**How effective is marginal healthcare expenditure?**

**New evidence from England for 2003/04 to 2012/13.**

20 May 2021

Authors

Stephen Martina, James Lomasb, Karl Claxtona, b and Francesco Longob

Affiliations

aDepartment of Economics

 University of York

 York

 YO10 5DD

UK

bCentre for Health Economics

 University of York

 York

 YO10 5DD

 UK

Email addresses

Stephen Martin: sdm1@york.ac.uk

James Lomas: james.lomas@york.ac.uk

Karl Claxton: karl.claxton@york.ac.uk

Francesco Longo: francesco.longo@york.ac.uk

Corresponding author

Stephen Martin: sdm1@york.ac.uk

**Abstract**

*Background*

The endogenous nature of healthcare expenditure means that instruments are often used when estimating the relationship between expenditure and mortality. Previous English studies of this relationship have largely relied on statistical tests to justify their instruments. A recent paper proposed that exogenous components of the resource allocation formula, used to distribute the national healthcare budget to local health authorities, be used as instruments.

*Objectives*

To estimate the relationship between healthcare expenditure and mortality by disease area for England from 2003/4 to 2012/13 using exogenous elements from the resource allocation formula as instruments for expenditure. To use these disease-specific estimates to estimate the marginal cost per quality-adjusted life year (QALY) for English NHS expenditure. To compare these estimates with those that relied on statistical tests to justify their instruments.

*Methods*

The two-stage least squares estimator is used to estimate the annual relationship between mortality and healthcare expenditure by disease area across 151 local authorities. These disease-specific outcome elasticities are combined with information about survival and morbidity disease burden in different disease areas to calculate the marginal cost per QALY for English National Health Service (NHS) expenditure.

*Results*

The results suggest an annual marginal cost per QALY of between £5,000 and £10,000. This is similar to that reported previously by studies that used statistical tests to justify their instruments.

*Conclusion*

These cost per QALY estimates are much lower than the threshold currently used by the UK’s National Institute for Health and Care Excellence (NICE) (£20,000 to £30,000) to assess whether a new pharmaceutical product should be funded by the NHS. Our estimates suggest that guidance issued by NICE is likely to do more harm than good, reducing health outcomes overall for the NHS. There may be legitimate reasons why such harms are deemed appropriate, but it is only through the type of empirical analysis in this paper that the reasons for these ‘harms’ are likely to be articulated and explicitly justified.

**Key points for decision makers**

* To generate an additional quality-adjusted life year requires between £5,000 and £10,000 of NHS expenditure
* NICE recommends that a new pharmaceutical product should be funded by the NHS if it costs less than £20,000 to £30,000 to generate a quality-adjusted life year
* The NICE threshold is too high and is likely to do more harm than good, reducing health outcomes overall for the NHS.

**Declarations**

Funding

This work is independent research funded by the UK’s National Institute for Health Research Policy Research Programme (NIHR PRP) through its Policy Research Unit in Economic Evaluation of Health & Care Interventions (EEPRU, grant reference 104/0001).

Conflicts of interest

This work is independent research funded by the UK’s National Institute for Health Research Policy Research Programme (NIHR PRP) through its Policy Research Unit in Economic Evaluation of Health & Care Interventions (EEPRU, grant reference 104/0001).  The views expressed in this publication are those of the authors and not necessarily those of the NHS, NHS England, the National Institute for Health Research or the Department of Health and Social Care*.* Although the results have been presented to NHS England and the Department and members have commented on the research, they had no involvement with: the study design; the collection, analysis and interpretation of the data; the writing of the paper; and the decision to submit the article for publication.

Conflict of interest statement

Stephen Martin, James Lomas, Karl Claxton and Francesco Longo declare that they have no conflict of interest.

Data availability

All of the raw data are in the public domain. The healthcare expenditure, population size and ‘distance from target’ data are available from the programme budgeting data set. This is available for 2012/13 from <https://www.networks.nhs.uk/nhs-networks/health-investment-network/news/2012-13-programme-budgeting-data-is-now-available> [accessed 14 July, 2020] and previous years’ data are available via links on the above webpage. The socio-economic variables have been constructed from the 2001 and 2011 Population Censuses. The 2011 values are available from the Office for National Statistics at <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuskeystatisticsforlocalauthoritiesinenglandandwales> [accessed 14 July, 2020] and the 2001 values are available from <https://www.nomisweb.co.uk/home/census2001.asp> [accessed 14 July, 2020]. Current mortality data are available from the NHS Digital website at <https://digital.nhs.uk/data-and-information/publications/statistical/compendium-mortality/current/years-of-life-lost> and historical data are available on request from the Clinical Indicators team at NHS Digital (email address: clinical.indicators@nhs.net). The 2009/10 MFF and prescribing cost age indices are available from the Department of Health’s resource allocation book for 2009/10 at [http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/Allocations/DH\_091850](http://webarchive.nationalarchives.gov.uk/%2B/www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/Allocations/DH_091850) (accessed June 23, 2020). The 2011/12 MFF and prescribing cost age indices are available from the Department of Health’s resource allocation book for 2011/12 at <https://www.gov.uk/government/publications/exposition-book-2011-2012> [accessed 14 July, 2020]. The IMD 2007 is available from https://webarchive.nationalarchives.gov.uk/20100411141238/http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/ [accessed 14 July, 2020] and IMD 2010 is available from <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010> [accessed 14 July, 2020].

Code availability

Our estimation code is not available but estimation was undertaken using the *ivreg2* command within the commercially available software package *Stata 15*.

Ethics approval

Ethical approval is not required.

Author contributions

All authors (SM, JL, KC, FL) contributed to the concept and design of this paper. SM led on the analysis and drafting, and the final paper was edited and approved by all four authors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting these criteria have been omitted.

**How effective is marginal health care expenditure?**

**New evidence from England for 2003/04 to 2012/13[[1]](#footnote-1)**

**1 Introduction**

Early attempts to quantify the relationship between health outcomes and expenditure found rather mixed results, with some studies finding the anticipated positive effect while others struggled to identify any association [1]. Although vaguely based on the idea of a health production function, most early empirical studies failed to account for the endogeneity of expenditure (most likely due to reverse causality) and often used international datasets to undertake cross-country studies [2].

More recent research has focused on the analysis of national data (for Australia, the Netherlands, Spain and Sweden) and these studies have usually attempted to control for the endogeneity of expenditure [3-6]. Interest in this topic has been stimulated by its increasing policy relevance. The results of such studies can often be used to calculate the marginal productivity (opportunity cost) of healthcare expenditure, and such estimates can inform decisions by regulatory agencies about the maximum that the health care system can afford to pay for the benefits offered by new pharmaceutical products.

