ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Health Policy Analysis

Estimating the Marginal Productivity of the English National Health Service From 2003 to 2012

James Lomas, PhD,[1,*] Stephen Martin, PhD,[2] Karl Claxton, PhD[1,2]

[1]Centre for Health Economics, University of York, York, England, UK; [2]Department of Economics and Related Studies, University of York, York, England, UK

ABSTRACT

Objective: Estimates of the marginal productivity of the health sector are required for a wide range of resource allocation decisions. Founding these estimates on robust empirical analysis can inform these decisions and improve allocative efficiency. This article estimates the annual marginal productivity of the English NHS over a 10-year period (between 2003 and 2012).

Methods: Data on expenditure and mortality by program budget category are used in conjunction with socioeconomic and demographic variables from the censuses for 2001 and 2011. This article applies an econometric strategy that employs an established instrumental variable approach, which is then subjected to a number of sensitivity analyses. The results of the econometric analysis, along with additional data on the burden of disease, are used to generate an estimate of the marginal productivity for each of the study years.

Results: We find that an additional unit of health benefit has cost between £5000 and £15 000 per quality-adjusted life-year from 2003 to 2012. Over this period these estimates (all in current prices) have increased at a faster rate than NHS price inflation, suggesting an increase in real terms.

Conclusions: These results are discussed in the context of the existing literature, and the potential policy implications for decisions about resource allocation are explored.

Keywords: allocative efficiency, econometric modeling, health opportunity costs, productivity, program budgeting.

VALUE HEALTH. 2019; 22(9):995–1002

Introduction

Although useful for resource allocation decisions, reliable estimates of the size of the causal link between healthcare expenditure and health outcomes are difficult to obtain. This is partly because of several empirical challenges, including the heterogeneity of observational units and that mortality might be influenced by expenditure and also influence it (reverse causality).[1] For these and other reasons, several studies have failed to identify a strong and consistent relationship between healthcare expenditure and health outcomes (after controlling for other factors).[2]

In a bid to overcome these econometric challenges and to provide policy-relevant estimates of marginal productivity for national decision making, recent studies have started to employ instrumental variable (IV)–based regression approaches using subnational rather than cross-country data.[3-8] A subnational approach has considerable advantages over the use of aggregate country-level data; for example, it permits the inclusion of a broader range of variables because numerous sources of data can be linked and available data are not constrained by the need for international comparability.

Several subnational studies use English data that reflect local-level information on expenditure, outcomes, and other factors.[3-7] Although these studies use similar datasets, they differ in how the effect of expenditure on outcomes is identified using IVs. One approach directly estimates the elasticity of all-cause mortality with respect to health expenditure.[6] Here, the IVs are chosen on the basis that the per-capita budget assigned to each health authority is the product of the national per-capita budget and 4 adjustments reflecting local circumstances, 3 of which are plausibly unrelated to mortality and are therefore suitable as instruments. These four adjustments are the local age index, local additional needs index, local input price index, and local distance from target Index. The authors argue that all are potentially exogenous with the exception of the local additional needs index.

The other approach[3-5] uses data on expenditure and outcomes in different disease areas (program budget categories, PBCs)

Conflict of interest: The authors have no conflicts of interest to declare.

* Address correspondence to: James Lomas, PhD, Centre for Health Economics, University of York, York YO10 5DD, England, United Kingdom. Email: james.lomas@york.ac.uk

https://doi.org/10.1016/j.jval.2019.04.1926

reported at a local level (for English primary care trusts, PCTs). An expenditure equation is estimated to quantify how the overall budget is allocated across PBCs, and an outcome equation is estimated to quantify the elasticity of PBC-specific mortality with respect to PBC-specific healthcare expenditure. The IVs used in this approach and also employed in this article reflect factors such as socioeconomic deprivation and the availability of informal care in the community, which are believed to influence healthcare expenditure but plausibly only indirectly affect mortality through their impact on expenditure. This second strategy for finding IVs has also been employed in the analysis of Australian data where an elasticity of all-cause mortality with respect to health expenditure of −2.2 is reported.[7,8] A third approach has recently been explored.[9] It is essentially a hybrid of the two approaches described in the text. This third approach employs IVs for total expenditure in order to estimate elasticities for program budget category (PBC)–specific mortalities instead of all-cause mortality. This approach is a promising avenue for future research as one benefit is that it is no longer assumed that mortality in a PBC is unaffected by spending in other PBCs.

The disease-specific elasticities of mortality with respect to expenditure are interesting results in themselves, but they do not fully express the marginal productivity of the NHS in the most useful way possible. This is because the NHS is concerned with not only extending life but also improving the quality of life. Therefore, a measure of marginal productivity should ideally reflect the effect of NHS expenditures on extending survival (resulting from reduced mortality) and improving health-related quality of life. Such a measure can be obtained by combining health outcome and expenditure elasticities with additional information about disease-specific life expectancy and morbidity burden by age and sex of the patient population.[5] Claxton et al[5] reports a cost per life-year of £25 214 for 2008 (ie, financial year 2008 and 2009); this reflects an estimate of marginal productivity that captures the effect of NHS expenditure on extending survival only. But, by using the effect of expenditure on the mortality as a surrogate for the effect on a measure of health burden that also includes morbidity burden, the study[5] also reports a cost per QALY of £12 936 for 2008, and this reflects the likely impact of expenditure at the margin on both mortality and morbidity.

