Office of Health Economics Research

Research Paper

Exploring Variations in the Opportunity Cost Cost-Effectiveness Threshold by Clinical Area:

Results from a Feasibility Study in England

March 2019

Karla Hernandez-Villafuerte, Bernarda Zamora, Yan Feng, David Parkin, Nancy Devlin, Adrian Towse

Exploring Variations in the Opportunity Cost Cost-Effectiveness Threshold by Clinical Area: Results from a Feasibility Study in England

Karla Hernandez-Villafuerte[a], Bernarda Zamora[b], Yan Feng[c], David Parkin[b], Nancy Devlin[d] and Adrian Towse[b]

[a]German Cancer Research Center (DKFZ), [b]Office of Health Economics, [c]Queen Mary University of London, [d]University of Melbourne

March 2019

Please cite this report as:

Hernandez-Villafuerte, K., Zamora, B., Feng, Y., Parkin, D., Devlin, N., and Towse, A. (2019). *Exploring Variations in the Opportunity Cost Cost-Effectiveness Threshold by Clinical Area: Results from a Feasibility Study in England*. OHE Research Paper, London: Office of Health Economics. Available at: https://www.ohe.org/publications/exploring-variations-opportunity-cost-cost-effectiveness-threshold-clinical-area

Corresponding author: Adrian Towse. atowse@ohe.org

The Office of Health Economics (a company limited by guarantee of registered number 09848965; a charity with registered charity number: 1170829)
Southside, 7th Floor
105 Victoria Street
London SW1E 6QT
United Kingdom

For further information please contact
Adrian Towse
Director Emeritus and Senior Research Fellow
Tel: +44(0)207 747 1407
atowse@ohe.org

About OHE Research Papers

OHE Research Papers are intended to provide information on and encourage discussion about a topic in advance of formal publication. They are subject to internal quality assurance and undergo at least one external peer review, usually by a member of OHE's Editorial Panel. Any views expressed are those of the authors and do not necessarily reflect the views or approval of the OHE's Editorial Panel or Research and Policy Committee, or its sponsors.

Once a version of the Research Paper's content is published in a peer review journal, that supersedes the Research Paper and readers are invited to cite the published version in preference to the original version.

Funding and Acknowledgements

This study was commissioned and funded by the Association of the British Pharmaceutical Industry.

Steering Group

The authors would like to thank those individuals who attended meetings of the project Steering Group. It should be noted that individuals' participation in the Steering Group was as subject experts rather than representatives of their respective organisations.

John Cairns, Paul Catchpole, Warren Cowell, Nick Crabb, Mireia Jofre-Bonet, Sarah Karlsberg, Alistair McGuire, David Taylor.

Table of Contents

ABSTRACT

Background

The English National Health Service (NHS) is tax-funded and is expected to meet a budget constraint each year. Adopting a new net cost-increasing medical technology into the NHS therefore means potentially diverting resources from some other use. The benefits of the new technology have to be weighed up alongside the benefits foregone from those other uses i.e. their opportunity cost.

The National Institute of Health and Clinical Excellence (NICE) assesses the incremental cost-effectiveness ratio (ICER) of new medical technologies on behalf of the NHS. A technology is deemed to be good value for money (or not) by comparing this ratio against a cost effectiveness threshold which is intended to reflect this opportunity cost.

In 2015, Claxton et al published a seminal paper estimating an average "marginal" cost effectiveness threshold for the English NHS by estimating the marginal effect of changes in healthcare budget on the reduction of mortality rate among patients' population, across different clinical areas, termed Programme Budget Categories (PBCs).

Aims

The aim of this paper is to provide further empirical evidence on the relationship between health outcomes and health expenditures in England. The methods in this paper aim to address two limitations of the Claxton et al (2015) work, by:

- going beyond a focus on preventing premature deaths by examining health outcomes, aligned with NHS priorities, defined in terms of five outcome domains;
- examining the relationship between mortality and health expenditures at different parts of the mortality distribution, rather than assuming the relationship to be linear.

In addition the paper explores the extent of inefficiency in the NHS which, at the margin, can also impact on the opportunity cost faced when adopting a new technology.