A recent study estimated the relationship between health outcomes and healthcare expenditure for 23 programme budget categories (PBCs) (disease areas) across 152 English health authorities [7]. This used disease-specific expenditure to estimate outcome and expenditure equations for each of the nine disease areas with a mortality rate, and an expenditure equation for each of the other 14 programmes. The expenditure equations show how a change in the total budget is split across different care programmes, and the outcome equations reveal how this change in programme expenditure affects programme health. The analysis used socio-economic variables as instruments for the potentially endogenous expenditure variable and the authors largely relied on statistical tests to defend their choice of instruments. With the aid of two assumptions – that the impact of a change in expenditure on the QALY burden of disease is the same as on the mortality burden, and that the average impact on mortality in those areas with a mortality-based measure can be used as a proxy for the average impact in those areas without a mortality-based measure – the study reported an annual marginal cost per QALY of between £5,400 and £14,400 for all English National Health Service (NHS) expenditure [7].

Since the completion of that study, a new approach to the instrumentation of healthcare expenditure has been proposed [8]. This ‘funding rule’ approach notes that the per capita budget allocated to each health authority is the product of the national per capita budget and various adjustments reflecting local circumstances. Where the variables used to make these adjustments have no direct link with mortality, the proposal is that they be used as instruments for expenditure [8].

Although the funding rule (FR) approach to instrumentation can be applied to disease-specific mortality, the initial FR study only reported a single all-cause outcome elasticity (=-0.706 for 2005/06) [8]. This is unfortunate because it is not possible to convert an all-cause outcome elasticity into a cost per quality-adjusted life-year (QALY) without assuming that, for a small change in NHS expenditure, the total health (QALY) effect estimated using all-cause mortality is the same as the sum of the health (QALY) effects across all disease-specific areas when separate outcome equations are estimated for each of these disease areas. Our results suggest that this assumption is not met (the disease-specific results generate more QALYs in total than the all-cause result) and so we believe that disease-specific results provide a more secure basis for cost per QALY estimates.

Following the initial FR study, a small pilot study obtained plausible results using the FR approach for four disease areas using English healthcare data for the same single year (for 2005/06), but it presented no cost per QALY estimates [9]. A plausible cost per quality-adjusted life-year (QALY) estimate for all NHS healthcare expenditure requires disease-specific elasticities for as many conditions as mortality data permit.

In addition to better instrumentation, the FR approach has several other advantages over the use of socio-economic variables as instruments [9]. For example, by using total rather than disease-specific expenditure, the FR approach avoids the need to estimate any expenditure equations and hence it does not rely on the availability of disease-specific expenditure nor on health authorities providing an accurate split of total expenditure across different care programmes. Moreover, the use of total rather than disease-specific expenditure means that results will reflect the health benefits in each disease area generated by expenditure in all disease areas, and these may be important for patients with co-morbidities (eg an effect on deaths recorded as cardiovascular disease may reflect expenditure in non-cardiovascular areas such as diabetes).

Given the improvements offered by the FR approach, it is important to examine whether the cost per QALY estimates implied by the new approach confirm or reject the earlier results obtained using the less acceptable socio-economic instruments. To facilitate this comparison, we use the FR approach to estimate outcome equations for all-cause mortality and for nine disease areas, annually, over the ten-year period from 2003/04 to 2012/13. We use the resulting elasticities, together with information about survival and disease burden in all 23 care programmes, to calculate the implied marginal cost per QALY and the health opportunity costs associated with £10m of NHS expenditure. Ten annual estimates have been calculated to check the plausibility of any one year’s estimates and to examine time trends in the estimates. We also compare these FR-based results with those reported previously that used socio-economic variables as instruments [7].

**2 Methods**

***2.1 Institutional context***

The English NHS is a centrally planned and publicly funded health care system. Its income comes almost entirely from national taxation. Access to the Service is usually via general practitioners who act as gatekeepers to secondary care and pharmaceuticals. With some minor exceptions, the service is free at the point of consumption for patients. The Service is organized geographically, with responsibility for the local management of the NHS delegated to local health authorities. In the ten years to April 2013, these authorities were known as Primary Care Trusts (PCTs). Using various resource allocation formulae developed by the Department of Health (one for hospital services, one for pharmaceuticals and one for primary care), PCTs were assigned a fixed annual budget by the national ministry within which they were supposed to meet expenditure on almost all types of health care. The guiding principle behind the allocation formulae was that the need for medical care in a given population could be systematically calculated, allowing resources to be allocated in a fair and transparent manner, so that areas in equal need received the same budget. Well over two-thirds of all funding was for hospital and community health services (HCHS) with most of the remainder split between prescribing and primary care.

***2.2 Theoretical framework***

Most health outcome and expenditure studies estimate an equation of the form

 $y\_{i}= α+ βx$i$ + γN\_{i}+ ε\_{i}$(1)

where $y$ is some measure of health outcome (we use the under 75 age/sex standardised years of life lost rate) , $x$ is a measure of healthcare expenditure (we use expenditure per resident), N is the need for healthcare and $ε$ reflects everything not included elsewhere in the specification. Interest focuses on the size of the estimate of $β$as this tells us how responsive outcomes are to a small change in the level of expenditure. Healthcare expenditure is likely to be endogenous due to the presence of reverse causality, that is, it is likely to be affected by (historic) mortality. This reverse causation implies that expenditure is likely to be correlated with the error term in equation (1), that the ordinary least squares (OLS) estimator will be both biased and inconsistent, and that the OLS coefficient on expenditure will under-estimate the size of the negative effect on mortality.

To obtain consistent estimates, researchers often use instrumental variable analysis (also known as two-stage least squares (2SLS)). This involves finding one or more exogenous variables (‘instruments’) that are associated with the exogenous part of the variation in the endogenous regressor (the level of healthcare expenditure) but which are otherwise unrelated to the dependent variable (the mortality rate). If Z is the selected instrument then estimation proceeds in two steps. The following first-stage equation is estimated using OLS:

 $x\_{i}= δ+θZ\_{i}+ρN\_{i}+φ\_{i}$ (2)

The instrument (Z) and the controls from the equation of interest (N) are used to predict the ‘good’ (or exogenous) variation in the endogenous regressor (expenditure), that is, that part that is not correlated with the error term in equation (1). At the second stage, this predicted level of expenditure ($\hat{x}$i) replaces the actual level of expenditure ($x$i) when using OLS to estimate equation (1).

For 2SLS to generate consistent estimates certain assumptions have to be met. First, the instrumental variable (Z) should be a ‘strong’ (good) predictor of the endogenous variable; ‘weak’ instruments can generate estimates with wide confidence intervals and cause the 2SLS estimator to be badly biased towards the OLS estimate. The usual rule of thumb ‘test’ for strong instruments is that the F statistic associated with the instrument(s) in the first-stage regression should be about 10 or better, and hence we report the Kleibergen-Paap F test statistic for all first-stage estimations in this study [10].

The second assumption is that the only reason for the relationship between the instrument and the outcome is the first-stage so that $Cov\left(Z\_{i}, ε\_{i}\right)=0$. Although this is called *the* exclusion restriction assumption, this assumption has two parts. The first part is that the instruments are as good as randomly assigned so that, conditional on the controls (N), they are independent of the outcome [11]. This requires that there are no unmeasured/unobserved confounders that affect both the instrument and the outcome variable. The second part of the exclusion restriction assumption is that, conditional on the controls (N), the instrument has no effect on the outcome other than through the first-stage channel [11].