This article builds on and extends a recent study.[5] That study reported cost per QALY estimates for 3 years (2006, 2007, and 2008) using PCTs as the unit of analysis, and all census-based variables reflected 2001 data. Mortality data, available in 3-year periods, ceased to be available for PCTs after 2008 (ie, 2008 to 2010) and hence, to facilitate updates, the unit of analysis used here is the "upper-tier" local government geography, hereafter local authority (LA). Moreover, the 2011 census is used to update the census-based variables. This article reports estimates of the marginal productivity of the English NHS, annually, for the 10-year period between 2003 and 2012. It applies the methodology used in previous work[5] to new data, constructed for a different unit of analysis, and undertakes a range of additional sensitivity analyses.

The plan of the article is as follows. The datasets are described in the Methods section, along with an overview of the empirical approach to estimation. In the Results section, annual marginal productivity results for the 10-year period (both point estimates and key percentiles of the distribution) are expressed in terms of the amount of resource used to produce a unit of health benefit (cost per QALY) and the volume of health benefits produced using a unit of resource (QALYs per unit of expenditure). These results are discussed in the Discussion section, before the Conclusions section. Additional information about the methods and results from the sensitivity analyses are presented in the Appendix (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1926).

Methods

Data

This article uses 3 sources of data to examine the relationship between NHS expenditure and mortality. *NHS expenditure* (adjusted for unavoidable cost-factors) by geographically defined local health authorities, PCTs, is available for 23 PBCs, annually, for financial year 2003 and 2004 to financial year 2012 and 2013 (hereafter referred to as 2003 and 2012, respectively), and this includes virtually all NHS expenditure on inpatient care, outpatient and community care, and pharmaceutical prescriptions. This dataset does not include non-NHS expenditures on health. This makes the resulting estimate useful mainly for decisions that concern the NHS and would not be directly relevant for decision making in the comparably small UK private healthcare sector. *Mortality rates* (standardized years of life lost rates, SYLLR) are available for 10 PBCs at the LA level averaged over 3-year periods from 2003 through 2005 to 2012 through 2014 (hereafter referred to as 2003 and 2012, respectively). Finally, *UK census data* for 2001 and 2011 are used to construct a dozen socioeconomic variables that were used by Claxton et al.[5] These variables include measures of the proportion of residents born outside the European Union; the proportion of the working-age population employed in managerial and professional occupations; the proportion of households that are owner occupied; and the proportion of the population that provides unpaid care. Further details about the census-based variables can be found in the Appendix (see section A2.1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1926). In the absence of intermediate observations, values for 2003 to 2010 are linearly interpolated from those observed for 2001 and 2011, and the value for 2012 is assumed to be the same as that recorded by the census for 2011. These socioeconomic variables are available as potential controls in the second-stage equation and as potential instruments for the first stage.

We adopt English LAs as a consistent geographical unit of analysis across these different sources and years of data. Mortality and census variables are directly available at LA level. However, the remaining variables (including the PBC expenditure data) are only available at PCT level, and these data are mapped from PCT to LA level using a tool developed by the UK Department of Health. The sensitivity of the results to the mapping tool was investigated as part of preliminary work, using data from 2008 where both PCT-level and LA-level are available, and the results were largely robust to inaccuracies resulting from the mapping process.[10]

Econometric Strategy

Our modeling framework derives from an underlying conceptual model that assumes that each PCT manager receives a fixed annual budget and allocates it across the 23 PBCs so as to maximize social welfare subject to a health production function. The optimal level of spending for a given PBC is a function of the total PCT budget, the need for healthcare spending in that disease area, environmental factors that affect health in that PBC, and need for healthcare spending and environmental factors that affect health in other PBCs. Health within each PBC is assumed to be a function of healthcare expenditure within that specific PBC only.

This framework suggests the estimation of an expenditure equation (1) for each of the 23 PBCs, and an outcome equation (2) for the 10 PBCs for which mortality data are available.

Accordingly, for the jth PBC we have:

$$x_i = \beta_0 + \beta_1 n_i + \beta_2 m_i + \beta_3 y_i + \varepsilon_i \quad (1)$$

$$h_i = \gamma_0 + \gamma_1 n_i + \gamma_2 x_i + \omega_i \quad (2)$$

All variables are log-transformed before estimation in accordance with standard practice in this literature and consequently all coefficients are interpreted as elasticities. The i subscript denotes the unit of observation (LA as opposed to PCT given data availability), y is the overall budget, h is mortality in the jth PBC, x is the expenditure on the jth PBC, n is the need for healthcare in the jth PBC, m is the need for care in other PBCs, β and γ arguments are parameters to be estimated (β_3 is referred to as an expenditure elasticity, γ_3 as an outcome elasticity), and ε and ω are error terms.

