal costs, but there is no reason to expect increases in productivity to consistently raise or lower marginal costs. This is because extra productivity has two effects, increasing the amount an employee can produce in an hour, but also increasing overall output, which acts like a budget increase and tends to reduce the value produced by the last pound spent when there is diminishing marginal productivity. So when NHS productivity improves, we cannot be sure whether the health gain from additional expenditure goes up, because PCTs are now able to spend the extra money more productively, or down, because a more productive PCT is already able to perform the most worthwhile activities within its original budget and so achieves less extra health benefit from extra spending.

While there have been large recorded increases in NHS budgets since the time for which the £12,936 estimate was calculated, there is little evidence of increased productivity, and nor would we necessarily expect there to be. So even if productivity increases unexpectedly happen to have a systematically inverse effect to that of rising budgets, we would predict that the effects of budget increases would dominate, and that the threshold would have risen.

CHERP81 argues that a comparison of threshold estimates from two different years shows that a budget increase over that time did not lead to a rise in the calculated threshold, but CHERP81's estimates are much too uncertain to reliably test the relationship between threshold and budget using just one year's changes in budget and threshold.

Finally, CHERP81 argues that not all budget increases affect PCT budgets, and that budget increases might end up being directed to less productive areas. But the reverse is also true of budget *decreases* due to approval of a new technology. The general problem with this argument is that it ignores how CHERP81 has designed its model by assuming that all changes to NHS spending directly and consistently affect PCT budgets.

In future it might be useful to have a more detailed model of how particular budget changes affect particular kinds of expenditure. CHERP81 estimates the threshold using a model that says all budget changes, up or down, flow through to the PCTs and PBCs. Increases and decreases in the budget available to fund PCTs should be treated symmetrically, and the CHERP81 best estimate “implicitly assumes that any budgetary shock affects only PCT funding and leaves non-PCT funding unchanged.” It is therefore not appropriate to point to the possibility that budget increases might accrue to non-PCT spending, having assumed away this possibility in arriving at the central threshold estimate. When CHERP81 considers a more complex spending model, which has some portion of budget changes not directly affecting the PCTs it concludes that using this model would increase its threshold estimate by 17.7%, relative to the simpler model it actually uses to calculate the threshold.

We conclude that the positive impact of budget increases on the threshold over the period 2008 to 2014 will exceed any negative effects from productivity increases and therefore that the optimal threshold for 2014 is likely to lie above its estimated nominal value for 2008/9.

## **Analysis**

### **Increased overall expenditure unambiguously increases the optimal policy threshold**

General NHS expenditure grows every year in nominal terms. To the extent that this is due to input price inflation it can be expected to be reflected in increases in the nominal opportunity cost per QALY over time. CHERP81 argues that this outcome may be offset in practice by productivity improvements but, as we explain below, that is unlikely to be the case generally.

CHERP81 estimates a constant marginal elasticity relationship between PBC expenditure and mortality. This relationship implies diminishing returns in the absolute number of deaths averted in response to absolute increases in expenditure.

Section 5.5 of CHERP81 makes an argument to the effect that marginal productivity is diminishing faster than the rate implied by its core model, on the basis of partitioning PCTs into high and low spending groups and comparing their marginal levels of productivity. We can reach the same conclusion by noting that CHERP81 estimates an increasing returns to scale relationship between expenditure and mortality for several PBCs, so that doubling expenditure would reduce overall mortality by more than half. Since this would quickly lead to predictions of negative overall mortality for large but not inconceivable increases to expenditure,<sup>19</sup> it is clear that the level of diminishing returns

---

<sup>19</sup> The estimated outcome elasticity for the respiratory PBC, for instance, predicts negative mortality for budget increases over 49% - see CHERP81 Table 3.4. We do not suggest this represents a flaw in the CHERP81 model, which is intended to capture marginal effects.

implied by this model at the margin understates the rate at which returns to increases expenditure will diminish in response to non-marginal changes.

The core CHERP81 model for marginal changes predicts that the additional benefit of an increase in general NHS expenditure, measured in terms of pounds rather than percentages, is decreasing as the general NHS budget rises. In addition, we agree with CHERP81 that the rate at which productivity decreases is actually faster than that predicted by the basic model, because there must be some additional decrease in marginal productivity when budget changes are large. As a result, any large (non-marginal) increase in the overall budget will unambiguously increase the average marginal cost per QALY as marginal productivity falls, and therefore increase the optimal value of the policy threshold.

CHERP81, in section 5.6, endorses this conclusion:

*"If overall expenditure increases ... then, other things equal, the threshold would also be expected to increase"* (p88)

### **Increases in productivity have an ambiguous effect on the optimal policy threshold**

Section 5.6 of CHERP81 argues, in relation to productivity changes that:

*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.* (p88)

And therefore that:

*It is not necessary the case that the threshold will rise with overall expenditure or even with NHS prices.* (p88)

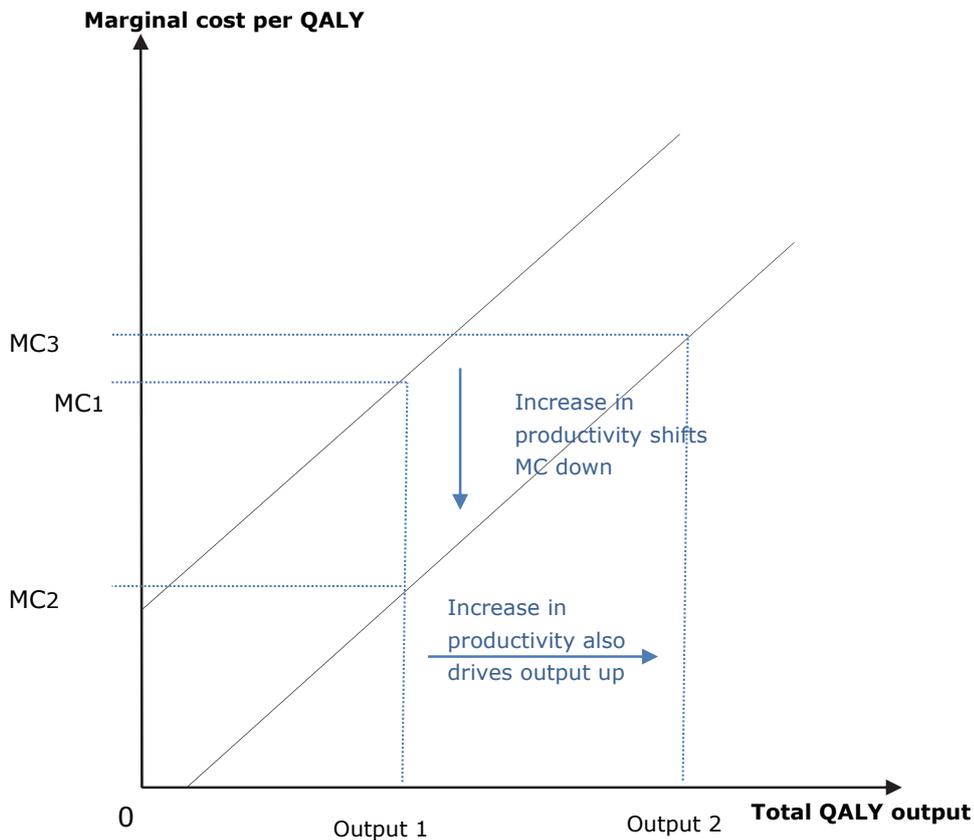
However, the expected effect of changes in *average* productivity on *marginal* productivity is ambiguous and therefore there is no reason to expect that productivity changes will have any countervailing effect on the rate of increase in the threshold, or any *a priori* predictable effect on the threshold at all.