The exclusion restriction assumption cannot be tested if the number of instruments is the same as the number of endogenous regressors. However, if there are more instruments than endogenous variables, we can test whether the additional instruments are valid. This test compares the 2SLS estimator with all instruments to 2SLS using a subset that just identifies equation (1). If all instruments are valid, the estimates should differ only as a result of sampling error [12]. If we fail to reject the null hypothesis that all instruments are valid, we can have some confidence in the set of instruments used. However, there remains the possibility that more than one instrument is endogenous, and that the 2SLS estimators using a full and reduced set of instruments are biased in the same direction and with the same degree of bias [12]. Where there are two or more instruments, we report the Hansen-Sargan test of instrument validity for every estimated second-stage equation.

***2.3 The funding rule approach to instrumentation***

Initial English studies used socio-economic variables as instruments for healthcare expenditure [13-15]. More recently, an alternative approach has been proposed [8]. This notes that between 2003/04 and 2012/13 the share of the national budget apportioned to each PCT was governed by the Department of Health’s allocation formula or ‘funding rule’. Strictly speaking, there were different formulae for different types of healthcare activity but the basic structure of each formula was the same and hence the discussion below talks of a single formula or funding rule.

The Department’s funding rule reflected each PCT’s need for expenditure and this, in turn, reflected the PCT’s population size, its age profile, local input prices, and other ‘need for health care’ factors. Periodically, the national ministry revised its funding rule and this, together with data updates, generated a new target allocation for each PCT. The new funding rule might generate a large change in the target allocation for some PCTs and, to avoid sudden large reductions in actual budgets, such changes were usually incorporated into annual budgets over a number of years [16]. Thus each PCT’s budget could be specified as:

Local budget per person= (National budget per person) x (Local age-cost index) x

 (Local additional needs index) x (Local input price index) x (Local DFT Index) (3)

where:

(a) the age-cost index reflects the pure age/gender impact of the local population’s demographic profile on healthcare costs;

(b) the additional needs index reflects local deprivation and other factors, such as the recent mortality rate, likely to influence the need for health care;

(c) the input price index (the market forces factor (MFF)) reflects prices in the local health economy; and

(d) the distance from target (DFT) index reflects how far each PCT’s actual budget allocation is from its target allocation.

It is argued that three of the four adjustment factors in equation (3) can be used as instruments because these factors will be correlated with expenditure but are unlikely to be correlated with mortality, particularly with the inclusion of appropriate controls in equation (1) [8]. The additional needs index is not an appropriate instrument because, as constructed, it contains a measure of historical mortality which would introduce an element of reverse causality.

We saw in section 2.2 that valid instruments should be both exogenous (uncorrelated with unobservable determinants of mortality) and excludable (from the second-stage). Let us consider whether the proposed instruments meet these requirements. The MFF (input price index) adjustment reflects prices in the local health economy. It is designed to compensate health authorities for the unavoidable higher costs they incur when purchasing labour and other goods/services. If the MFF adjustment is perfect for every health authority then all

PCTs will be able to buy the same bundle of inputs and, although a useful predictor of nominal expenditure, this instrument can, in principle, have no impact on mortality because it has no impact on real expenditure. However, in practice, the MFF adjustment will not be perfect and any errors in the adjustment will lead to differences in the volume of real resources available to health authorities (we assume that this error is small relative to the adjustment for local prices). We have no reason to believe that these errors in the MFF adjustment will have any effect on mortality other than through their effect on expenditure (excludability). The MFF index reflects characteristics of the local (health) economy that could potentially be correlated with unmeasured determinants of mortality and its exogeneity is therefore conditional on health needs and socio-demographic characteristics employed in the estimated specification [17].

Similar arguments can be made for the age-cost index. This is designed to compensate health authorities for the unavoidable additional expenditure they incur due to the demographic profile of their population. If the age-cost adjustment is perfect for every health authority then all PCTs will be able to offer the same level of care irrespective of whether their population is a particularly old or young one. Again, this (age-cost) index will be a useful predictor of nominal expenditure but, if the adjustment is perfect, this instrument can have no impact on mortality because it has no impact on real expenditure. However, the age-cost adjustment will not be perfect and any errors in the adjustment will lead to differences in the volume of real resources available to health authorities. We have no reason to believe that these errors will have any effect on mortality other than through their effect on expenditure (excludability). The age-cost index reflects the impact of the local population’s demographic profile on healthcare costs. As is the case for the MFF, this profile could potentially be correlated with unmeasured determinants of mortality and its exogeneity is therefore conditional on the control variables employed in the estimated specification.

The DFT index will reflect the various funding formulae and ‘pace of change’ policies implemented under several governments of various political persuasions over the past thirty years.  While there are undeniably policy choices involved, such as the setting of the ‘pace of change’ (POC) adjustment that transitions PCTs towards their target, over the study period the POC policy focussed on providing a minimum basic budget uplift for all PCTs with a larger increase for those that were most under-target. We have no evidence to suggest that these policy choices were made on the basis of other factors such as outcomes (excludability) [18]. Moreover, health authority allocations usually include a relatively small component that seeks to address health inequalities directly and it is at this point that outcomes are considered rather than at the POC policy stage [19]. We also have no evidence to suggest that, conditional on our controls, the DFT index will be correlated with unmeasured/unobserved determinants of mortality (exogeneity).

Finally, of course, we test the empirical validity of our instruments using the appropriate statistical test (the Hansen-Sargan test).

***2.4 Estimation strategy***

To facilitate comparison with other studies, we too start with an analysis of expenditure data for 2005/06. We use three funding rule variables (an age index, an MFF index and a DFT index) from the resource allocation formulae used to distribute budgets to PCTs as instruments for expenditure. When estimating equation (1) theory suggests that it is important to control for healthcare need because need will be strongly correlated with mortality and expenditure. However, theory provides little guidance as to the precise appropriate controls. Hence following other studies and to make our results comparable with theirs, we identify a dozen socio-economic variables – such as the proportion of owner-occupied households – as potential controls for the need for health care and we use backward and forward selection techniques to reduce multicollinearity problems [7, 9, 13-15].

For each disease area, we start by estimating (1) with all socio-economic variables included as controls. This ‘full’ specification is merely a starting point (based on an evolved body of work in this area) from which we seek to identify a plausible specification that passes the usual statistical tests.The least significant regressor is removed from the specification and the equation is re-estimated. This process – of dropping the least significant regressor and re-estimating -- continues until there are only significant controls remaining (the expenditure term is forced to be ever-present). This specification becomes our preferred result if the estimated coefficients correspond with our theoretical priors (eg the coefficient on expenditure is not positive) and it passes the appropriate statistical tests (eg the instruments are valid and strong). However, if this is not the case, the specification is adjusted (eg an invalid instrument is removed) and the equation re-estimated. When the specification requires no further adjustment it becomes our preferred specification for this estimation year (2005/06).