Methods

Two methods are used to explore the marginal relationship between health expenditure and health outcomes: Data Envelopment Analysis (DEA) and Quantile Regression (QR). DEA allows the incorporation of multiple outcomes (not just mortality) and the measurement of efficiency and scale elasticity, while QR allows us to look for non-linearities in the relationship between spending and mortality.

DEA was applied to health outcomes and health expenditure data from 151 Primary Care Trusts (PCTs) in England across seven PBCs. Two environmental variables were selected (the deprivation index and budget shortfalls against formula) to adjust for factors affecting efficiency that were outside of the control of PCT managers, using a three step procedure. The QR method was applied to estimate the mortality rate as a function of health expenditure and a set of covariates using data from 151 PCTs in England across six PBCs. The method recognises the non-negative, highly asymmetric and leptokurtic distribution of health expenditure. Point estimates of the mortality elasticity to health expenditure are compared at different parts of the mortality distribution.

Finally, we compare the ranking of PCTs according to the DEA efficiency scores and the outcome elasticities estimated in the QR approach.

Results

Results from DEA show that efficiency varies across PCTs and PBCs. PCTs achieve a range of health outcomes which cannot be adequately explained by concentrating on reductions in the mortality rate. The results from QR analysis provide evidence of heterogeneity across PCTs and PBCs regarding the way health resources are used to improve outcomes. The results suggest that the marginal effect of health expenditure on mortality rate is not constant across PCTs and PBCs.

The comparison of PCT rankings from DEA and QR analysis are consistent and robust. In general, efficient PCTs (based on the DEA results) tend to have a lower absolute value of mortality elasticity to health expenditure (based on the QR results). A plausible explanation for these results is that PCTs operating efficiently in a PBC tend to have lower rates of mortality, and for most disease areas, the lower the mortality, the harder it is to achieve additional reductions.

Policy Implications

Estimation of an opportunity cost-based cost-effectiveness threshold using a health production function approach involves many assumptions about the behaviour of the implied function. These are compounded by the nature of the programme budgeting data that are used for estimation. This study uncovers further problems with making assumptions that may undermine attempts to obtain a simple singular system-wide threshold estimate.

This study provides empirical evidence of production inefficiency, that is the inability of some PCTs to achieve the best practice performance found in others. This means that estimates of the opportunity cost of introducing new technologies based on average performance could be (i) biased and (ii) subject to far greater variation than normally assumed. Moreover, the PCTs who are found to be inefficient vary between PBCs, confounding further the plausibility of estimates based on averages. There is evidence for some PBCs that some apparent inefficiencies result from adoption of a different underlying production

function technology, casting further doubts on the assumption of a common production function for all that underlies a common threshold.

The implications of this for setting a cost-effectiveness criterion for NICE and other NHS bodies are therefore not straightforward. However, they suggest that the direct translation from estimated levels of historic opportunity cost to cost-effectiveness thresholds for future investment is not justified. The average estimates generated by current research use a very large number of empirical and theoretical assumptions that are equally hard to justify, particularly when different approaches and assumptions produce radically different estimates.

One way to approach this problem is to accept that there are multiple sources of information relevant to the setting of cost-effectiveness criteria and that these may not be capable of being synthesised using scientific methods alone, but involve political judgements. A possible source of information would be an NHS mandated, targeted and supported survey of opportunity costs in terms of services at the local level, to generate routine data on this issue; ad hoc academic studies such as Appleby et al (2008) are not adequate for that purpose. Evidence should also be incorporated on the likely effects of any criteria that are set. An alternative would be an independent public body specifically tasked with assessing the evidence, commissioning evidence where it is lacking, and publishing the evidence and the body's deliberations and conclusions (Appleby, Parkin and Devlin, 2007).

1. INTRODUCTION

In a previous paper we reviewed the challenges in using econometric models to approximate the opportunity costs of adopting a new medical treatment through the estimation of the marginal relationship between health outcomes and health care expenditures (Hernandez-Villafuerte, Zamora and Towse, 2018). In principle, exploring this relationship should provide a good estimation of the opportunity costs of adopting health technologies. Getting good data is the main challenge. Another critical issue is the definition of health outcomes used. A further issue is methodological: if we accept mortality rate as one outcome measure, as has been considered in different models (Claxton et al., 2015; Martin, Rice and Smith, 2008; 2012), the application of a linear regression model can result in a limited summary of the effect of health expenditure on mortality across health locations.