To see this, note that increased average productivity implies only that greater total health outcomes can be achieved by using the same quantity of health care inputs, but it does not necessarily imply, once expenditure is reallocated under a fixed budget, that the *marginal* productivity of the last pound spent will have increased. The resources released by the increased productivity in activities up to the margin could have been used to treat previously extra-marginal patients or offer new services for which the marginal cost per QALY gained is above the pre-existing average cost per QALY. The effect of general productivity gains on the marginal cost per QALY is therefore ambiguous and, unless this ambiguity is resolved, will not help in predicting future changes in the optimal policy threshold.

To see this consider Figure 1 below, showing the relationship between marginal cost per QALY and total output. While an increase in overall productivity will increase productivity, on average, at any margin, pushing the marginal cost curve down and

reducing marginal cost per QALY from MC1 to MC2, the increase in productivity across the output up to the margin has the effect of increasing output holding inputs constant, from Output 1 to Output 2. The effect of the expansion in output, given increasing marginal costs per QALY (arising from decreasing marginal productivity of healthcare resources), are, in this simple illustration, to increase marginal costs from MC2 to MC3.

**Figure 1: Marginal cost per QALY as a function of total QALY output**



In this example, the effect of the increase in output is sufficient to outweigh the effects of increased marginal productivity at constant output, contrary to the CHERP81 assumption that rising average productivity increases marginal productivity. The case shown in this diagram will not necessarily hold in other circumstances: all we can say is that the effects of increases in productivity on marginal productivity, and therefore marginal cost, are ambiguous. As such, the role of average productivity changes plays no part in our discussion of likely changes in marginal productivity through time until we know more about the overall direction of the relationship.

Even if we believed that average productivity increases would systematically increase marginal productivity within a given NHS budget rather than having the ambiguous effect predicted above, there is no evidence available about NHS productivity trends over time in terms of the average cost of producing QALYs, as there are no data on the quantity of QALYs produced by the NHS in total. ONS and other studies of trends in NHS productivity measured in terms of activity – some quality adjusted, some not – rather than health gains, have found little or no improvement (Castelli et al, 2011; Jones and Charlesworth, 2013; Massey, 2012). In its most recent, December 2012, study of

productivity in publicly funded health care the Office for National Statistics concludes: "For most of the 15 year period since 1995, productivity has remained broadly constant. The period 2003 to 2006 saw some improvement in productivity which has since levelled off or slightly declined" (Massey, 2012). This is perhaps unsurprising in view of the face to face nature of much health care: ten minutes of time with a patient will always take ten minutes and labour costs per minute tend to rise over time (Baumol, 1995).

### **The optimal future NICE threshold depends on the NICE threshold when data were gathered**

Any change to the policy threshold will lead to a transfer of funds into general NHS expenditure (in the case of a lower threshold) or out of general NHS expenditure (in the case of a higher threshold). As a result of diminishing marginal returns (see our detailed discussion above), the optimal value of the policy threshold is an increasing function of the overall level of NHS expenditure. So, the £12,936 marginal cost per QALY depends on the balance between general NHS and technologies funding associated with the actual threshold used by NICE in 2008, which was stated by NICE as being between £20,000 and £30,000. Implementing the lower threshold would mean that the assumptions on which the £12,936 figure was based would no longer be accurate.

To put it in more general terms: any decision to significantly lower the policy threshold to a value calculated under the operation of an earlier, higher, threshold will, itself, increase the optimal value of the policy threshold.

So, the optimal value of the policy threshold depends endogenously on its own value, because of the influence the choice of threshold has on NHS spending and therefore on marginal NHS productivity. The CHERP81 best estimate of the threshold does not take into account this relationship and therefore understates the optimal value of the threshold. Although this effect may be small, it should be acknowledged.

### **How much do increases in the NHS budget increase PCT spending?**

In considering how the threshold will change through time, CHERP81 further suggests

*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. ...Therefore, it is growth in expenditure on more 'discretionary' parts of NHS expenditure ... which [is] most relevant....Therefore, it is not self evident that the threshold has grown over recent years, despite real increases in the NHS budget<sup>20</sup>*

We do not agree, since we believe that the PCT behaviour implied by the CHERP81 model requires us to treat the injection of funds into the NHS via budget increases and the removal of funds from the NHS via approval of new technologies as two sides of the same coin, and to model them accordingly. The CHERP81 model of how PCT spending will vary in response to budget shocks does not leave room for a distinction between

---

<sup>20</sup> CHERP81, p. 88.

'discretionary' and 'non-discretionary' expenditure, as changes to expenditure are calculated as arising automatically from constant elasticities. Even where an increase in expenditure is funded from a particular budget, CHERP81 assumes that expenditures will readjust so that the actual impact of any increased expenditure is spread across PBCs at a fixed rate. Consistent with this approach, we need to treat increases in NHS expenditure, even where those increases are notionally allocated to a particular goal, as effectively increasing PCT expenditure and therefore decreasing marginal productivity.

In practice not all changes to the NHS budget will find themselves reflected in changes to individual PCT spending. The approach adopted to non-PCT general NHS expenditure in CHERP81 Appendix B (page 86) is as follows:

*The cost of a life (year) estimates presented above are based on the impact of a 1% exogenous change in total net PCT spend. ...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. ... 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.*

It is important to be consistent in the treatment of the role of non-PCT funding in response to positive and negative budget shocks. If the cost per QALY estimate assumes away changes to non-PCT spending in response to disinvestment, it cannot simultaneously be argued that budget increases may be subsumed within non-PCT areas or allocated in a non-discretionary fashion to activities which do not displace marginal PCT investment opportunities. In the absence of any information about responsiveness of PCT budgets to overall budget changes, or of PCT productivity to the non-discretionary proportion of their budget we should assume, as CHERP81 does elsewhere, that rising budgets "only affect PCT funding and leave non-PCT funding unchanged" (p30) and that such increases are allocated according to CHERP81's calculated expenditure elasticities.