Rather than re-derive the precise empirical specification for each programme for the following year (i.e., for 2006/7), we update the dataset and use the preferred specification for 2005/06 to re-estimate each outcome equation for 2006/07. We use this as our preferred result for 2006/07 if this re-estimation produces a result that passes the appropriate statistical tests and generates coefficients in line with our theoretical priors. If the re-estimation fails to do this, we re-estimate the equation having adjusted the specification as suggested by the initial result. This process – of using the empirical specification for the previous year and applying it to updated data – is applied to each year’s dataset both going forward from 2005/06 to 2012/13, and going back from 2005/06 to 2003/04. In addition to nine disease-specific outcome equations, we also use this estimation strategy to obtain an all-cause specification for each year.

Ideally, the same empirical specification would be relevant for a particular disease area for each data year but sometimes the specification requires adjustment (table A1 in appendix 0 provides a qualitative indicator of the degree to which the preferred specification for the current year has been changed from that for the previous year). There may be good reasons why adjustment is required. For example, the general population and the population at risk in each disease area will be changing every year and this may affect the relationship between instruments, expenditure, controls and mortality. Moreover, we only have observations for the socio-economic controls for two years (for 2001 and 2011) and values for the intervening years are estimated using interpolation. A changed specification might reflect inaccuracies associated with interpolation rather than any change in the underlying data generating process.

When estimating regressions, the values for all variables are logged so that regression coefficients can be interpreted as elasticities. All observations are weighted by the size of the area’s population and robust standard errors are reported. Estimation is undertaken using the *Stata* statistical packageand the *ivreg2* command [20]. In addition to the weak instrument and valid instrument tests mentioned above, we also report a test for whether expenditure is endogenous and a reset test (Pesaran-Taylor) for mis-specification [20, 21].

***2.5 Using the outcome elasticities to estimate the cost per QALY***

The outcome elasticities obtained by estimating equation (1) for each disease area with a mortality indicator indicate the proportionate effect on mortality in that particular disease area in one year, due to a proportionate change in total expenditure. These elasticities are used as a surrogate for the likely effect of a change in expenditure on a more complete measure of the health, which combines survival and morbidity effects (i.e., changes in the QALY burden of disease). In those PBCs without significant mortality (see table A1 in appendix 1), where mortality elasticities cannot be estimated, the proportional effect of a change in expenditure on the QALY burden of disease is assumed to be the same as the overall proportional effect on the burden of disease across those PBCs where the mortality effect can be estimated (this is the extrapolation assumption). The estimated outcome elasticities, together with these surrogacy and extrapolation assumptions, are used to calculate the change in the total QALY burden across all PBCs associated with a small change in health care expenditure, and hence the marginal cost per QALY. We also calculate the number of QALYs gained through a small (£10m) increase in the total NHS budget.

The assumptions of surrogacy and extrapolation have been found to be conservative with respect to the likely health effects of changes in expenditure [22, 23]. Both, however, require estimates of the total QALY burden of disease for the population with disease (in a PBC) in a particular expenditure year. Each PBC (with the exceptions of PBCs 16, 22 and 23) is a subset of the approximately 1500 non-procedural, 3-digit ICD 10 codes. Therefore, the QALY burden of disease is first estimated at ICD code level before being aggregated to each PBC (see appendix 1 for additional stepwise explanation, sources and specific links to the previous publication of these methods and data).

Office for National Statistics (ONS) data, which report the age of death, for all deaths, by gender for all ICD codes each year, combined with evidence about the age and gender distribution of the at-risk population in each ICD code, is compared to the life expectancy of the matched general population, to provide a measure of the survival burden of disease (YLL). The impact of disease on quality of life for those surviving with disease must first account for evidence of quality of life ‘norms’ for the general population by age and gender of the at-risk population in each ICD code. Quality of life decrements due to disease (using pooled data from the Health Outcome Data Repository (HoDAR) and the Medical Expenditure Panel Survey (MEPS)) are then applied to these age and gender ‘norms’.

Therefore, the total QALY burden of disease (for each ICD code aggregated to PBC) in one year is the sum of the YLL adjusted for quality of life norms and the quality of life impact of disease on average for the incident and prevalent population in one year. These are updated each year using the routinely reported ONS data, which provides age of death for all deaths, by age of death and gender for all ICD codes, and life expectancy for the general population matched to the age and gender of the at-risk population in each ICD code (using ONS life tables). Estimates of the YLL and total annual QALY burden by PBC are reported in table A1 in appendix 1 for each year of expenditure data.

**3 Data**

***3.1 Expenditure data***

NHS expenditure data were obtained from the NHS programme budgeting website (see <https://www.networks.nhs.uk/nhs-networks/health-investment-network/news/2012-13-programme-budgeting-data-is-now-available>). This dataset includes all expenditure by PCTs including that on inpatient and outpatient care, pharmaceuticals and primary care. We want to estimate outcome equations for as many years as possible, from the first availability of the expenditure data (for 2003/04) to the abolition of Primary Care Trusts in March 2013. The only geography for which mortality data is available over this entire period is by local authority and hence we use local authorities, rather than PCTs, as our unit of analysis. PCT expenditure data are converted to a LA geography using information from a mapper that reflects population sizes as at mid-2010. These time-fixed weights are used to map expenditure from PCTs to LAs according to the proportion of the PCT population in each LA.As Table 1 reveals, the average level of expenditure per person by LA for 2005/6 is £1,340, but this varies across LAs from just over £1,000 to just under £2,000 per person. Further details about the expenditure data are in appendices 2 and 3.

***3.2 Mortality data***

To study of the impact of expenditure in 2005/6 on health outcomes, we use the under 75 years (age standardised) years of life lost rate (SYLLR)) calculated by pooling deaths data over the three-year period 2005, 2006 and 2007 as our measure of mortality [24]. The mortality data was also pooled over a three-year period when calculating SYLLR for the other study years. This mortality data was obtained via the NHS Digital Clinical Indicators portal (see <https://digital.nhs.uk/data-and-information/publications/statistical/compendium-mortality/current/years-of-life-lost>). Descriptive statistics for the various SYLLRs for 2005/2006/2007 are in Table 1. For example, the all-cause SYLLR averaged 471 years of life lost per 10,000 head of population. However, this rate varied considerably across LAs, ranging from 257 to 736 years of life lost per 10,000 head of population. Equally large geographical variations in mortality rates are evident for some disease-specific areas. Further details about the mortality data are in appendices 2 and 3.

***3.3 Instruments***

During the study period the Department of Health employed a different resource allocation formula for each of the main types of healthcare (HCHS, primary care, pharmaceuticals) funded by the NHS. Each formula has its own age, needs, and MFF indices. These formulae were only revised intermittently so that time-varying MFF and age indices are not available for every study year (and, anyway, these will only change very slowly). We make the best use of what is available.