It may be the case that m in equation 1 and x in equation 2 are endogenous because other program need is proxied using the mortality rate in these other programs, and expenditure in a disease area may be related to unobservable factors that influence the level of health outcomes (eg, a high level of historical mortality). For these reasons, the OLS estimation of equations 1 and 2 would likely result in bias even if the analysis controlled for observable healthcare need (n). Of note, for equation 1, the endogeneity of m is important even though the coefficient of interest, β_3, is estimated on γ. One approach to controlling for this endogeneity is to use suitable instruments for the endogenous variables.[3–7] However, theory provides no guidance as to the specific IVs that should be used. In general, we need IVs that are associated with either PBC-specific expenditure or other-PBC mortality (this requirement is referred to as "instrument relevance"), but not PBC-specific mortality either directly or through some unobserved variable (this requirement is referred to as "instrument validity"). Instrument relevance can be directly tested, typically by requiring an F test of excluded instruments in the first stage and requiring that the test statistic exceeds 10.[11] Instrument validity cannot be directly tested, and expert judgment is required, but when an equation is overidentified (there are more excluded instruments than endogenous variables) then an overidentification test can be helpful, although it may lack power in rejecting the null hypothesis of joint validity in some circumstances.[12]

Each wave of data is analyzed separately. The preferred empirical specification for each PBC and for each year is identified using the following method. We use the preferred empirical specifications reported by others[5] for 2008 as the starting point for the estimation of outcome and expenditure equations for 2009. If the 2008 specification performs satisfactorily when re-estimated with 2009 data, then this becomes our preferred specification for 2009 too. A specification is deemed satisfactory if it passes a battery of statistical tests (including an endogeneity test, the Hansen-Sargan overidentification test, and the Kleibergen-Paap F test for instrument strength) and meets 3 priors: that expenditure reduces mortality ($\widehat{\gamma}_2<0$) and that expenditure on a given PBC increases with overall budget ($\widehat{\beta}_3<0$) and decreases with other PBC need ($\widehat{\beta}_2<0$).

If the specification fails a test, then it is revised (to make good this failure) and re-estimated. If this approach—of specification revision and re-estimation—fails to reveal an acceptable specification, then the entire equation is re-estimated with covariates and IVs using a backward step-wise procedure. In the rare circumstances where it proves impossible to obtain a satisfactory specification and outlier expenditure values are found, the sample is trimmed in an attempt to find an acceptable specification. This was undertaken for 7 regression specifications out of a total of 310 reported as part of this article. Further details of this estimation process are provided in the Appendix Section A1.1 (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1926). Once a specification has been obtained for 2009, analysis of 2010 draws upon the preferred econometric specification from 2009, and so on, up to 2012. The model specification for 2008 draws upon that for 2009, and then that for 2007 draws upon 2008, and so on, until an appropriate model was specified for each wave of data as far back as 2003 and up to 2012.

Translating Mortality Effects Into Quality-Adjusted Life-Years

The outcome and expenditure elasticities obtained via the estimation of equations 1 and 2 for those 10 PBCs with an outcome indicator, together with information about the number of life-years lost in each program, could be used to estimate the cost of a life-year.[3] The problem with such an estimate is that it would implicitly assume that expenditure in those programs with a mortality indicator has no effect on morbidity and that NHS expenditure on those programs without a mortality indicator has no health effect at all. These assumptions appear wholly implausible, especially when almost half of NHS expenditure is on programs without a mortality indicator and morbidity is a significant aspect of disease burden where a mortality indicator is available.

Complete morbidity and quality-of-life data by PBC and local area are not available, so direct estimates of a more comprehensive measure of the likely health effects of health expenditure (eg, QALY effects) are not possible. Nevertheless, previous work has linked estimated effects on mortality to the likely QALY effects of changes in NHS healthcare expenditure using plausible assumptions. For example, one study[5] drew on a number of data sources to estimate the QALY burden of disease for each 3-digit ICD10 code within each PBC. Data from the World Health Organization global burden of disease study[13] were used to estimate the duration and incidence of disease (by age and sex), Office for National Statistics data provided mortality conditional life expectancies by age and sex, quality of life norms by age and sex were based on data from the Health Survey for England, and the impact of disease on these quality of life norms were provided by Health Outcomes Data Repository supplemented with information from the Medical Expenditure Panel Survey.

They calculate the QALY burden of disease for each PBC by summing (over all relevant ICD10 codes) the product of the per-patient QALY burden and the size of the population with the disease (prevalent and incident) in 1 year. For each PBC with an outcome elasticity, the estimated change in its QALY burden associated with, say, a 1% change in the overall budget, can be calculated by forming the product of the outcome elasticity, the expenditure elasticity, and the QALY burden for the PBC (ie, effects on the mortality burden of disease are used as a "surrogate" for effects on the broader QALY burden). For those PBCs without an outcome indicator, the authors[5] calculate the average proportionate effect of a change in expenditure on the mortality burden of disease in those PBCs where mortality-based outcome elasticities can be estimated, and this average is used as a proxy for the outcome elasticity for those PBCs without a directly estimated outcome elasticity (the proportionate effects on burden of disease are extrapolated from where they can be observed to where they cannot).

Therefore, using the same approach to estimating the QALY burden of disease, combined with these surrogacy and extrapolation assumptions, we are able to estimate the total QALY

Table 1. Estimated outcome and expenditure elasticities, by program budget category 2003 to 2012.