### **What can we learn from a single observed change to the estimated threshold?**

The key CHERP81 estimate of the threshold is for the financial year 2008/9. CHERP81 also estimates values for 2006/7 and 2007/8 in order to see how the threshold changes over time as total NHS spending changes: see Table 3 below. CHERP81 rejects comparison between 2006/7 and later years because of major changes in the way that costs were apportioned when allocating them to PBCs (which reinforces our earlier concerns as to the potential variance of threshold estimates due to cost allocation issues within the data), and compares only the 2007/8 threshold estimate with that for 2008/9.

**Table 3. CHERP81 “best estimate” of the threshold 2006/7-2008/9**

<b>Money of the day</b>	<b>2006/7</b>	<b>2007/8</b>	<b>2008/9</b>
“Best estimate” cost per QALY gained	£10,187	£13,554	£12,936
<i>% change on previous year</i>	-	+33.1%	-4.6%

Sources: CHERP81 Table 4.20 for 2006/7; Table 5.3 for 2007/8 and 2008/9

This comparison reveals that the (undiscounted) best estimate of the threshold cost per QALY falls by 1.6% in cash terms in that one year period – from £13,554 to £12,936. Given that the estimate of threshold in any single year is highly uncertain relative to its underlying “true” value – as discussed in detail above and in CHERP 81 Appendix C – we do not believe any significant probative value can be assigned to this number, particularly given the strong theoretical prediction both we and CHERP81 share of a negative relationship between overall budget and marginal productivity, and given that the total spend on PBCs increased throughout the period covered by Table 3 above.

Relevantly, CHERP81 notes, on p90:

*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 necessarily 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.*

We agree with the first sentence. We also agree that the nominal value of the threshold does not necessarily change in the same direction as the nominal size of the NHS budget. But we are not convinced that one can simultaneously conclude that falling budgets are likely to reduce the optimal threshold over time, while rising budgets will not necessarily increase it. These are perfectly symmetrical events in the linear, constant elasticity model of PCT decision making adopted by CHERP81. To suggest that the threshold is more likely to fall at a time when real budgets are falling is symmetrically to suggest that the threshold is more likely to rise at a time when real budget growth is rising and that the threshold is more likely to be flat when real budget growth is flat, since (*ceteris paribus*) none of the variables relevant to calculating the threshold are changing. There is no basis for the suggestion that increasing marginal returns take effect when growth in budgets, rather than budget themselves, is falling, or when budgets or budget growth is constant.<sup>21</sup>

---

<sup>21</sup> Note that total nominal NHS England spending grew 14.3% from 2008/09 to 2011/12 (latest data, see HM Treasury, 2013b), which represents a 7.2% real terms rise, and has been flat in real terms since 2011/12.

Further, we note that, *ceteris paribus*, a rise in the nominal prices of NHS inputs matched by a rise in the nominal NHS budget just sufficient to pay for that would be expected to leave the real terms threshold unchanged and therefore the nominal threshold to have risen in line with the input price inflation. Over the period 2008/9 to 2012/13 (latest data available) the GDP deflator rose 9.8% and is expected by HM Treasury to have risen by another 2.3% over the following year (HM Treasury, 2013a).

## HOW SHOULD WE ADDRESS UNCERTAINTY ABOUT THE THRESHOLD?

### Summary

CHERP81 demonstrates that deciding whether to approve a new health technology depends not only on the opportunity cost of a QALY, but also on the degree of uncertainty in the threshold estimate. It also indicates that the consequence of symmetrical uncertainty is an asymmetric response. There is a difference between the consequences of underestimating the threshold (the true, underlying value is above the central estimate) and overestimating it (the true value is below the central estimate) by any given £ amount. This means, for example, that we would rather incorrectly reject a technology which is, say, £2,000 per QALY under the “true” value of the threshold (underestimate the threshold) than incorrectly accept one which is £2,000 per QALY above the true threshold (overestimate the threshold).

CHERP81’s analysis is however only correct if the likely size of underestimates and overestimates is the same. If overestimates of the threshold are more common but *likely to be smaller in size* than underestimates then it is no longer certain that we would be better off setting the threshold for accepting new technologies below the estimated expected opportunity cost of a QALY.

Because the positively skewed distribution of the threshold calculated by CHERP81 does tend to produce underestimates larger than the overestimates on average, we cannot be sure whether the uncertainty implies that the threshold used to evaluate technologies should be above or below the estimated opportunity cost of a QALY. We can be sure that, should the threshold need to be lower, it will be by a smaller amount than suggested by analysis of unskewed errors.

More generally, CHERP81 modelling of uncertainty focuses on only one of the three sources of uncertainty they identify: namely that from statistical sources of uncertainty, i.e. uncertainty around the parameters. CHERP81 uses assumptions about their distribution to arrive at probabilities of being above or below NICE’s current stated threshold range of £20,000 to £30,000. It also discusses structural uncertainty, i.e. the possibility that the model is not correctly specified, but it is not easy to translate this into something quantifiable. Whilst we find no issue with the model specification, there is inevitably some uncertainty. There is a third set of “other assumptions” which we have discussed earlier in the paper. The potential size and direction of these effects is discussed, and in some cases CHERP81 gives values for one variable sensitivity analysis to estimate the impact on the estimate of the threshold if some of the “other assumptions” were not valid. However, the effect of several of the other assumptions being wrong is not stated. Some other key underlying assumptions, such as the relationship between length of life and quality of life gains, are not tested at all.

## Analysis

Section 5.4 of CHERP81 demonstrates a positive expected QALY loss from approving technologies with certain ICERs up to the threshold in the presence of symmetrical uncertainty about the threshold estimate itself. If this model of uncertainty reflects the situation faced by policy makers, then we agree with CHERP81 that it calls for an optimal policy threshold below the central estimate of the NHS opportunity cost per QALY.

While CHERP81's conclusion follows logically from its representation of uncertainty, we are not sure that this model, which relies on comparing the effects of equally-sized errors, is an accurate summary of the question to be answered by policy makers.