The Department of Health’s resource allocation books for 2003/4, 2004/5 and 2005/6 can be used to extract the relevant age and MFF indices for Hospital and Community Health Services (HCHS). However, these allocations are for the 303 PCTs in existence in April 2003. In October 2006 there was a major re-organisation of PCTs and their number was reduced from 303 to 152. As we do not have an appropriate mapper to convert the age and MFF indices for the 303 PCTs to 152, we consulted the first post-merger resource allocation book for proxies for these variables. The resource allocation book for 2009/10 contains an MFF for HCHS and we use this as our first instrument for the initial study year (2005/06) having mapped it to our local authority geography.

The HCHS resource allocation formula for 2009/10 also contains an age index but this is combined with an additional need index and the latter includes an element of historical mortality. Hence as our second instrument we use the age index from the prescribing cost resource allocation formula for 2009/10 which is a pure age index and uncontaminated with any measure of mortality. Annual distance from target (DFT) data for each PCT’s total allocation across all funding streams is available for all study years and this is our third instrument.

Table 1 shows that all three of our instruments average about one but that there is considerable variation around the country. For example, the MFF index suggests that unit costs in some LAs are between 7% lower and 15% higher than the average. Further details about the instruments can be found in appendices 2 and 3.

***3.4 Socio-economic variables***

The dozen potential socio-economic controls for the need for health care are in Table 1. These census-based variables are constructed using the 2001 census and then again using the 2011 census. Values for specific analysis years (e.g., for 2005) are constructed by linear interpolation between the values for the two censuses. For the final study year (2012) we use the same values as for the penultimate study year (2011). The descriptive statistics in Table 1 relate to the 2005 study year. They reveal that, on average, 10% of all residents are born outside the European Union, 29% of the working-age population are employed in managerial and professional occupations, and 65% of households are owner occupied. Again, these averages mask considerable variation across local authorities: the proportion of residents born outside the EU varies from less than 1% to more than 42%, and the extent of owner occupation ranges from 28% to 83% of all households. Statistics for the same variables for the other study years are in appendix 3.

**4. Results**

***4.1 Outcome elasticities***

The estimation strategy outlined above generates a large amount of statistical output when applied to ten disease areas, annually, over the study period. Here we provide detailed results for one year (for 2005/06) with summaries for the others, but detailed results for each of the other study years are in appendix 4.

*All-cause mortality for 2005/06*

The results in Table 2 illustrate the application of our estimation strategy to the estimation of equation (1) for all-cause mortality. This reports the second-stage regression for various specifications as suggested by our estimation strategy (first-stage results are in Table A10 in appendix 4). We begin with the inclusion of all potential socio-economic controls in the equation of interest (see column 1 of Table 2) and all three funding rule variables as instruments (see column 1 of Table A10). As indicated by our estimation strategy, the least significant regressor is removed from the specification and the equation is re-estimated. This process – of dropping the least significant regressor and re-estimating -- continues until there are only significant controls remaining. This process leads to the more parsimonious second-stage specification shown in column 2 of Table 2.

Unfortunately, the instrument set associated with this specification is not valid so we eliminate one instrument at a time and re-estimate the model (three times in all). The result in column 3 shows the result associated with the least invalid instrument set (this includes the age and MFF instruments) but this combination of instruments still just fails the Hansen-Sargan test at the 10% level. The specification in column 4 shows the (strongest instrument) result associated with the use of a single instrument (the MFF index). However, this specification still fails the reset test so we re-estimate this several times and, on each occasion, add back in one of our previously deleted socio-economic control variables. The results shown in columns 5 and 6 both pass the reset test and, as the final result has the strongest instrument set (the Kleibergen-Paap F statistic=26.7), this is our preferred outcome specification for all-cause mortality.

*Disease-specific mortality for 2005/06*

If we apply the same strategy to the estimation of the health outcome equation (1) for nine disease-specific mortality rates we obtain the second-stage results shown in columns 2 – 9 of Table 3 (first-stage results are in Table A11 in appendix 4). There is no result for epilepsy mortality because we were unable to find a plausible specification for this disease area for this year. The coefficient on expenditure is significant at the 10% level for all eight disease-specific areas. The endogeneity test statistics suggest that expenditure is clearly endogenous for all eight specialties. The Hansen-Sargan test statistics suggest that the instrument sets are valid. The Kleibergen-Paap F statistics suggest that, for six of the eight specifications, we have strong instruments (F statistic around ten or better) and that for the other two (for gastro-intestinal problems and for neonate mortality) the instrument sets are borderline strong (F-statistics of 8.74 and 8.95 respectively). The Pesaran-Taylor reset test statistics suggest no evidence of mis-specification. The coefficients on the instruments in the first-stage regressions (Table A11 in appendix 4) are usually significant and positive (as is anticipated).

*All-cause and disease-specific mortality for 2003/04, 2004/05, 2006/07 – 2012/13*

The estimation strategy outlined in section 2.4 is applied to each year’s dataset, both going forward from 2005/06 to 2012/13, and going back to 2003/04. The values of all variables are logged before estimation so that the regression coefficients can be interpreted as elasticities; they reflect the percentage change in mortality associated with a 1% change in NHS expenditure. The outcome elasticities for all-cause and disease-specific mortality are in Table 4. Confidence intervals associated with these point estimates are in Table A30 in appendix 5. In addition to the directly estimated outcome elasticities, Table 4 also reports the all-cause elasticity implied by the disease-specific elasticities.

***4.2 Implied cost per QALY and health opportunity costs***

The outcome elasticities reported in Table 4, together with information about the QALY burden of disease and the surrogacy and extrapolation assumptions outlined in section 2.5, enable us to estimate the marginal cost per QALY for total NHS expenditure. These cost per QALY estimates are shown in Table 5 together with the health opportunity costs (in terms of the number of QALYs forgone or gained) associated with a small (£10m) change in health care expenditure. The confidence intervals associated with the mortality elasticities were used to derive the confidence intervals for the cost per QALY and health opportunity cost estimates.

The cost per QALY point estimates and their associated confidence intervals in Table 5 are illustrated diagrammatically, using a solid line, in Figure 1. These estimates can also be presented as the number of QALYs gained or forgone associated with a small change in the level of NHS expenditure (say, £10m). These health opportunity cost estimates are also presented in Table 5 and are illustrated diagrammatically, using a solid line, in Figure 2.

One advantage of the diagrammatic form is that it permits a ready comparison with the cost per QALY and health opportunity cost estimates obtained using socio-economic instruments [7]. These estimates and their associated confidence intervals are reproduced diagrammatically using a dashed line in Figures 1 and 2.

All of the estimates in Table 5 and Figures 1 and 2 are at current prices. Outcome elasticities estimated over the pooled sample using constant (2012/13) price expenditure data are in appendix 6. This appendix also reports the annual cost per QALY at constant (2012/13) prices using (a) the cross-section results and (b) the pooled analysis.