Program Budget Category	2003		2004		2005		2006		2007	
	Outcome	Spend	Outcome	Spend	Outcome	Spend	Outcome	Spend	Outcome	Spend
Infectious diseases	−0.205	1.094*	−0.100	0.932*	−0.432	1.205*	−0.608	1.051*	−0.660†	1.387*
Cancer	−0.201†	1.711*	−0.224†	1.259*	−0.159‡	1.592*	−0.239*	1.219*	−0.273*	1.626*
Blood	n/a	0.652‡	n/a	0.952*	n/a	1.486*	n/a	1.037*	n/a	1.374*
Endocrine	0	0.653*	−1.843	0.573*	−1.035	0.663*	−1.464	0.630*	−1.491	0.455*
Mental health	n/a	1.333*	n/a	0.999*	n/a	0.991*	n/a	1.143*	n/a	1.103*
Learning disability	n/a	0.646‡	n/a	0.446‡	n/a	0.449‡	n/a	0.410	n/a	0.386
Neurological	−0.751‡	1.408*	−0.968†	0.929*	−0.325	1.220*	−0.869‡	0.382‡	−0.237‡	0.733*
Vision	n/a	0.833*	n/a	1.350*	n/a	1.127*	n/a	0.931*	n/a	1.106*
Hearing	n/a	0.694‡	n/a	0.526	n/a	0.762†	n/a	0.989†	n/a	0.951‡
Circulatory	−1.202*	1.873*	−1.375*	1.652*	−1.637*	1.477*	−1.404*	1.578*	−1.315*	1.614*
Respiratory	−1.666*	1.661*	−2.494*	1.253*	−2.217*	1.225*	−2.281*	1.287*	−1.564*	1.555*
Dental	n/a	0.717‡	n/a	0.848‡	n/a	1.224†	n/a	0.835†	n/a	0.420*
Gastro-intestinal	−1.493*	1.409*	−1.253*	0.928*	−1.014‡	1.076*	−1.255†	1.014*	−0.837†	1.490*
Skin	n/a	0.700*	n/a	0.595*	n/a	0.840*	n/a	0.701*	n/a	0.787*
Musculoskeletal	n/a	1.014*	n/a	0.567*	n/a	0.935*	n/a	0.628‡	n/a	0.733*
Trauma and injuries	0	0.556*	0	0.576†	0	0.897*	0	0.705*	−0.638	1.328*
Genito-urinary	−0.063	0.934*	−0.931‡	0.716*	−0.869‡	1.079*	−0.588	0.988*	−1.977	1.015*
Maternity and neonates	0	0.757*	−0.121	0.678*	−0.056	0.865*	−0.085	0.614†	−0.057	0.563†
Poisoning	n/a	2.327*	n/a	1.674*	n/a	1.735*	n/a	1.107*	n/a	1.674*
Healthy individuals	n/a	1.538†	n/a	0.709‡	n/a	0.507	n/a	0.709	n/a	1.296†
Social care	n/a	1.581*	n/a	1.313†	n/a	1.069‡	n/a	1.702*	n/a	1.669†
Other	n/a	0.681*	n/a	0.337*	n/a	0.532*	n/a	0.447*	n/a	0.553*

Note. Cells containing "n/a" or "0" are not incomplete, but indicate that no elasticity was estimated for that PBC year (or was estimated to be 0).
*$P < .01$.
†$P < .05$.
‡$P < .10$.

change associated with a change in total NHS expenditure for all PBCs and hence we can calculate the "cost per QALY." Clearly, linking the estimated effects on mortality to QALYs requires a number of assumptions to be made. For a more detailed summary of all assumptions, their justification, and a discussion of their likely impact on the central estimate of the cost per QALY, see Table 32 in Claxton et al.[5(p 83)]

Recently, the plausibility of these assumptions has been examined through structured elicitation from clinical experts,[14] and this work suggests that these assumptions are likely to be conservative with respect to the QALY effects of changes in expenditure (ie, the cost per QALY is likely to be lower than that estimated using these assumptions).

Results

Full regression results for the preferred outcome and expenditure specifications by PBC for each year, including specification test results, can be found in the Appendix Section A2.2, Tables A2.1 to A2.20 (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1926). The first stage regressions for the IV specifications can be found in the Appendix Section A2.3, Tables A2.21 to A2.40 (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1926). Of particular interest are (1) the coefficients on the expenditure variables in the outcome equations and (2) the coefficients on the budget variables in the expenditure equations. These coefficients (elasticities) are presented in Table 1.

Taking cancer in 2012 as an illustrative example, the expenditure elasticity is interpreted as saying that a 1% increase in overall NHS expenditure leads to a 1.027% increase in cancer expenditure. The outcome elasticity suggests that cancer mortality is reduced by 0.361% as a result of a 1% increase in cancer expenditure. As outlined in the section "Translating mortality effects into quality-adjusted life-years", these estimated elasticities can be combined with additional information about age, sex, life expectancy, and burden of disease of the patient population to produce estimates of the following: (1) volume of resources used to produce a unit of health benefit (cost per QALY) and (2) volume of health benefits produced using a unit of resource (QALYs per £ of expenditure). For the latter, we report the number of QALYs gained for £10 million, which represents a small amount of money relative to the overall level of NHS expenditure, and we label this the "health opportunity cost" (HOC) associated with £10 million of expenditure. All estimates are expressed in current prices and are shown in Table 2.

The cost per QALY and HOC estimates are generated first deterministically (column 1) and also probabilistically (columns 2 to 4). For the latter, draws are taken from independent normal distributions for each outcome and expenditure elasticity with mean and variance equal to the estimated coefficient and standard