The key point to note is that, as stated on CHERP81 p82:

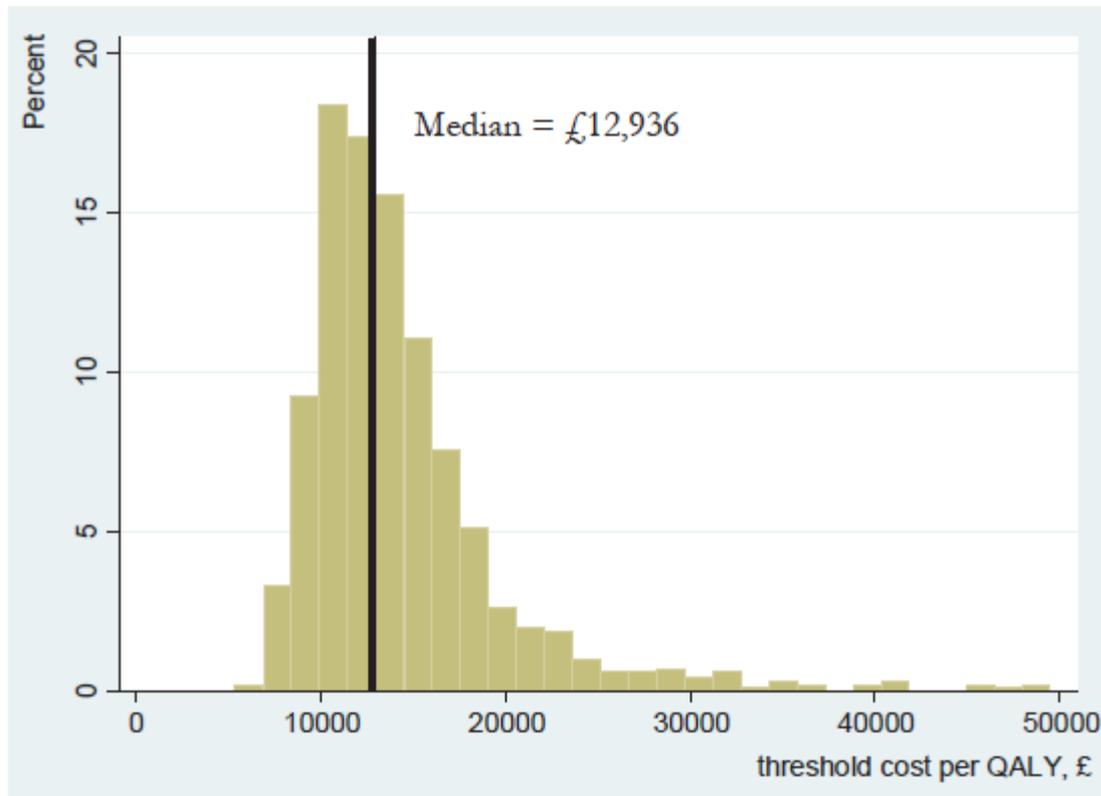
*How much lower a policy threshold should be set below the mean or expected value depends on [inter alia]...the skewness of the distribution of cost per QALY threshold (a positive skew tends to offset these effects - see Figure C8 in Appendix C ).*<sup>22</sup>

---

<sup>22</sup> This appears to be contradicted by the footnote on p81 of CHERP81 that:

*“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.”*

**Figure 2: Distribution of cost per QALY threshold, Figure C.8 from Appendix C to CHERP81**



Intuitively, the effect of skewness on the degree and direction of difference between the central estimate of the threshold and the optimal policy threshold will depend on the nonlinear effect of the size of errors above and below the “true” underlying threshold. Errors which arise in circumstances where the true value of the threshold is very low – a point close to the origin - (large overestimates of the threshold) lead to very large reductions in the optimal policy threshold relative to its central estimate. As the “true” value of the threshold approaches zero (equivalently, when the size of the overestimate approaches the central estimate of the threshold) then the number of QALYs forgone as a result of endorsing a product with an ICER equal to the central threshold estimate approaches infinity.

Since a positive skew in the uncertainty as to the threshold reduces the expected size of overestimates of the threshold, while increasing the probability that the threshold is overestimated it will, over most plausible ranges reduce, but not eliminate, the gap between the central threshold estimate and the optimal policy threshold.

Consider, by way of example, a modification of the distribution of uncertainty used on p81 of CHERP81 to demonstrate the effects of uncertainty. Instead of considering only equally sized (£10,000) and equally probable (50%) errors about the central estimate of £20,000, we model a skewed distribution with expected errors in each direction which are equal to each other and equal to the expected size of the CHERP81 example errors.

Assuming a decision regarding £1,000,000, the uncertainty in the CHERP81 example will lead to an opportunity cost of 66.7 QALYs<sup>23</sup> rather than the 50 QALYs calculated based on the central estimate of £20,000. The uncertainty adjusted threshold is £15,000.

Taking a negatively skewed distribution of uncertainty, with the expected size of the errors remaining equal to £5,000 (=£10,000x50% as in the CHERP81 example) but distributed such that there is a 70% possibility that the true threshold lies above the central estimate (is currently being underestimated) and a 30% chance that it lies below. We have a 30% chance that the true threshold is equal to £3,333 (=£20,000-£5,000/0.3) and a 70% chance that the true threshold is equal to £27,143 (=£20,000+£5,000/0.7). Under this negatively skewed distribution, the opportunity cost of a £1,000,000 decision is increased from 66.7 QALYs in the unskewed case to 115.8 QALYs<sup>24</sup> implying an uncertainty adjusted threshold of £8,600.

For the positively skewed case, we assume that there is a 70% chance the the true threshold lies below the central estimate (it is currently being overestimated), and therefore a 30% chance that the true value lies above the central estimate (it is currently being underestimated). Keeping the expected size of errors equal to those in CHERP81, this implies a 70% chance of a threshold of £12,857 (=£20,000-£5,000/0.7) and a 30% chance of a threshold of £36,667 (=£20,000+£5,000/0.3). In this, positively-skewed, case, the number of QALYs foregone as a result of spending £1,000,000 is 62.625 implying an uncertainty adjusted threshold of £16,000. This is fewer QALYs lost than the unskewed case (66.7) and far less than the negatively skewed case (115.8). The positive skewed distribution has an uncertainty adjusted threshold (£16,000) which is closer to estimate of £20,000 than of the unskewed case (£15,000). We can see that the negatively skewed case has a much bigger move away from the £20,000 estimate at £8,600. In general, consistent with this example and with the text of CHERP81, we can expect positively skewed distributions to reduce the gap between the central estimate and the optimal policy threshold.

An examination of Figure 2 above, which reproduces Figure C.8 from Appendix C to CHERP81, reveals that the distribution of the estimated threshold relative to its true underlying value (based only on those forms of error which are susceptible to being modelled in this fashion) displays a positive skew. This is as we would expect, since the marginal cost of a QALY in the NHS is bounded below at zero (since economic theory requires non-negative marginal costs for goods), but unbounded above. For this reason we would expect any non-statistical sources of error in the estimate of the threshold to reinforce the level of right-skewness observed in Figure 2.

Not only does a positive skew “tend to offset” the effects of the asymmetry in the effects of investment and disinvestment identified by CHERP81, it is possible to show that for a sufficiently high level of skewness the predicted effect is reversed and the effect of a, say, £10,000 per QALY underestimate of the threshold, measured in QALYs forgone, is greater than the effect of £10,000 per QALY overestimate.