**5. Discussion**

***5.1 Outcome Elasticities***

The disease-specific elasticities shown in Table 4 reveal that, although there is some across year variation, the elasticities for most disease areas are reasonably constant throughout the study period. This consistency is most notable for those disease areas with a relatively large number of deaths (such as cancer and circulatory disease). The disease-specific elasticities in Table 4 also demonstrate that the responsiveness of mortality to changes in expenditure varies considerably by disease area. For example, for every year circulatory disease appears to be more responsive than cancer to marginal changes in expenditure.

Two all-cause outcome elasticities are reported in Table 4. One of these has been directly estimated in the same way as the disease-specific elasticities, and the other is the elasticity implied by the disease-specific estimates. These confirm the patterns noted for the disease-specific elasticities: first, although there is some variation across years, both series are relatively consistent; and second, as these reflect total mortality, the variation in the estimates across time is smaller than the variation in the disease-specific estimates. It is also worth noting that, for all ten years, the directly estimated all-cause elasticity is smaller (absolutely) than the all-cause elasticity implied by the disease-specific elasticities. There may be some aggregation bias associated with the all-cause elasticity that is avoided through the use of more disaggregated data. A reviewer suggested that this result could also be because the disease specific result is the less robust result due to the cumulative effects of bias across the individual disease specific models. However, it is not obvious why the disease specific results should all be affected by bias in the same direction, nor why this bias should not affect the all-cause result.

Our FR-based elasticities can also be compared with those obtained when analysing a very similar English dataset but using socio-economic instruments selected using statistical tests. We find that our FR-based elasticities are similar to those reported previously using socio-economic instruments [7]. For example, over the ten-year study period we find that the average FR-based implied all-cause elasticity is -1.193 and this is almost identical to that reported using socio-economic instruments (=-1.248) [7]. In other words, the use of two different approaches to the identification of expenditure generates almost identical average (implied) all-cause outcome elasticities.

***5.2 Cost per QALY and health opportunity costs***

All of the cost per QALY estimates in Table 5 lie within a relatively narrow range, that is, between £5,800 and £9,900. The 95% confidence intervals associated with these point estimates overlap each other considerably and, across all years, fall between £4,200 and £22,400. Similarly, the point estimates of the health opportunity cost associated with £10m of NHS expenditure lie within a relatively narrow band for all of the ten years (between 1,000 and 1,800 QALYs), with all 95% confidence intervals ranging between 440 and 2,400 QALYs.

The diagrammatic presentation of our results in Figure 1 (solid line) emphasises the relative consistency of our results through time and that they are similar to those obtained using socio-economic instruments (dashed line) [7]. For eight of the ten years the two sets of cost per QALY estimates are very similar, with the socio-economic instruments generating a slightly larger cost per QALY in both 2008/09 and 2012/13. The two sets of health opportunity cost estimates shown in Figure 2 also demonstrate that two very different approaches to instrumentation generate similar results.

The cost per QALY estimates are in nominal (current price) terms and reveal an average annual growth rate of 5.4%. They can be deflated to constant prices using the average annual growth rate of the hospital and community health services (HCHS) pay and price index (2.8%) over the study period [25]. Having taken account of this, our results indicate that the estimated cost of a QALY gained by marginal NHS expenditure grew in real terms with an average real-terms annual growth rate of 2.6% (see Table A32 in appendix 6). This growth in the cost per QALY implies a fall in real marginal productivity. This may be attributed to diminishing marginal returns to the dramatically increased total NHS expenditure over this period, which followed an average real-terms annual growth rate of 4.2% (the average annual growth rate in current price NHS expenditure was 7.0% (see appendix 3) and the corresponding growth rate in the HCHS pay and price index was 2.8%) [25].

A recent paper used 28 clinical experts and 25 individuals with policy responsibilities to examine the validity of our surrogacy and the extrapolation assumptions [22, 23]. The results from the research suggest that the effect of a change in expenditure on the total QALY burden is, in proportionate terms, expected to be higher than (rather than equal to) those on the mortality burden. In other words, the surrogacy assumption is likely to underestimate the QALY benefits of a change in expenditure. The research also suggests that the extrapolation assumption is likely to underestimate QALY benefits; in particular, the health effects in disease areas without measured mortality effects are, on balance, expected to be greater than the assumed average of those effects where a mortality measure is available. This evidence suggests that the surrogacy and extrapolation assumptions are, if anything, conservative ones.

However, it is straightforward to examine the impact of the assumption that the effect of expenditure on the mortality burden of disease can be used as a surrogate for the effect on the broader QALY burden. If the effect on the QALY burden is twice as big as the effect on mortality, then the cost per QALY would be halved. Alternatively, if the effect on the QALY burden is half the effect on mortality, then the cost per QALY is doubled.

Interest in the cost per QALY for health expenditure is not confined to the UK. A study of

Spanish health care expenditure reports a cost per QALY of between 22,000€ and 25,000€ and anAustralian study suggests a figure of AU$28,000 per QALY gained [3, 5]. At current exchange rates both of these studies report slightly larger cost per QALY figures than we have found for England. Estimates are also available for the Netherlands (41,000€ per QALY) and Sweden (39,000€ per QALY) [4, 6].

***5.3 Policy implications***

An assessment of the marginal productivity of health care expenditure is of considerable policy interest. For example, the cost per QALY estimates presented above suggest that marginal increases in health expenditure, whether funded through additional taxation, borrowing or reallocation from other spending departments, appear good value when compared with estimates of the equivalent consumption value of health. Recent reviews of stated preferences suggest that £30,000 per QALY might be a reasonable lower bound on individual willingness to pay to improve their health by one QALY [26, 27].

There is limited evidence about the marginal productivity of other types of public expenditure that could start to inform UK Treasury decisions during periodic public expenditure reviews. However, a similar approach to that adopted here may generate marginal productivity estimates for other categories of public expenditure, particularly those designed to reduce mortality and/or improve the quality of life. For example, local authority expenditure on public health and social care pursue these objectives and evidence about the marginal productivity of both would usefully inform any assessment of where public money is best spent.

Some assessment of the marginal productivity of health care expenditure is also unavoidable when considering more granular choices within the health care system. For example, when considering whether a proposed investment (e.g., a programme of care, access to a new drug, procedure or intervention) will improve health outcomes, a comparison of the expected health benefits of the investment with the expected health opportunity costs of the additional resources required is necessary [28].

Importantly, the marginal productivity estimates presented here start to reveal whether the unavoidable implicit or explicit norms used in making these types of decisions are reasonable and evidence based. For example, since 2004, the National Institute for Health and Care Excellence (NICE), which issues guidance to the UK NHS, has published an explicit range for the cost-effectiveness ‘thresholds’ used in its deliberative decision-making process [29]. NICE states that the ‘threshold’ ought to represent the health opportunity costs of the additional NHS costs of a new technology. However, the stated range (£20,000 to £30,000 per QALY) was, in fact, founded on the values implied by the decisions it made between 1999 and 2003 and has been widely recognised for some time (including by NICE) as having little empirical foundation [30].