This paper presents additional empirical evidence of the relationship between health outcomes and health expenditures in England seeking to overcome two main limitations of the current models: (1) health outcomes are aligned with NHS priorities defined along five outcome domains which go beyond preventing premature deaths, and (2) the relationship between mortality outcomes and health expenditures departs from a linear model, presenting different point estimates of the relationship at different parts of the mortality distribution.

The first method presented in this paper: Data Envelopment Analysis (DEA), considers multiple health outcomes in a number of clinical areas, some of which may not be well reflected in improvements in QALYs or in mortality reduction; examples include the outcomes used in mental health disorders.

The second method we present, Quantile Regression (QR), has been mainly used in health economics to describe the distribution of health expenditures which depart from normality: healthcare expenditures are non-negative, highly asymmetric and leptokurtic. The relationships between covariates and costs are likely to be non-linear. A growing literature in econometrics has developed techniques to model the entire distribution of health expenditures conditional on a set of covariates, thus 'going beyond the mean' (Fortin, Lemieux and Firpo, 2011; Jones, 2011). A description of the conditional distribution of mortality as a function of health expenditures and other covariates is presented by a family of quantile conditional functions, resulting in point estimates of the mortality elasticity to health expenditure which can be compared at different parts of the distribution, for example, for Primary Care Trusts (PCTs) at the lower and upper tails of the mortality distribution.

2. METHODS

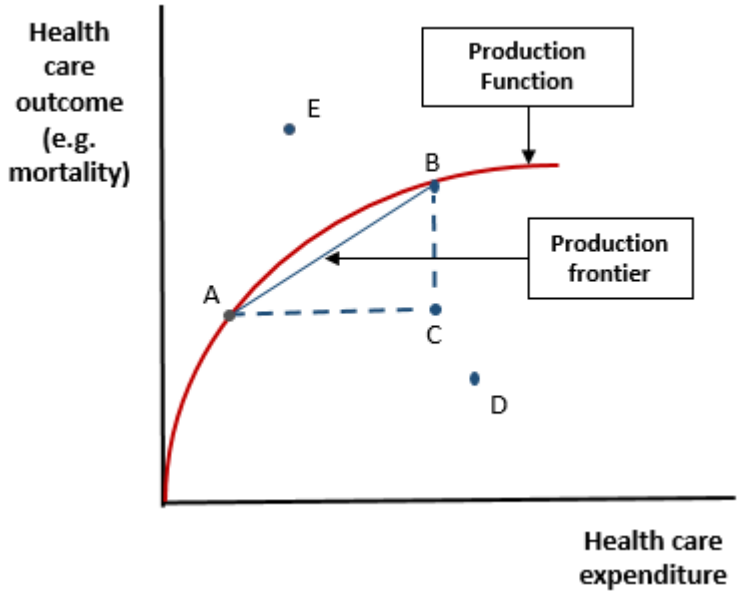
2.1. Data Envelopment Analysis

DEA is applied to estimate the effect of a marginal change in expenditures on the mix of health outcomes related to a specific disease area (henceforth called a Programme Budget Category (PBC) - see Appendix 1). DEA is a useful and robust methodology to evaluate the efficiency of the health system (Cylus, Papanicolas and Smith, 2015; Emrouznejad, Parker and Tavares, 2008; Hollingsworth, 2008; Pelone et al., 2014)[1].

[1] We reproduce here some of the discussion of the strengths and weaknesses of the use of DEA we set out in Hernandez-Villafuerte, Zamora and Towse, 2018, Section 3.2.