<sup>23</sup>  $50\% \times \text{£}1000,000/\text{£}10,000 + 50\% \times \text{£}1000,000/\text{£}30,000 = 50\% \times 100\text{QALYs} + 50\% \times 33.33\text{QALYs} = 66.7\text{QALYs}$

<sup>24</sup>  $30\% \times \text{£}1000,000/\text{£}3333 + 70\% \times \text{£}1000,000/\text{£}27,143 = 30\% \times 300\text{QALYs} + 70\% \times 36.84\text{QALY} = 115.8\text{QALYs}$

<sup>25</sup>  $70\% \times \text{£}12,857 + 30\% \times \text{£}36,667 = 70\% \times 77.78\text{QALYs} + 30\% \times 27.27\text{QALYs} = 62.6\text{QALYs}$

It is relatively straightforward to derive the optimal level of the policy threshold as a function of the distribution of errors in the estimate of the threshold cost per QALY and that of the mean ICER. We agree with CHERP81 that the optimal policy threshold should be set so as to minimise the expected number of QALYs forgone. In doing so it is important to bear in mind that modelled error, for which estimates are available, is likely to understate the size and skew of overall error. Any additional, unobservable errors in the estimate are, as with the observable errors, likely to be systematically positively skewed, because they are unbounded above, but bounded below at zero due to the implausibility of negative marginal costs implied by a negative estimate of the threshold.

We agree with CHERP81 of the need to take into account the actual distribution of errors. It would seem to us, however, that without obtaining such additional data there is no a priori reason to expect the policy threshold to lie significantly below the central opportunity cost estimate of the marginal cost per QALY in the NHS.

In addition to the skew of the errors modelled by CHERP81, the central threshold estimate may vary relative to its underlying "true" value as a result of errors in the assumptions made by CHERP81. These assumptions are not incorporated in the distribution shown in Figure 2 above, because there is no evidence of their probability distributions and, in some cases, because there is no evidence as to their potential size.

As far as is possible, given the absence of quantitative data in relation to some assumptions, it would be useful for future iterations of the CHERP81 research to report an overall aggregate estimate of uncertainty, based not only on structural and parameter uncertainty but also reflecting uncertainty arising from study design and assumed relationships between observed and unobservable variables. At a minimum, sampling uncertainty around threshold estimates based on plausible alternative assumptions, such as those presented in Table 5.1, should be reported.

## **OPTIMAL POLICY THRESHOLD WHERE THERE ARE NON-MARGINAL EFFECTS**

### **Summary**

CHERP81 argues that the threshold for accepting technologies that will have a large, i.e. more than marginal, impact on the budget should be higher than the opportunity cost of a QALY, because this value will go up when we remove a significant amount of funds from general NHS expenditure, due to the same diminishing marginal returns effect discussed above.

We agree with CHERP81's conclusion, but point out that the effects on productivity from approving expensive technologies are just the flip side of the effects of rising budgets, and no technologies have anything like enough budget impact to offset increases in the NHS budget between when CHERP81's data were gathered and today.

Overall, CHERP81 is not entirely consistent in applying the model it uses to estimate the health benefits of spending, which relies on the assumption that the effects of one more pound of expenditure are equal and opposite at the margin to those of one less pound,

to changes in the overall budget and nor is it consistent in its choices of when to consider the potential for non-marginal effects.

## Analysis

As we note above, in relation to the effect of budget changes on the optimal threshold, general NHS expenditure displays diminishing marginal returns in excess of those predicted by the constant budget mortality relationship modelled by CHERP81.

We agree with CHERP81 that these diminishing marginal returns will imply a lower optimal policy threshold for decisions which would result in relatively large transfers out of the general NHS budget. In fact this process of non-marginal disinvestment is directly analogous to a change in the NHS budget and, given the symmetry between investment and disinvestment assumed in CHERP81's threshold estimate, we would expect further investigation as to the effects of budget changes on the optimal threshold to reveal the correct level of adjustment required to deal with non-marginal disinvestment decisions.

The specific results of the CHERP81 analysis of non-marginal budget changes are less persuasive, in our view. The first investigation of the relationship between marginal productivity and level of spend is on p90 of Appendix B:

*To test this [diminishing marginal returns] 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 ...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.*

This outcome is somewhat concerning, given the theoretical requirement that NHS spending display diminishing marginal returns overall. With sufficiently different production functions between PCTs one might observe cross-sectional data showing increasing marginal returns even in the presence of diminishing marginal returns as individual PCTs have their funding increased, but this would require not only that PCTs do not share a single production function, but also that their production functions differ *systematically* with level of spend, which is inconsistent with the structure of their threshold estimation model.

Given that the CHERP81 model assumes both a common production function ( $f''$  in equation 6.2 on p29 of Appendix B linking expenditure and mortality) and the ability to extrapolate from cross sectional observations to counterfactual time-series predictions

for individual PCTs, evidence of such systematic deviations would undermine the core rationale for comparing different PCTs to predict how individual PCTs will behave in the future.

An alternative analysis of diminishing marginal productivity is conducted on p93 of CHERP81 Appendix B and summarised on p87, without reference to the increasing marginal returns result:

*...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... and those that are over target...The results confirm what would be expected ... 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 ... Although these cost per life year estimates are not based on the same calculations as Section 4.2.*

While it is comforting that this alternative approach does show diminishing marginal returns, it is unclear whether these results are compatible with the assumption, required to explain the earlier finding of increasing marginal returns, that high spending PCTs are systematically more productive than low spending ones at a given level of output. Such systematic heterogeneity in production functions makes it very difficult to draw inferences across PCTs for any purpose, including estimation of the marginal cost of a QALY.

We agree with CHERP81 that the NHS will display decreasing marginal productivity, but are concerned that the evidence to this effect presented by CHERP81 is inconsistent and seems to cast doubt on the methodology underpinning the estimate of the threshold itself, underlining our earlier discomfort with the use of cross section comparisons between structurally different entities to make counterfactual time-series predictions about within PCT change.

## OTHER APPROACHES TO INVESTIGATING THE THRESHOLD

### Summary

CHERP81 assumes that data on differences between different PCTs at one point in time can be used to estimate how any single PCT will perform as its expenditure changes through time. In other words, it uses cross sectional data to answer the counterfactual time-series question: what is the opportunity cost of different budget changes within PCTs? While it is possible that CHERP81's method could produce accurate estimates of the marginal opportunity cost of a QALY, provided all PCTs are essentially the same on average, it would be useful to directly measure how individual PCTs respond to budget changes, to see if the responses really can be predicted by assuming that PCTs with *lowered* budgets resemble PCTs with *lower* expenditure on a given PBC.

An alternative to the "top-down" approach adopted by CHERP81 is a "bottom-up" approach, where individual PCTs are investigated directly for their responses to budgetary changes.

While there is a range of potential challenges in linking spending decisions to budget changes and interpreting the results as a "threshold", this approach avoids many difficulties associated with the CHERP81 methodology.