More recent evidence suggests that the ‘thresholds’ implied by the decisions NICE actually makes are, in fact, much higher [31]. The estimates of marginal productivity in this paper suggest that guidance issued by NICE is likely to do more harm than good, reducing health outcomes overall for the NHS [32]. There may be legitimate reasons why such harms might be deemed appropriate, but it is only through the type of empirical analysis in this paper that the reasons for these ‘harms’ are likely to be articulated and explicitly justified.

Other established norms are also evident in published economic evaluations. For example, in the US, thresholds of $50,000 to $100,000 per QALY have become increasingly cited but are also widely recognised as having little evidential foundation [33]. The type of ‘thresholds’ previously promoted by the World Health Organisation became widely cited but do not reflect evidence of health opportunity costs [34]. Therefore evidence-based challenges to these implied and/or explicit norms seem warranted.

***5.4 Study limitations***

Our results are reasonably consistent; the point estimates suggest a cost per QALY of between £5,800 and £9,900 (CI: £4,200 - £22,300). That we obtain similar results for several time periods could be interpreted as evidence of the reliability of our findings. Alternatively, we may have been estimating the same relationship using ten different draws from the same data generating process.

The backward selection of covariates is one of several approaches to the identification of a more parsimonious specification. It involves the application of standard statistical tests -- which assume a single test of a pre-specified model -- to a sequence of steps to identify relevant explanatory variables, and this may mean that the reported p-values are incorrect[35]. We readily acknowledge this limitation but we persevere with this approach so that our results are comparable with a similar study that used the same approach with socio-economic instruments [7].

One issue that we have not addressed is the potential lagged effect of expenditure on mortality. We have the impact of expenditure in, say, 2005/06 confined to mortality in 2005/06/07. This precludes any lagged effect on mortality in, say, 2007/08. Similarly, the mortality effect in 2005/06/07 only reflects expenditure in 2005/06 and not, say, spending in 2004/05. Therefore, our estimates will reflect the full impact of expenditure on mortality if these two lagged effects cancel each other out for each disease group although, of course, the length of this lag may be longer for some areas (eg cancer) than for others (circulatory disease).

We should acknowledge the possibility that both mortality and expenditure might be driven by local differences in some unspecified variable that we have omitted from the study (i.e., that one of the assumptions required for 2SLS to be a consistent estimator is not met).  If this were true, any claims about the causality of the relationship we have estimated between health care expenditure and mortality would be weakened.  This possibility is inevitably true of any statistical analysis where the precise details of the true underlying causal mechanisms are not known with certainty.  We have examined all plausible possibilities based on the limited guidance that theory can offer, the previous empirical work that has been undertaken in this area, and all the sources of data that are currently available.

We believe that the funding rule approach adopted here offers a substantial advance on what has gone before and that the estimates we provide can inform, and bring more accountability to, important and unavoidable social choices about health care expenditure and how it is allocated. Nevertheless, the assumptions and methodological choices that we have had to make do generate uncertainty and this is not captured in the reported confidence intervals.

**6. Conclusions**

This paper builds on previous attempts to estimate the marginal benefit associated with English NHS expenditure. Previous studies used socio-economic variables as instruments for disease-specific expenditure but a new approach proposes the use of exogenous variables in the formula used to allocate funds to local health authorities. We have applied this approach to generate annual cost per QALY estimates for English NHS expenditure from 2003/04 to 2012/13. These estimates suggest an annual cost per QALY of between £5,800 and £9,900. These figures are close to those reported previously using a similar dataset but socio-economic rather than funding rule variables as instruments.

Currently, there is limited evidence about the marginal productivity of other types of public expenditure but the approach outlined here has obvious applications to other types of expenditure and, in particular, to specific components of healthcare expenditure (e.g., pharmaceuticals) and to other types of healthcare related expenditure (e.g., public health and social care).

The most recent results reported here are for 2012/13. Clinical Commissioning Groups (CCGs) replaced PCTs as commissioners of local health services in England from April 2013. CCGs lost about one-third of their commissioning remit with responsibility for primary care, specialised commissioning and public health transferred to more central administrators. We have no reason to believe that this re-allocation of commissioning responsibility will affect the relationship between total NHS expenditure and mortality, but the identification of separate expenditure totals for specific types of healthcare since 2013 may permit the estimation of the marginal productivity of these types of activity.

**References**

1. Nolte, E and McKee, M (2004). Does health care save lives? Avoidable mortality revisited. Research Report. The Nuffield Trust, London. Available at: https://www.nuffieldtrust.org.uk/research/does-healthcare-save-lives-avoidable-mortality-revisited [accessed 21 June 2020].

2. Nixon, J. and Ulmann, P (2006). The relationship between health care expenditure and health outcomes. European Journal of Health Economics, 7(1), pp.7-18. Available at: https://link.springer.com/article/10.1007/s10198-005-0336-8 [accessed 21 June 2020].

3. Edney, L.C, Afzali, H. H. A., Cheng, T. C., and Karnon J. Estimating the Reference Incremental Cost-Effectiveness Ratio for the Australian Health System. PharmacoEconomics, 2018, 36(2), pp.239–252. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29273843 (accessed 21 June 2020).

4. Van Baal, P, Perry-Duxberry, M, Bakx, P, Versteegh, M, van Doorslaer, E, and W Brouwer. A cost‐effectiveness threshold based on the marginal returns of cardiovascular hospital spending. Health Economics, 2019, 28, 1, 87-100. Available at: <https://doi.org/10.1002/hec.3831> (accessed 21 June 2020).

5. Vallejo-Torres, L., Garcia-Lorenzo, B., and Serrano-Aguilar, P. Estimating a cost-effectiveness threshold for the Spanish NHS. Health Economics, 2018, 27(4), pp.746-761. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3633> (accessed 21 June 2020).

6. Siverskog, J. and Henriksson, M. Estimating the marginal cost of a life year in Sweden’s public healthcare sector. Eur J Health Econ (2019) 20: 751-762. Available at: <https://doi.org/10.1007/s10198-019-01039-0> (accessed 21 June 2020).

7. Lomas, J., Claxton, K., and Martin, S. Estimating the marginal productivity of the English National Health Service from 2003/04 to 2012/13. Value in Health, 2019, 22, 9, 995-1002. DOI: <https://doi.org/10.1016/j.jval.2019.04.1926> (accessed 21 June 2020).

8. Andrews, M., Elamin, O., Hall, A.R. Kyriakoulis,,K and Sutton, M. Inference in the presence of redundant moment conditions and the impact of government health expenditure on health outcomes in England. Econometric Reviews, 2017, 36(1–3), pp.23–41. DOI: https://www.tandfonline.com/doi/full/10.1080/07474938.2016.1114205 (accessed 21 June 2020).

9. Claxton, K., Lomas, J. and Martin, S. The impact of NHS expenditure on health outcomes in England: Alternative approaches to identification in all-cause and disease specific models of mortality. Health Economics, 2018, 27(6), pp.1017-1023. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3650> (accessed 21 June 2020).