For simplicity, the concept of the DEA is explained using an illustration based on four economic concepts: the *production function*, *technical efficiency*, *economic efficiency* and the *production frontier*. A **production function** shows the maximum outcome combination that can be produced during a specific time period given fixed resource inputs and a fixed production technology. **Technical efficiency** is achieved when a production unit achieves this maximum output for a given level of physical resource inputs; it is not possible for the unit to produce more output without using more inputs, or to use fewer inputs without producing less output. If the volume of resource inputs is aggregated using a common value base, their unit cost, then it is possible to construct a production function having a single input. An example of this is shown in Figure 1 where the red line represents the maximum level of a particular health care outcome that can be achieved with different levels of health care expenditures. In this case, the production function shows not only technically efficient production, in terms of physical resources, but also **economic efficiency**, meaning the lowest possible cost of producing a given output level, or the greatest possible output for a given budget.

Figure 1. Economic efficiency



Units producing at points A and B in Figure 1 are efficient, but those at points C and D are not; for point C, for example, it would be possible to produce greater health outcomes with the same expenditure, as in point B, or have lower expenditure for the same outcomes, as in point A. Any point on the function between A and B would represent an unambiguous improvement, in terms of efficiency, for a unit at Point C. Point E, and any other point above the production function, represents a level of production that cannot be achieved with the current production technology.

Decision making units on points C and D are not efficient relative to A or B, which are the two PCTs on the production function. Additionally, we can also say that D is not efficient relative C, since it has higher expenditure and lower output. A key aim of DEA is to estimate relative efficiency, measured by the distance that an inefficient unit is from the technically efficient production function. In this regard, the distance from D to the

production function is higher than the distance from C. The distance between a line and a point is not unambiguously defined, so DEA uses a *Farrell radial measure*, which restricts the comparison to the point on the production function where outputs and inputs are reduced in the same proportion as in the inefficient unit's real level of these.

Production functions represent production points that may not be observed in the real world – for example there may be no production units that have the production levels that could theoretically be achieved. Instead, DEA and other estimation methods are based on the concept of a **production frontier**, also known as a *best-practice* frontier. Within a group of production units, this consists of all units that are not inefficient relative to any other unit. In Figure 1, ignoring the impossible point E, producers at points A and B are both more efficient than those at points C and D, but not inefficient relative to each other. They therefore form the production frontier, shown as a blue line, from which the relative inefficiency of the producers represented by points C and D can be measured.

Because DEA compares each producer, in our case English healthcare areas at Primary Care Trust (PCT) level, with only the best producers, known as *peers*, problems related to analyses based on averages are avoided. DEA does not assume any functional form for the production function. An advantage of DEA is that takes the two dimensional idea represented in Figure 1 and translates it into an approximation of a production function that allows the analysis of both multiple inputs, and multiple outcomes. This is important in the analysis of health systems since it allows more than one health outcome indicator to be included, and therefore can better capture the complexity of each PBC and the effect of a change in health expenditures on the population.

The opportunity cost of an effective new health care technology can be regarded as a effect of reducing the budget available for other technologies, offset by an improvement in health outcomes. For producers A and B, a new technology offering the same incremental cost-effectiveness ratio (ICER) as the current most marginal technology would enable them to substitute the technologies without impacting on overall cost-effectiveness. A new technology that has a better ICER than the current most marginal would enable them to generate more output from the current budget. However, producers C and D could adopt the new technology even if it has a worse (higher) ICER than current technologies, if they are able to improve efficiency.

2.1.1. DEA Assumptions

We used input oriented DEA (Bogetoft and Otto, 2011). We are interested in approaching the opportunity cost of the adoption of a new treatment in terms of estimating any decrease in health outcomes that might result from the displacement of resources needed to fund the new treatment. This is exactly the idea behind input oriented DEA. It allows us to observe how much the inputs of PCTs that are not efficient could in principle be decreased without affecting the outcomes delivered. In other words, to observe how much of the inputs (resources) can be displaced before there is an opportunity cost in terms of outcomes. The efficiency score estimated will indicate the proportional reduction of inputs needed by an inefficient PCTs to be efficient, while its outcomes are held constant. This cannot be captured by the output oriented DEA since in this case we would observe how much health output could increase without increasing inputs.