### Analysis

#### Understanding PCT behaviour

The theoretical model used in CHERP81 involves each PCT maximising its welfare function by allocating funds across the 23 PBCs, subject to its lump-sum budget. This implies that "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" (CHERP81, Appendix B, page 8).

Many of the marginal QALYs bought in different PBCs are estimated in CHERP81 to cost far more than £12,936. This is made clear in Table 4, below. Each QALY bought in Maternity and Neonates (PBC 18 & 19), for example, is estimated to cost almost £3m. CHERP81 argues that this is unimportant because some of the PBCs with high cost-per-QALY estimates have relatively low "elasticities of the threshold", where the elasticity of the threshold is defined as "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" (CHERP81 p76). We would argue that the huge range of estimates across PBCs means that readers should be wary of putting a large degree of emphasis on a single average cost per QALY figure as a predictor of the displacement that would result from a change in budgets or a new NICE mandate to fund a health technology, in the absence of any evidence to demonstrate that PCTs allocate expenditure in a way consistent with the existence of a predictable cost per QALY threshold.

The relationship estimated by CHERP81 depends on the observed correlation between the budgets of different PCTs and their spending decisions, at one point in time,

carrying over to the decisions of individual PCTs when faced with a change in their budgets across time.

**Table 4. Individual PBC costs per QALY**

PBC	Elasticity of the threshold*	Cost per QALY
Cancer	0.35	£16,997
Circulatory	1.43	£7,038
Respiratory	3.05	£1,998
Gastro-intestinal	0.58	£7,293
Infectious diseases	0.21	£20,829
Endocrine	0.80	£3,124
Neurological	1.45	£5,480
Genito-urinary	0.14	£43,813
Trauma & injuries	0	NA
Maternity & neonates	0.00	£2,969,208
Disorders of Blood	0.19	£28,305
Mental Health	0.93	£49,835
Learning Disability	0.03	£78,854
Problems of Vision	0.07	£76,850
Problems of Hearing	0.12	£19,070
Dental problems	0.13	£55,916
Skin	0.03	£174,775
Musculo skeletal	0.47	£20,254
Poisoning and AE	0.01	£163,766
Healthy Individuals	0.01	£1,483,012
Social Care Needs	0	NA
Other	0	NA

Source: CHERP81 Table 5.2, p76

However, an individual PCT's response to a budget change will not necessarily be to mimic better or worse funded PCTs, because expenditure decisions may occur through different institutional mechanisms and pursuing different balances of objectives. PCTs may consistently behave in different ways to one another. CHERP81 does not in practice provide a model of marginal decisions, it provides a cross-sectional analysis across single years of how different PCTs look with different levels of spending, and assumes that can function as a model for how, in aggregate, marginal decisions will be made.

### A "bottom-up" approach

One set of alternative approaches to estimating the marginal opportunity cost of NHS spending decisions is to deal directly with the marginal investment and disinvestment

decisions taken by health care organisations responding to changes in their budgets. This kind of approach was taken by Appleby et al (2009).

It is possible to generate data on marginal spending decisions in the NHS relatively cheaply. In Scotland this is already done: NHS Boards (who take the role of PCTs/CCGs in England) were asked in 2010, 2012 and 2013 by the Health and Sport Committee of the Scottish Parliament about plans for service investments and disinvestments at the margin (see for example: HSC, 2012). Importantly, Scottish Health Boards are also asked about desirable service developments which were ultimately unfunded, allowing researchers to identify investment plans which were rejected before being made explicit in documentation. Generating this type of data makes possible direct observations of marginal decisions and thus estimation of the opportunity cost of NHS spending.

Another way of identifying the opportunity cost is to analyse the responses of health care commissioners to exogenous “shocks” to their budgets that have occurred in the past. An example of such a shock would be NICE’s approval of a cost-increasing technology, which must be funded within PCTs’ current budgets. Assuming that PCTs treat their budgets as completely fixed, PCTs would be expected to disinvest from some other technology they currently fund in order to accommodate the new technology. This would allow researchers to identify the cost effectiveness of the displaced service and thus estimate the opportunity cost of the decision in practice.

Such “bottom-up” approaches can reveal directly how PCTs behave in real time in response to real budget changes, without the need to proxy these decisions indirectly by cross-section comparison with different PCTs who allocate their fixed budgets very differently between PBCs at a given point in time.

### **Potential caveats to alternative approaches**

#### ***Margins, thresholds and opportunity costs***

PCT decision makers may not use cost per QALY evidence. As is noted in CHERP81, although Appleby et al (2009) were able to identify a number of marginal spending decisions, it was found that these were not made using cost effectiveness analysis. Instead, the Directors of Public Health at the six PCTs examined cited clinical and other non-economic factors as the basis for their investment and disinvestment decisions. As the authors of CHERP81 point out, estimating a cost-per-QALY threshold based on decisions made without the use of cost-per-QALY evidence require careful interpretation.

If PCTs were found not to use QALYs as part of their planning process, this would have important implications for how researchers model PCT behaviour. The underlying theory assumed in CHERP81 as well as other research (that health care managers will, in effect and on average, buy services in decreasing order of cost effectiveness) assumes that health care organisations operate efficiently. The finding in Appleby et al (2009) that non-economic factors are often the basis for spending undermines this assumption.

How health care organisations spend their budgets may be a more complex process than solving a maximisation problem where the objective is to maximise health gain (measured in QALYs or otherwise), subject to a fixed budget constraint. For example, if we relax the assumption that PCTs are operating efficiently or “at the margin”, then

decreasing a PCT's budget over time may have no impact on health – the PCT would simply provide the same service more efficiently. We might still estimate the opportunity cost of budget changes but these opportunity costs might be felt outside the domain of “health” – e.g. by the discomfort felt by NHS staff in working in new ways or under increased pressure to increase efficiency. Also if budgets are not absolutely fixed, so that a new NICE mandate is partly or wholly met by PCTs increasing the size of their budget overspends or reducing their underspends, the opportunity cost will fall on the rest of the economy outside the NHS as the Treasury adjusts its net borrowing (in which case QALYs would no longer be the appropriate unit of comparison).

### ***Causality***

CHERP81 also argues that establishing a causal relationship between changes to a budget and subsequent disinvestment could be difficult. While we acknowledge this potential problem, bottom-up qualitative approaches potentially offer more insights into causation than top-down quantitative methods.

### ***Variance by locality***

As correctly noted in CHERP81, alternative approaches may well result in a number of different estimates of the threshold which may vary markedly by locality. This represents a challenge for NICE if it is to use this information to construct a threshold on a national level. The CHERP81 solution is to estimate the average cost per QALY in the NHS in England derived from cross-section differences between PCTs at a point in time. We would argue that this does not solve the problem but answers a different question and assumes that approach provides an adequate proxy for marginal decision made in response to budget changes. Following a bottom-up approach would produce a sample of estimated marginal opportunity costs in practice in a number of localities, which could be appropriately averaged to arrive at an overall measure of opportunity cost.