Table 1. Continued

2008		2009		2010		2011		2012	
Outcome	Spend	Outcome	Spend	Outcome	Spend	Outcome	Spend	Outcome	Spend
−0.549*	1.471*	−0.310‡	0.968*	−0.256	1.006*	−0.305*	0.841*	−0.362*	0.749*
−0.287*	0.784†	−0.345*	0.502†	−0.220*	0.438	−0.430*	0.961†	−0.361*	1.027†
n/a	0.995*	n/a	1.060*	n/a	0.332	n/a	0.876*	n/a	1.119*
−1.607†	0.498*	−1.075†	0.708*	−0.174	0.696*	−0.199	1.116*	−0.499	0.951*
n/a	0.995*	n/a	0.899*	n/a	0.973*	n/a	1.194*	n/a	1.023*
n/a	0.329	n/a	0.647†	n/a	1.208†	n/a	0.741‡	n/a	0.000
−0.304	0.897*	−1.357	0.850*	−0.374	0.557*	−1.415	0.703*	−0.009	0.856*
n/a	0.701*	n/a	0.934*	n/a	0.997*	n/a	1.279*	n/a	1.411*
n/a	1.637*	n/a	1.273*	n/a	0.808‡	n/a	1.231*	n/a	1.523*
−1.384*	1.784*	−1.842*	0.494‡	−1.692*	1.013*	−1.611*	1.491*	−1.464*	1.285*
−1.671*	0.752†	−2.103*	0.576*	−2.006†	1.192*	−1.743*	1.360*	−1.704*	0.928*
n/a	0.428†	n/a	0.765*	n/a	0.229	n/a	0.843*	n/a	0.855*
−1.146†	0.520‡	−1.989‡	0.387‡	−1.425†	1.040*	−2.000†	1.033*	−1.904†	0.997*
n/a	0.907*	n/a	0.890*	n/a	0.422‡	n/a	0.681*	n/a	1.158*
n/a	0.738*	n/a	0.295	n/a	0.489†	n/a	0.456†	n/a	0.725*
0	1.344*	0	1.090*	−0.064	0.589†	0	1.024*	0	1.058*
−0.024	0.733*	−2.997	0.878*	−2.83	0.631*	−0.494	0.598*	−0.160	0.855*
−0.030	0.963*	−0.166‡	0.653*	−0.04	0.342	−0.136	0.481*	−0.106	0.833*
n/a	2.102*	n/a	0.658†	n/a	1.078†	n/a	0.631†	n/a	1.124*
n/a	1.049	n/a	1.246†	n/a	1.359†	n/a	1.748*	n/a	1.172‡
n/a	1.192‡	n/a	0.844	n/a	1.592†	n/a	1.859*	n/a	1.613*
n/a	0.338*	n/a	0.564*	n/a	0.520*	n/a	0.518*	n/a	0.585*

error squared, respectively. In total, 20 000 draws are made to generate a probability distribution for the overall result. The mean, 5th, and 95th percentiles of the generated distributions are presented in Table 2. The deterministic and probabilistic point estimates differ because of the nonlinear function that combines the estimated elasticities and additional information about survival and health-related quality of life. In particular, 2 features of the model are responsible for the nonlinearity: the extrapolation assumption and the adjustment to expenditure elasticities so that the changes in expenditure in all PBCs sum to the overall change in expenditure simulated (1% of total expenditure). This confidence interval reflects sampling uncertainty and is determined entirely by the standard errors of the estimated expenditure and outcome elasticities. The confidence interval does not reflect other sources of uncertainty arising from assumptions relating to the effect of expenditure on morbidity or any bias resulting from IV regressions with IVs that are not perfectly valid.

The numerical results shown in Table 2 are plotted in Figures 1 and 2. Figure 1 presents the deterministic point estimate for the cost per QALY along with its 90% confidence interval, whereas Figure 2 presents the deterministic point estimate in terms of the number of QALYs per £10 million of expenditure along with its 90% confidence interval.

Figure 1 shows that all point estimates of the cost per QALY lie between £5000 and £15 000, and Figure 2 shows that all point estimates for the HOC lie between 690 and 1860 QALYs. Although the point estimates are suggestive of an increase in the cost per QALY over the study period, the associated confidence intervals overlap each other considerably, and so it is not obvious that there has been a significant change in the cost per QALY. A similar argument can be made about the HOC. In addition, the HOC estimates do not decrease monotonically through time even though the cost figures are in current prices and NHS inflation averaged about 3% per annum during the study period.[15]

Discussion

Claxton et al[5] report a cost per QALY of between £10 000 and £15 000 for 2006, 2007, and 2008. This article contributes to the literature by extending the number of years analyzed and presents more recent estimates; our results suggest a cost per QALY of between £5000 and £15 000 for 2003 to 2012. The extension of the study period is only possible by changing the unit of analysis to LA level. In addition, this article uses more appropriate IVs by interpolating values from the 2001 and the 2011 censuses. Finally, this article describes a process where additional waves of data can be added to the dataset and the preferred specification for the previous year can be used to inform the specification for the additional wave.

Table 2. Marginal productivity for 2003 to 2012.

		Point estimate (deterministic)	Point estimate (probabilistic)	Fifth percentile	95th percentile
2003	Cost per QALY	£6381	£6381	£5048	£8534
	Health opportunity costs of £10 million (QALYs)	1567	1567	1172	1981
2004	Cost per QALY	£5389	£5377	£4110	£7517
	Health opportunity costs of £10 million (QALYs)	1856	1860	1330	2433
2005	Cost per QALY	£7613	£7635	£5611	£11 619
	Health opportunity costs of £10 million (QALYs)	1314	1310	861	1782
2006	Cost per QALY	£6844	£6838	£5139	£9878
	Health opportunity costs of £10 million (QALYs)	1461	1462	1012	1946
2007	Cost per QALY	£9747	£9765	£7689	£13 043
	Health opportunity costs of £10 million (QALYs)	1026	1024	767	1301
2008	Cost per QALY	£12 960	£13 271	£8390	£32 881
	Health opportunity costs of £10 million (QALYs)	772	754	304	1192
2009	Cost per QALY	£9887	£9920	£6802	£17 296
	Health opportunity costs of £10 million (QALYs)	1011	1008	578	1470
2010	Cost per QALY	£10 225	£10 214	£7073	£17 153
	Health opportunity costs of £10 million (QALYs)	978	979	583	1414
2011	Cost per QALY	£8997	£8985	£6520	£13 945
	Health opportunity costs of £10 million (QALYs)	1112	1113	717	1534
2012	Cost per QALY	£14 410	£14 411	£11 182	£19 861
	Health opportunity costs of £10 million (QALYs)	694	694	504	894

QALY indicates quality-adjusted life year.