In addition, we assume **variable returns to scale** (VRS) instead of increasing returns to scale (IRS), constant returns to scale (CRS) or decreasing returns to scale (DRS). The returns to scale concept is linked to the production function. It explains the relationship

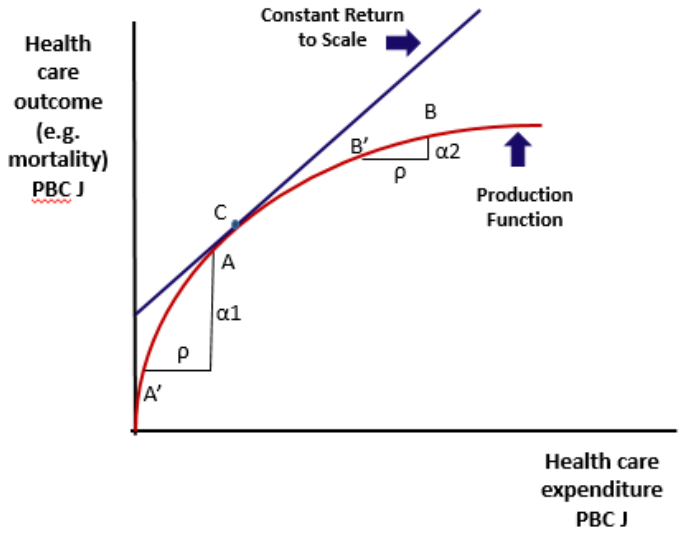
between the rate of increase in health outcomes and the rate of increase of the inputs (resources). The assumption of CRS is that any possible combination of health outcomes and inputs can be scaled up or down in a constant ratio. In other words, output increases by the same proportional change as all inputs change. The CRS assumption requires that the effect of a decrease in expenditures would have the same constant proportional effect on health outcomes regardless of the PCT, which is unlikely. The results in Hernandez-Villafuerte, Zamora and Towse, 2018 suggest that English local health locations cannot all be considered to behave similarly to one another. Based on the patterns of per-capita expenditures by PBC and PCT, the authors identified at least two distinct groups of PCTs each year and a number of PCTs that belong to the same distinct group over time. This empirical evidence suggests that PCTs have different production functions to one another, therefore, differences in returns to scale can be expected.

The VRS assumption in the DEA model indicates that PCTs may exhibit increasing, constant or decreasing returns to scale, giving the potential to better fit the data. If a PCT shows DRS the health outcomes will increase/decrease less than the input (resources). IRS indicates that health outcomes will tend to increase/decrease faster than the input. The assumption of VRS allows for the health outcomes produced (the ratio of health outcomes per input unit used), to change as output increases, for example, first to increase, then to be constant and then to decrease, as shown by the red line in Figure 1.

In order to test that the VRS assumption is the correct assumption to be used in the estimation, we used the Simar and Wilson (2002) and (Simar and Wilson, 2011) returns to-scale test for input-oriented DEA models. In addition, the Kruskal-Wallis rank sum test was also estimated. Both tests – the Kruskal-Wallis rank and the returns to-scale test for input-oriented DEA models - led to the same results.

A closely related concept is the idea of **scale elasticity**. Scale elasticity is the proportional change in outcomes resulting from a change in inputs. In Figure 2 we look at clinical area PBC j. If A and B need to reduce expenditures in an amount equal to ρ, the reduction in health outcomes experienced by PCT A, marked as α1, would be larger than for PCT B (α2). PCTs such as PCT A have IRS in clinical area PBC j, such that a 1% decrease in expenditures will decrease health outcomes by more than 1%. In this case the value of the scale elasticity will be higher than 1. The converse of this is that increasing economies of scale means that 1% decrease in expenditures will decrease health outcomes by less than 1%, such as in the case of PCT B. For a PCT like B the value of the scale elasticity in PBC j will be lower than 1. Finally, a situation in which a 1% decrease in expenditures will be translated in 1% decrease in health outcomes is constant returns to scale and a scale elasticity equal to 1. This is represented by a PCT such as C in Figure 2.