### ***Data limitations***

Finally, it is important to note the data problems associated with the alternative approaches, as discussed in CHERP81. Addendum 3 to Appendix C of CHERP81 attempts a bottom-up estimation of the cost-per-QALY of marginal mental health services for depression and schizophrenia. The purpose of this exercise was to inform the direction of bias of the cost-per-QALY estimate for this PBC (given the unavailability of mortality data). In order to identify investment in mental health, the authors used recent NICE guidance documents; for disinvestment, they consulted experts in the field.

CHERP81 found it difficult to find accessible data on investment and disinvestment decisions in specific areas of mental health, and therefore relied on the opinions of clinical experts. In addition, for many of the treatments that were identified by NICE guidance, cost effectiveness data were not available. Finally, since no reliable method of ensuring the marginality of decisions examined was put in place, the results reported by CHERP81 are non-marginal and therefore uninformative with respect to marginal costs. However, we would again argue that, although past studies have encountered difficulties in implementing this type of approach, those problems can be overcome, for example by using the data collected by the Health and Sport Committee in Scotland.

## WAYS FORWARD

As we noted earlier, the CHERP81 report is a complex and impressive attempt to arrive at a usable estimate of the displacement effect of NICE decisions. Given the incomplete nature of the data available, CHERP81 is forced to rely on a series of strong assumptions in extrapolating from what we can observe to what we really want to know. The number and size of the assumptions required means that the estimate presented is highly sensitive to plausible alternative assumptions and hence is very uncertain as a representation of the actual opportunity cost of a QALY in the NHS in England in 2008/9. This means the estimate can be improved by collecting additional data.

In addition to trying other approaches as discussed in the previous section, there are ways in which the CHERP81 analysis might helpfully be developed. We group possible next steps into three categories:

1. Adjustments that can be made within the existing analysis and approach
2. Additional data that could be collected
3. Improvements to the modelling approach.

### Adjustments that can be made within the existing analysis and approach

We have set out several points at which we think the central estimate could be put in the context of the estimates generated by using other assumptions in the report, including:

- Assuming the overall effect of the assumption of ignoring future health benefits and so not lowering the threshold, is probably smaller than CHERP81 states and may be “optimistic” overall if future quality of life benefits are properly discounted and current mortality differences are estimated at a “steady state” equilibrium. The true central estimate of the threshold could be somewhere between the most conservative value (assuming no net future effects, a low - but not lower bound, see footnote 12 - estimate of post-intervention lifespan and health consistent with normal patient averages for patients who avoid death) and the CHERP81 “central” estimate (assuming no net future effects and improved disease burden and normal healthy mortality for patients who avoid death). This would imply that the threshold should lie somewhere between £12,936 and £63,844, these figures being taken from the “All 23 PBCs” estimates in column 3 row 1 and column 2 row 3 respectively of Table 5.1 on p72, a realistic, but still potentially conservative estimate would be £30,270, taken from row 1 column 2.
- Concentrating on the discounted figures. The figures in the preceding paragraph are all undiscounted and are therefore inappropriate for comparison with ICERs calculated using discounted figures. The discounted central estimate reported by CHERP81 of £13,141 assumes that all quality of life improvement generated by PCT spending is enjoyed immediately, which we have argued above seems implausible. Discounting the central and relevant upper bound estimate based on the assumption that quality of life improvement follows the same timepath as life

extension yields a discounted central estimate of £13,724, alters the realistic but conservative estimate to £32,114 and raises the upper bound figure to £67,664 based on a discount rate of 3.5% employed annually over a 4.5 year period beginning immediately following the expenditure of funds. This discount rate does not take into account the probability that some portion of current year benefit arises from past spending and represents the imputed future effects of current spending. If half of all health benefits from current spending are enjoyed in future periods, declining at a constant rate, then the threshold would be expected to rise by a further 3.5%.

- Relaxing the assumption that PCTs achieve quality of life improvement equiproportional to their success in reducing mortality would further raise the threshold. Greater than proportional improvements in unobservable quality of life would lower the threshold.

### **Additional data that could be collected**

Section 5.8 of CHERP81 sets out areas of research and of additional data collection that could improve the estimate of the threshold in the future. These include (pp99 -100):

- To “update estimates of the threshold with more recent and future waves of expenditure and mortality data”
- To obtain direct evidence on quality of life gains from changes in expenditure: “We have demonstrated that these methods of analysis can be applied to quality of life data collected as part of [Patient Reported Outcomes Measures] 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.”
- Using routinely collected primary care outcomes data for Mental Health – a key PBC because of its expenditure size of and because most of it targets quality of life not reducing mortality. “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.”
- “Improved and more recent estimates of incidence ..and duration of disease..from the recently published updated Global Burden of Disease study..Alternatively estimates could be based on [Clinical Practice Research Data Link] CPRD.”
- “Estimating a more complex lag structure based on the evolving panel data [to] provide evidence about the duration of the health effects of changes in expenditure.”

These would all be helpful. The key to improving the estimate is obtaining quality of life data. As CHERP81 indicates, rolling out the PROMs programme, utilising existing outcomes datasets, and expanding the routine collection of outcomes data are key. Such a programme of data collection is also essential if NHS England and CCGs turn out to be able to understand the effectiveness of their programmes of expenditure.

In addition, given the importance of the allocation of unestimated residual spending, we would suggest a formal series of face validity tests, similar to the examination of proportional mental health spending through time which led to CHERP81 change in distributional assumptions, to determine whether the expenditure elasticities including assumed PBC expenditure match observed changes through time.

### **Possible improvements to the modelling approach**

We have argued that it is not possible to derive a reliable relationship between expenditure and QALYs by assuming a fixed relationship between quality of life improvement and life extension. There are two ways to address this issue.

The first would be to model PCTs using a health production function, rather than assuming they rationally pursue particular goals, and to attempt to determine a stable, statistical relationship between expenditure and QALYs. This approach would not require any assumptions as to *why* PCTs make the choices they do, but it would also not allow missing data to be estimated based on assumed relationships between different goals.

Without a testable model of organisational behaviour, we have to be able to observe *all* of the relevant inputs and outputs into the health production process, across the full range of relevant outputs. Assuming that, for instance, quality of life improvement will have a proportional relationship with life extension, or that lower-spend PCTs in a PBC are just a budget increase away from performing like higher-spend PCTs, is not appropriate. These need to be estimated directly. A production function approach to estimating PCT QALY output would require time-series data on quality of life, life extension, idiosyncratic input quality, and probably much else besides.

Whilst more of this data will become available over time (as set out above in the discussion of future research), comprehensive data will be difficult to obtain in the near future, leaving this approach difficult to pursue in the medium term.