The results for 2012 (the most recent year) indicate that an additional £10 million of NHS expenditure generates 694 QALYs. Expressing this as a ratio of incremental cost to incremental health effect yields an estimate of the marginal productivity of NHS expenditure (£14 410 per QALY). Nevertheless, as with most point estimates, there is uncertainty associated with it, and we find that the 90% confidence interval for the cost per QALY ranges from £11 182 to £19 861. Because only the incremental health effect is estimated with uncertainty and this measure appears in the denominator of the (cost per QALY) ratio, the confidence interval is not symmetric around the point estimate.

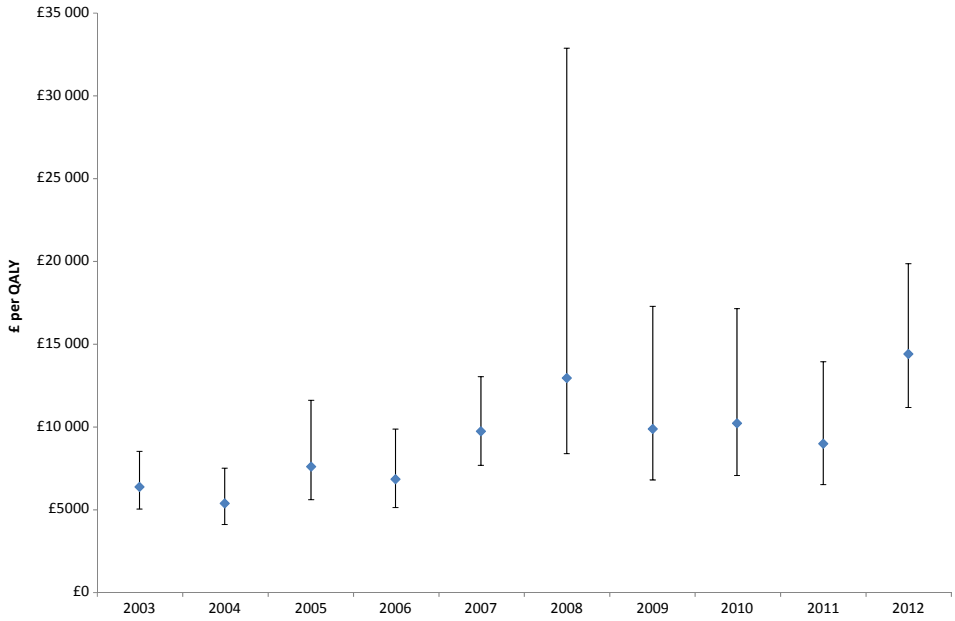
Although the reported point estimates should be used to calculate expected changes in health resulting from changes in expenditure, the degree of uncertainty in our estimates helps to assess the robustness of claims as to whether implicit or explicit established norms are compatible with this evidence or not. 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: £20 000 per QALY to £30 000 per QALY.[16] Although NICE makes clear that the threshold ought to represent the health opportunity costs of the additional NHS costs of a new technology, this range was, in fact, founded on the values implied by the decisions it made between 1999 and 2003.[17] The evidence from this article suggests that the NHS's marginal productivity is significantly higher (the cost per QALY is significantly lower) than that implied by NICE's stated guidance.

An assessment of the marginal productivity of healthcare expenditure is of general policy interest beyond guidance for health technology appraisal. For example, some judgment about the likely health effects of increasing or reducing public expenditure on healthcare is at the heart of debates about whether public expenditure should be increased to offer additional funding to the NHS or whether existing overall levels of public expenditure should be reallocated across spending departments. The aforementioned estimates suggest that marginal increases in health expenditure, whether funded through additional taxation, borrowing, or reallocation from other spending departments, appear as good value when compared with estimates of the equivalent consumption value of health (with recent reviews suggesting that £30 000 per QALY might represent a reasonable lower bound for this).[18,19] Where resource allocation decisions have been made on the implicit basis that the estimate of marginal productivity of the NHS is £30 000, or even £40 000 per QALY,[20] these decisions may have been suboptimal in terms of population QALYs.[21] QALYs may of course not be the sole objective of healthcare expenditure, and decisions may be made that lead to reductions in health (as measured by QALYs), but are judged worthwhile because of other considerations. A framework for analysis to inform decisions such as these is illustrated elsewhere.[22]

Similar studies from other healthcare systems vary widely in terms of methods used but produce similar results in the sense that the estimated marginal productivity is usually lower than the value implied by healthcare decision-makers.[7,18,23,24] In this way our results are consistent with the existing literature with marginal productivity estimated to be much lower than currently explicitly stated norms[25] or the values implied by actual decisions.[20]