Figure 2. Scale Elasticity



If PCT A in Figure 2 is efficient in the production of health outcomes in two PBCs, one with a scale elasticity higher than 1 and the other with a scale elasticity lower than 1, the PBC with a scale elasticity higher than one has an implied higher opportunity cost in comparison with the PBC with scale elasticity lower than one. In other words, more output will be lost by cutting back expenditure in the first area rather than in the second.

Scale efficiency can be approached by comparing the DEA efficient scores estimated by assuming DRS with those estimated by assuming CRS. CRS is illustrated at point C on the production function illustrated by the blue line presented in Figure 2.

From the point of origin up to point C the production function exhibits IRS. Beyond point C it exhibits DRS. Those PCTs whose results indicate that they are below optimal scale efficiency are in the increasing or constant returns to scale part of the production function (PCTs such as A in Figure 2), meaning they have a scale elasticity higher or equal than 1. Although the data reported in the figures presented in the Results section are those of the DEA assuming VRS, DEA efficient scores using DRS and CRS are also estimated and compared. Given that we are interested in the opportunity cost of PCTs by PBC, we will be particularly focused on identifying those PCTs that are in the IRS part of the productivity function. These are the efficient PCTs with the highest opportunity costs, the percentage lost in health outcome as a result of a percentage decrease in expenditure will be the highest in this group (scale elasticity higher than 1).

DEA efficiency scores are sensitive to extreme values or **outliers**. Therefore, the test statistic to identify outliers was applied (Bogetoft and Otto, 2011).

$$\frac{\sum_{h=1}^{K/k}(E(h,K\backslash k)-1)^2}{\sum_{h=1}^{K/k}(E(h,K)-1)^2} \qquad (1)$$

Here, K is both the set and the number of PCTs in the data set, and k is a potential outlier. K/k is the set K excluding the PCT k. Additionally, E(h,K) is the efficiency of the PCT h when all PCTs are used, and E(h, K\k) is the efficiency when PCT k does not enter into the estimation. The test therefore compares the average efficiency of the other PCTs when PCT k is not considered with the average efficiency of the other PCTs when PCT k

is part of the evaluation. Because E(h,K\k) >= E(h,K) this ratio is always less than or equal to 1, and the smaller the ratio, the larger the impact of k, i.e. small values will be an indication that k is an outlier. All PCTs for which the estimated value is less or equal to **0.975** are consider outliers.

2.1.2. Three steps procedure to consider the environmental variables

An assumption of the basic DEA model is that PCTs work in homogeneous environments. This assumption is not valid when performance is influenced by variables beyond the control of the PCTs managers. In the DEA, these variables are commonly called environmental variables. Understanding the influence of these variables is important to identifying genuine opportunities for managers to improve efficiency levels.

Different methods have been used to consider the influence of the environmental variables on efficiency. The most commonly used is to include the environmental variables together with the inputs and outcomes in the model. This method has two main problems. First, it is necessary to know the direction of the effect of each environmental variable a priori. Second, PCTs with worse environmental conditions will be considered more efficient. A second method that has been applied is a two-step procedure where the efficiency scores are regressed against the environmental variables in a second step. A criticism of this method is the fact that the range of the adjusted efficiencies is not distributed between zero and one, which hinders their interpretation. Some authors have overcome this problem by utilising a non-linear regression model in the second step. However, the adjusted efficiency scores are not easily linked to the production frontier estimated in the DEA, which hampers the possibility of estimating scale efficiencies that consider the effect of the environmental variables. Therefore, the feasibility of applying the three-stage procedure proposed by Fried et al. (2002) to estimate the efficiency of the PCTs while considering environmental variables is tested.

In the first stage, DEA is applied to health outcomes and inputs only, to obtain initial measures of the PCTs performance. In the second stage, stochastic frontier analysis (SFA) is used to regress the first stage performance measures against a set of environmental variables. This provides - for each input - a three-way decomposition of the variation in performance into a part attributable to environmental effects, a part attributable to inefficiency, and a part attributable to statistical noise. In the third stage, inputs are adjusted to account for the impact of the environmental effects and the statistical noise uncovered in the second stage, and DEA is used to re-evaluate PCTs efficiency. A more detail explanation of the three step procedure is presented in Appendix 2.