10. Staiger D, and Stock J. Instrumental variables regression with weak instruments. Econometrica, 1997, 65:557–586.

11. Angrist, Joshua & Pischke, Jörn-Steffen. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

12. Wooldridge J. M. (2010). Econometric analysis of cross section and panel data. Second edition. MIT Press.

13. Martin, S., Rice, N. and Smith, P.C. Does health care spending improve health outcomes? Evidence from English programme budgeting data. Journal of Health Economics, 2008, 27(4), pp.826–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18261812 (accessed 21 June 2020).

14. Martin, S., Rice, N. and Smith, P.C. Comparing costs and outcomes across programmes of health care. Health Economics, 2012, 21(3), pp.316–337. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21322086 (accessed 21 June 2020).

15. Claxton, K., Martin, S., Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith PC and Sculpher M. Methods for the estimation of the cost-effectiveness threshold. Health Technology Assessment, 2015 Feb, 19(14), pp.1-503, v-vi. Available at https://www.ncbi.nlm.nih.gov/pubmed/25692211 (accessed 21 June 2020).

16. Department of Health. Resource Allocation: Weighted Capitation Formula. Sixth Edition. 2008. Available at: https://webarchive.nationalarchives.gov.uk/20130124044910/http://www.dh.gov.uk/prod\_consum\_dh/groups/dh\_digitalassets/@dh/@en/documents/digitalasset/dh\_091848.pdf (accessed 21 June 2020).

17. Propper, Carol, and John Van Reenen. "Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labor Markets on Hospital Performance." *Journal of Political Economy* 118, no. 2 (2010): 222-73. DOI:10.1086/653137 (accessed 19 February, 2021).

18. Carr-Hill, R. A., Geoffrey Hardman, Stephen Martin, Stuart Peacock, Trevor A. Sheldon and Peter C. Smith. Interfaces, Jan. - Feb., 1997, Vol. 27, No. 1, Franz Edelman Award Papers (Jan. - Feb., 1997), pp. 53-70. Available from: <https://www.jstor.org/stable/25062211> (accessed 19 February, 2021).

19. Smith, Peter C. Resource allocation and purchasing in the health sector: the English experience. Bulletin of the World Health Organization, 2008, 86:884–888. Available from: <https://www.who.int/bulletin/volumes/86/11/07-049528.pdf> (accessed 19 February, 2021).

20. Baum, C.F., Schaffer, M.E., Stillman, S. (2010). ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. Available from: <http://ideas.repec.org/c/boc/bocode/s425401.html> (accessed 21 June 2020).

21. Pesaran, M.H. and L.W. Taylor, *Diagnostics for IV regressions.* Oxford Bulletin of Economics and Statistics, 1999. **61**(2): p. 255-265.

22. Soares, M, Sculpher, M and Claxton, K (2020). Health Opportunity Costs: Assessing the Implications of Uncertainty Using Elicitation Methods with Experts. Medical Decision Making, vol 40, no 4. See [https://doi.org/10.1177%2F0272989X20916450](https://doi.org/10.1177/0272989X20916450) (accessed 25 January, 2021).

23. Soares MO, Sculpher MJ, Claxton K. Authors’ Response to: “Health Opportunity Costs and Expert Elicitation: A Comment on Soares et al.” by Sampson, Firth, and Towse. Medical Decision Making. February 2021. DOI:10.1177/0272989X20987222 (accessed 16 March 2021).

24. Lakhani A, Olearnik H and Eayres D eds (2006). Compendium of clinical and health indicators: data definitions and user guide for computer files. London: National Centre for Health Outcomes Development. An updated version (December 2015) is available at: https://files.digital.nhs.uk/60/236EC2/Compendium%20User%20Guide%202015%20Dec%20Annex%203%20V1.pdf (accessed January 23, 2021).

25. Curtis, Lesley A. (2014). Unit Costs of Health and Social Care 2014. Other. Personal Social Services Research Unit, Canterbury Kent. See <https://kar.kent.ac.uk/46138/1/sources-of-information.pdf> (accessed 25 January, 2021).

26. Vallejo‐Torres, L., García‐Lorenzo, B., Castilla‐Rodríguez, I., Valcárcel‐Nazco, C., García‐Pérez, L., Linertová, R., Polentinos-Castro E, Serrano‐Aguilar, P. On the estimation of the cost‐effectiveness threshold: Why, what, how? Value in Health, 2016, 19(5), 558–556. DOI: <https://doi.org/10.1016/j.jval.2016.02.020> (accessed 21 June 2020).

27. Ryen, L. and Svensson, M. The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. Health Economics, 2015, 24(10), pp1289-1301. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3085> (accessed 21 June 2020).

28. National Audit Office. Investigation into the cancer drugs fund. HC 442, Session 2015-16. Available at: https://www.nao.org.uk/wp-content/uploads/2015/09/Investigation-into-the-Cancer-Drugs-Fund1.pdf (accessed 21 June 2020).

29. National Institute for Clinical Excellence. Guide to the Methods of Technology Appraisal, 2004, London. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment\_data/file/191504/NICE\_guide\_to\_the\_methods\_of\_technology\_appraisal.pdf (accessed 21 June 2020).

30. Rawlins M. D. and Culyer A.J. National Institute for Clinical Excellence and its value judgments. BMJ, 2004, 329(224), July. Available at: https://www.bmj.com/content/329/7459/224 (accessed 21 June 2020).

31. Dakin H, Devlin N, Feng Y, Rice N, O'Neill P, and Parkin D. The influence of cost-effectiveness and other factors on NICE decisions. Health Economics, 2014, 24 (10), pp.1256–1271.

Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3086 (accessed 21 June 2020).

32. Claxton K, Sculpher M, Palmer S, and Culyer A. J. Causes for concern: Is NICE failing to uphold its responsibilities to all NHS patients? Health Economics, 2015, 24(1), pp1-7. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3130> (accessed 21 June 2020).

33. Neumann P.J., Cohen J.T. and Weinstein M.C. Updating cost-effectiveness--the curious resilience of the $50,000-per-QALY threshold. The New England Journal of Medicine, August 28 2014, 371(9), pp796–797. Available at: https://www.nejm.org/doi/full/10.1056/NEJMp1405158 (accessed 21 June 2020).

34. Bertram M.Y., Lauer J.A., Joncheere K., Edejer T., Hutubessy R., Kienya M. and Hill S. Use and Misuse of Thresholds Cost–effectiveness Thresholds: Pros and Cons. Bulletin of the World Health Organization, 94, 2016, pp925-930. Available from: https://pdfs.semanticscholar.org/353e/9dd1b2ee92effae44f7982863a212f0c0fb2.pdf (accessed 21 June 2020).

35. Harrell, F. 2015. Regression Modelling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer Series in Statistics.

1. This work is independent research funded by the UK’s National Institute for Health Research Policy Research Programme (NIHR PRP) through its Policy Research Unit in Economic Evaluation of Health & Care Interventions (EEPRU, grant reference 104/0001).  The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care*.* [↑](#footnote-ref-1)