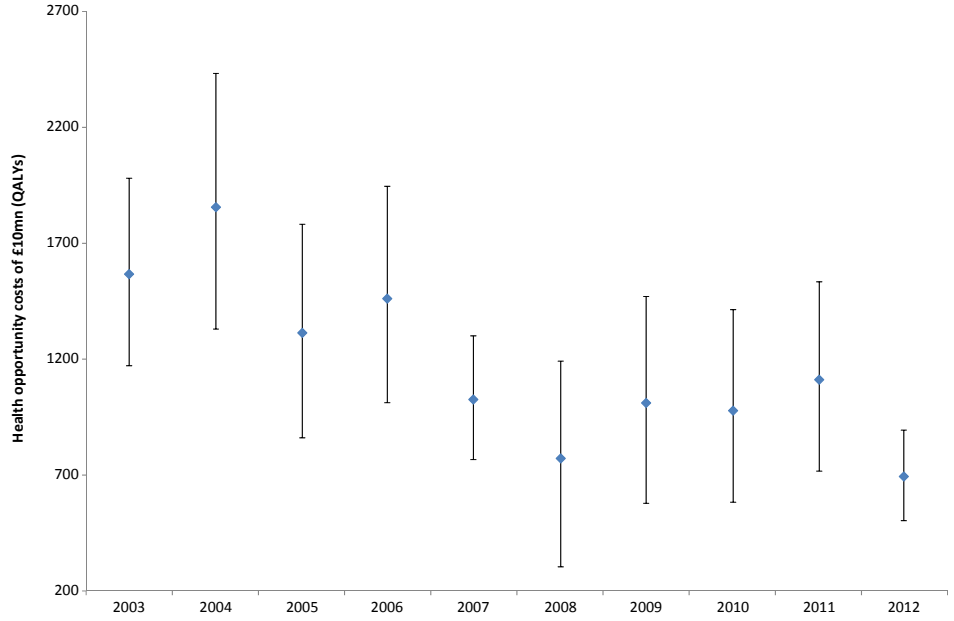
Looking at the results from this article over time, the cost per QALY in 2012 is more than double that of 2003, but some of this will reflect input price inflation faced by the NHS, roughly 28% over the same period.[15] Nevertheless, these results suggest that real productivity at the margin has fallen, which is consistent with diminishing marginal returns to health expenditure (expenditure has increased in nominal and real terms over most of the ten year period analysed[26]). Identifying factors that can explain this trend is difficult because there are many possible explanations for these changes, and it is left to future research to unpick the mechanisms underlying them.[27]

Figure 1. Marginal productivity for 2003 to 2012 expressed as cost per QALY with 90% confidence intervals.



QALY indicates quality-adjusted life year.

Figure 2. Marginal productivity for 2003 to 2012 expressed as QALYs per £10 million with 90% confidence intervals.



QALY indicates quality-adjusted life year.

At the core of this analysis is the econometric estimation of expenditure and outcome elasticities. A key component is the use of IVs to identify a causal effect. Our theoretical model does not tell us which IVs should be used and the validity of those selected cannot always be tested. Although for some PBCs there may be concern about the role of risk factors such as smoking and their association with socioeconomic status, generally speaking the mechanisms of cause and effect underlying the observed associations between socioeconomic status and PBC-specific mortalities are largely unknown. Nevertheless, the likely presence of endogeneity suggests that we should attempt to use some econometric approach beyond OLS, such as IV regression, to obtain consistent estimates of the causal effects. As such we rely on an approach where we are guided by statistical tests in addition to a priori plausibility that instruments could be valid. We have used appropriate statistical tests to guide model specification throughout as part of a clearly specified and documented protocol before the analysis was undertaken (see Appendix Section A1.1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1926). Nonetheless, the tests for validity can lack power to reject the null that IVs are

appropriately excluded from the second stage of the IV regression, particularly when all IVs might be thought to influence the endogenous regressor in the same kind of way.[12]

Nevertheless, the results of the just-identified sensitivity analysis and the poor performance of an OLS strategy, reported in Appendix Sections 1.2 and 1.3 (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1926), respectively, combined with other related work provides greater confidence and insight into how these considerations might influence a reasonable interpretation of the results in this article. First, in related work,[5] a sensitivity analysis[28,29] was undertaken to examine the impact of contaminated IVs (IVs that are not perfectly excluded from the second stage of the IV regression), which showed that contamination introduces additional uncertainty into the elasticity estimates, but not bias. Second, the implied all-cause elasticities using the approach taken to identification in this article (see Appendix Section A1.2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.04.1926) are comparable with the directly estimated all-cause elasticities in the literature.[6,9] The elasticities for key PBCs are also comparable.[9] The fact that similar results are obtained when a very different approach to identification is taken generates confidence in the census-based instruments that they are plausibly valid and that the results are not highly specific local average treatment effects. This is especially important in the more common context where the identification strategy pioneered by Andrews et al[6] is not possible. Taken together these considerations provide reassurance that our IV strategy is appropriate and that the estimates provided in this article are not seemingly biased in a particular direction. It also suggests that there is inevitably additional structural uncertainty that is not reflected in the confidence intervals reported in Table 2 and Figures 1 and 2.