The outliers are identified and excluded by using the equation (1) during the first stage of the analysis. Therefore, in the first stage the DEA is estimated twice: (1) including all observations, and (2) excluding those observations classified as outliers.

All the estimations are conducted on the R.3.4.1 software where the Benchmarking package was selected to conduct the DEA and the SFA estimations.

2.2. Quantile Regression

Quantile Regression (QR) allows us to measure heterogeneity of the effects of health expenditure on mortality without the need for distributional assumptions. We use QR to explore differences across PCTs in returns to spend in terms of reducing mortality with a particular focus on understanding any differences between PCTs with low and high

mortality rates in each clinical area. This heterogeneity can be used to create a ranking or effect of health expenditure on mortality at PCT level. This ranking can be used to compare the effect of health expenditure on mortality with the benchmarking of PCTs according to DEA efficiency scores.

QR provides a more complete picture of covariate effects by estimating a family of conditional quantile functions, as set out by Koenker and Bassett (1978). These are models in which quantiles of the conditional distribution of the response variable are expressed as functions of observed covariates.

Our QR estimator is based on the following equation:

$$Q_h(\tau|n_i, x_i) = \alpha(\tau) + \beta(\tau)x_i + \gamma(\tau)n_i + w_i \quad (2)$$

where $Q_h(\tau|n_i, x_i)$ is the τ-*th* quantile on mortality rate *h*, conditional on health expenditure per head *x*, and *n* represents the need for healthcare in the PBC. The effect of health expenditure on mortality is measured by the coefficient β. This effect can be assessed at different points τ of the mortality distribution.

The unique feature of the QR is that it allows estimation of a slope coefficient for each point of the distribution of the dependent variable through the estimation of a family of quantile regressions. In our case, we can choose different percentiles, and also we can find 151 points of the mortality distribution representing the PCT mortality rate, where each quantile in equation (2) is defined as:

$$\tau_i = \frac{SYLLR\ rank\ for\ PCT(i)}{151} \quad (3)$$

τ has been rounded to three decimals, with $\tau_1 = 0.007$, and $\tau_{150} = 0.993$. For the PCT at the top of the mortality ranking $\tau_{151} = 0.9999$.

Besides providing a more complete picture, QR results in more robust estimations in the presence of non-normally distributed errors and outliers. QR preserves the conditional quantiles in transformations of the variables such as the logarithmic transformation, while the conditional mean changes with a non-linear transformation.

Classical linear regression estimates the conditional sample mean by minimising the sum of square residuals. The estimated effects of covariates are the same along the data distribution. If, instead of estimating the mean, we estimate a point at any part of the distribution, this is more likely to give us smaller absolute residuals at each part of the distribution. This should ideally be done without splitting the sample into different groups which entails sample selection problems. These different point estimates can therefore be obtained as conditional quantiles by QR. For the 0.5 quantile which separates the sample in two, half the observations are below the median mortality rate and half the observations are above median mortality. The parameters measuring the effects on the conditional median are obtained by minimising the sum of absolute residuals. Different quantiles can be obtained by minimising a sum of asymmetrically weighted absolute residuals, with the median (quantile 0.5) obtained by minimising the unweighted absolute value. For example, if an underestimate is marginally three times more costly than an overestimate, or if we are interested in the upper tail of the mortality distribution, we can estimate the 0.75 quantile which leaves 3/4 of observations below and 1/4 above the quantile. The point estimate of this conditional quantile can be obtained by minimising the sum of absolute residuals but penalising underpredictions more than overpredictions. The negative residuals or overpredictions have a weight of 0.25, with a weight of 0.75 for positive residuals or underpredictions. Analogously,

different conditional quantiles can be obtained for different weights, where the weight assigned to underpredictions represents the quantile.

The starting point for QR is identical to that of linear regression: with the specification of an outcome function linking mortality caused by a disease with the expenditure budgeted for this disease category (PBC expenditure) and including other covariates which are correlated to both PBC expenditure and mortality, such as different health needs due to different demographic composition and burden of disease, or other socioeconomic variables. We take the preferred specifications of the outcome model presented by Lomas, Martin and Claxton (2018) for six PBCs where available mortality data represent mortality for all the diseases in the PBC, as described in the next section.