There is, however, an alternative, second, way in which QALY gains might reliably be predicted. This would be to start from a more detailed, testable model of PCT behaviour – one which goes beyond saying that PCTs maximise unobservable “health gain”,  $h$ .

The later chapters of CHERP81 suggest that not all budget increases are created equal – that changes in funding can get lost in specific initiatives and that a simple model of fixed expenditure elasticities may not be sufficient to model how changes to the overall NHS budget find their way into programme budgets in particular PCTs. We believe this is true and much more broadly applicable.

Further, it may be that PCTs act according to different imperatives, and hence experience different outcomes, when budgets go up as opposed to when they go down – i.e. there are discontinuities in expenditure elasticity.

Past funding decisions may be “sticky”, so that differences in expenditure today can predict differences in the future – i.e. there is path dependency in expenditure elasticity.

CHERP81 argues, in response to evidence of apparently increasing returns to scale, that:

*... if we drop the assumption that all PCTs are equally efficient ... 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.*

In other words, there is evidence of differences in efficiency between high and low spending PCTs – i.e. systematic heterogeneity in PCT behaviour.

All of these stylised observations about PCT structure and behaviour are potentially amenable to testing, parameterisation and incorporation into an econometric model of PCT objective functions. A more flexible model of response to budget changes like that which CHERP81 suggests in relation to changes to the overall NHS budget, time-series analysis and cluster analysis would allow the estimation of PCT behaviour which is consistent with, and testable against, observed PCT behaviour.

A model of PCT behaviour which captures the key effects of silo budgeting, historical bias and structural differences between PCTs is much more difficult to implement than one which assumes: equiproportionality of mortality reduction and quality of life gain; symmetry of response to increases and decreases in expenditure; and cross sectional analysis predicting longitudinal pathways. Correctly implemented, however, they open up the possibility of making more defensible assumptions about unobserved variables. This is because we would have some sense of how those variables relate to available measures in the minds of PCT managers, and we could use partial time-series data to arrive at a fleshed-out and testable model of PCT behaviour.

### **Implications for policy**

We believe that the uncertainty of the estimate provided by CHERP81 is sufficiently large so as to render it a questionable guide to changes in policy until more data are collected. Such additional data would hopefully serve to narrow the range of values for the key elements of the threshold prediction and to justify or refute certain key assumptions CHERP81 has relied upon. The difficulties evident in CHERP81 underline the importance of attempts to directly identify marginal services at a PCT level.

In the interim, we believe that an overall downward bias has been introduced into the CHERP81 “best estimate” of the marginal cost of a QALY in 2008/9 by a number of the assumptions made in its estimation. This could be addressed using existing data in most cases by changing assumptions to give a better sense of the degree of uncertainty of the current estimate. Such an exercise would highlight the importance of collecting additional data to address the degree of uncertainty before an estimate derived using this methodological approach could be confidently used in policy making. Decision makers will be able to have more confidence in the applicability of a revised estimate to policy making. For these reasons, we do not think use of the figure of £12,936 per QALY in NICE decision making is currently justified. A revised estimate may be higher or lower.

We suspect, on the basis of the analysis we present here, that it will be substantially higher but we may be wrong.

## REFERENCES

- Appleby, J., Devlin, N., Parkin, D., Buxton, M. and Chalkidou, K. (2009) Searching for cost effectiveness thresholds in the NHS. *Health Policy*. 91(3), 239-245.
- Appleby, J., Harrison, T., Foot, C., Smith, A., Gilmour, S. (2011) *Explaining variations in primary care trusts' spending on cancer services*. London: The King's Fund.
- Baumol, W. (1995) *Health care as a handicraft industry*. London: Office of Health Economics.
- Castelli, A., Laudicella, M., Street, A. and Ward, P. (2011) Getting out what we put in: Productivity of the English National Health Service. *Health Economics, Policy and Law*. 6(3), 313-335.
- Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P.C. and Sculpher, M. (2013) *Methods for the estimation of the NICE cost effectiveness threshold*. CHE Research Paper 81. Revised following referees' comments. York: Centre for Health Economics, University of York.
- DH (Department of Health). (2011) *A new value-based approach to the pricing of branded medicines – Government response to consultation*. London: Department of Health.
- DH (Department of Health). (2012) *Overview of the programme budgeting costing methodology*. London: Department of Health. Available at: <http://www.dh.gov.uk/health/2012/08/programme-budgeting-methodology/>
- DH (Department of Health). (2013) *The mandate: A mandate from the government to NHS England: April 2014 to March 2015*. London: Department of Health.
- Hansen, P. (2012) Health sector decision-making: More than just cost per QALY calculations. *Journal of Health Services Research and Policy*. 17(3), 129-130.
- HSC (Health and Sport Committee), Scottish Parliament. (2012) *Health and Sport Committee 9th Report, 2012 (Session 4): NHS Boards Budget Scrutiny*. SR paper 171. Available at: [http://www.scottish.parliament.uk/S4\\_HealthandSportCommittee/Reports/heR-12-09w.pdf](http://www.scottish.parliament.uk/S4_HealthandSportCommittee/Reports/heR-12-09w.pdf) [Accessed 25 November 2013].
- HM Treasury. (2013a) *GDP deflators at market prices, and money GDP: September 2013*. Available at: <https://www.gov.uk/government/publications/gdp-deflators-at-market-prices-and-money-gdp-march-2013>. [Accessed 13 November 2013].
- HM Treasury. (2013b) *Statistical bulletin: Public spending statistics February 2013*. London: Her Majesty's Treasury.
- Jones, N.M. and Charlesworth, A. (2013) *The anatomy of health spending 2011/12*. Research report. London: Nuffield Trust.
- 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(4), 826-842.
- Martin, S., Rice, N. and Smith, P. (2008b) *Further evidence on the link between healthcare spending and health outcomes in England*. London: The Health Foundation.
- Martin, S., Rice, N. and Smith, P. (2008c) *The link between health care spending and health outcomes for the new English Primary Care Trusts*. CHE Research Paper 42. York: Centre for Health Economics, University of York.
- Massey, F. (2012) *Public service productivity estimates: Healthcare, 2010*. 7 December 2012. Newport: ONS.
- NAO (National Audit Office). (2008) *Good governance report: Review of programme budgeting*. London: NAO.

NICE (National Institute for Health and Care Excellence). (2008a) *Guide to the methods of technology appraisal*. London: NICE.

NICE (National Institute for Health and Care Excellence). (2008b) *Social value judgements: Principles for the development of NICE guidance*. London: NICE.

ONS (Office for National Statistics). (2010) *Public service output, inputs and productivity: Healthcare*. UK Centre for the Measurement of Government Activity. Newport: ONS.

Shah, K., Praet, C., Devlin, N., Sussex, J., Appleby, J. and Parkin, D. (2012) Is the aim of the English health care system to maximize QALYs? *Journal of Health Services Research and Policy*. 17(3), 157-163.