Conclusions

Given the interest among economists in decision making at the margin, and the longstanding interest in the productivity of publicly funded institutions like the NHS, it is surprising that so few studies have sought to estimate the marginal productivity of the NHS. Its usefulness is not limited to decisions within the healthcare sector but is essential to inform the allocation of scarce public resources across sectors too. This article has shown how econometric analysis can be used to provide estimates of the marginal productivity of the NHS, with results expressed as either cost per QALY or QALYs per unit of expenditure. The results show that despite the inflation, the cost per QALY has remained relatively stable over time, with point estimates of the amount of resources, in nominal terms, to produce an additional unit of health benefit ranging from £5000 to £15 000 per QALY over the period between 2003 and 2012.

Acknowledgments

This article is independent research funded by the 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 PR-PRU-1217-20401). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. The views expressed are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health and Social Care or its arm's length bodies, or other UK government departments. Any errors are the responsibility of the authors.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2019.04.1926.

REFERENCES

1. Gravelle HS, Backhouse ME. International cross-section analysis of the determination of mortality. *Soc Sci Med*. 1987;25(5):427–441.
2. Gallet CA, Doucouliagos H. The impact of healthcare spending on health outcomes: a meta-regression analysis. *Soc Sci Med*. 2017;179:9–17.
3. Martin S, Rice N, Smith PC. Does health care spending improve health outcomes? Evidence from English programme budgeting data. *J Health Econ*. 2008;27(4):826–842.
4. Martin S, Rice N, Smith PC. Comparing costs and outcomes across programmes of health care. *Health Econ*. 2012;21(3):316–337.
5. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19(14):1–503.
6. Andrews M, Elamin O, Hall AR, Kyriakoulis K, Sutton M. Inference in the presence of redundant moment conditions and the impact of government health expenditure on health outcomes in England. *Econom Rev*. 2017;36(1-3):23–41.
7. Edney LC, Haji Ali Afzali H, Cheng TC, Karnon J. Estimating the reference incremental cost-effectiveness ratio for the australian health system. *Pharmacoeconomics*. 2018;36(2):239–252.
8. Edney LC, Haji Ali Afzali H, Cheng TC, Karnon J. Mortality reductions from marginal increases in public spending on health. *Health Policy*. 2018;122(8):892–899.
9. Claxton K, Lomas J, 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. *Heal Econ*. 2018;27(6):1017–1023.
10. Claxton K, Lomas J, Martin S. Switching to local authorities (LAs) as the unit of analysis (2008/09 expenditure). https://www.york.ac.uk/media/che/documents/PCT to Local Authority unit of analysis_08_09.pdf. Accessed January 18, 2018; 2017.
11. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica*. 1997;65(3):557.
12. Kovandzic T, Schaffer ME, Kleck G, Schaffer ME, Kleck G. Estimating the causal effect of gun prevalence on homicide rates: a local average treatment effect approach. *J Quant Criminol*. 2013;29:477–541.
13. WHO. The global burden of disease: 2004 update. 2008. http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/. Accessed January 18, 2018.
14. Soares MO, Sharples L, Morton A, Claxton K, Bojke L. Experiences of structured elicitation for model-based cost-effectiveness analyses. *Value Health*. 2018;21(6):715–723.
15. Curtis L. *Unit Costs of Health and Social Care*. Canterbury: Personal Social Services Research Unit, University of Kent; 2014.
16. NICE. Guide to the Technology Appraisal Process (reference N0514). 2004. http://webarchive.nationalarchives.gov.uk/20080205132341/http://nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/guide_to_the_technology_appraisal_process_reference_n0514.jsp. Accessed December 6, 2015.
17. Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. *BMJ*. 2004;329(7459):224–227.
18. Vallejo-Torres L, García-Lorenzo B, Castilla I, et al. On the estimation of the cost-effectiveness threshold: why, what, how? *Value Health*. 2016;19(5):558–566.
19. Ryen L, Svensson M. The willingness to pay for a quality adjusted life year: a review of the empirical literature. *Health Econ*. 2015;24(10):1289–1301.
20. Dakin H, Devlin N, Feng Y, Rice N, O'Neill P, Parkin D. The influence of cost-effectiveness and other factors on nice decisions. *Health Econ*. 2015;24(10):1256–1271.
21. Claxton K, Sculpher M, Palmer S, Culyer AJ. Causes for concern: is NICE failing to uphold its responsibilities to all NHS patients? *Health Econ*. 2015;24(1):1–7.
22. Sculpher M, Claxton K, Pearson SD. Developing a value framework: the need to reflect the opportunity costs of funding decisions. *Value Health*. 2017;20(2):234–239.
23. Ochalek J, Lomas J, Claxton K. Estimating health opportunity costs in low-income and middle-income countries: a novel approach and evidence from cross-country data. *BMJ Glob Heal*. 2018;3(6):e000964.
24. Vallejo-Torres L, García-Lorenzo B, Serrano-Aguilar P. Estimating a cost-effectiveness threshold for the Spanish NHS. *Health Econ*. 2018;27(4):746–761.
25. NICE. Guide to the methods of technology appraisal. 2013. https://www.nice.org.uk/process/pmg9/chapter/foreword. Accessed July 31, 2017.
26. Nuffield Trust. NHS in Numbers. https://www.nuffieldtrust.org.uk/resource/nhs-in-numbers. Accessed January 18, 2018.
27. Paulden M, O'Mahony J, McCabe C. Determinants of change in the cost-effectiveness threshold. *Med Decis Mak*. 2017;37(2):264–276.
28. Small DS. Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *J Am Stat Assoc*. 2007;102(479):1049–1058.
29. Conley TG, Hansen CB, Rossi PE. Plausibly exogenous. *Rev Econ Stat*. 2012;94(1):260–272.