The estimation method accounts for the potential endogeneity of PBC expenditure per head. This can occur because the level of spend per head reflects in part the level of health outcome achieved in the past (poorer health outcome areas may get more expenditure), so that there is a reverse causation. We implement an instrumental variables (IV) method in two stages. We take into account the outcome model specification for each PBC and the set of instrumental variables proposed by Claxton et al. (2015), which are socioeconomic variables reflecting deprivation and availability of informal care in the community.

First, we replicate the Lomas et al. (2018) outcome models by linear regression, including tests for the null hypotheses of exogeneity of PBC spend per head, and Hansen overidentification tests to validate the choice of instrumental variables. If the exogeneity test does not reject the exogeneity hypothesis, we estimate the conditional mean by OLS and the QR regression model using the PBC spend per head as explanatory variable with the rest of variables included in the outcome model. In the other case, if the endogeneity test results in rejection of the null hypotheses, we estimate the model by Generalized Method of Moments (GMM) estimator for the linear regression conditional mean model. Then, the QR model is implemented in two stages. In the first stage, PBC spend per head is predicted as a function of the instruments: the rest of the explanatory variables in the outcome model and the excluded instruments have been validated by the overidentification test in the GMM estimation. The second stage implements the QR model using predicted PBC spend per head and the rest of explanatory variables in the outcome model.

3. DATA AND DEFINITION OF VARIABLES

Between 2008 and 2012 the NHS in England was divided geographically into 151 PCTs, which in turn were grouped into 10 Strategic Health Authorities (SHAs). The PCTs were responsible for spending around 80% of the NHS England budget. Since the beginning of 2013 this structure was replaced with the creation of 213 Clinical Commissioning Groups (CCGs). Given the difference in the composition and percentage of budget handled between the PCTs and the CCGs the analysis presented here focuses on PCTs level.

Expenditure data for PCTs and for each clinical area are available from the Programme Budgeting Aggregate PCT Expenditure for all 23 programmes and subcategories for financial years 2008/09 to 2012/13.[2]

[2] Available at https://www.networks.nhs.uk/nhs-networks/health-investment-network/news/2012-13-programme-budgeting-data-is-now-available

Outcome data have been downloaded using the NHS Digital Indicator Portal (https://digital.nhs.uk/), which includes mortality statistics from ONS (Compendium of Population Health Indicators), outcome data from the CCG Outcomes Indicator Set and from the Quality Outcomes Framework, among other indicators related to employment and deprivation.

3.1. Mapping health outcomes to PCT level

We use health outcomes data recorded at district level local authority (LA) (326 LAs), or 152 top tier LAs, or at CCG level (211 CCGs). However, PBC expenditure data are available at PCT level for 151 PCTs for fiscal years 2008/09 to 2012/13. Then, we need health outcomes data at PCT level. Therefore, we need to map the health outcomes from different geographies (LA/top tier LA/CCG level) to PCT level. We use the mapping provided by the National Audit Office (NAO) (National Audit Office, 2015) based on the 2012 population to map outcomes from CCGs to PCTs. We have constructed a similar mapping based on population data from the Census 2011 to map outcomes from LAs and top tier LAs to PCTs.

3.1.1. Mapping method

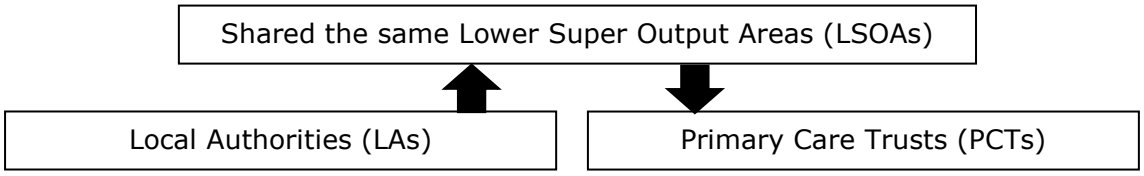
Mapping health outcomes based on the distribution of population across geographies implicitly assumes equal distribution of population health within a LA or CCG. So, we can use the same outcome values for the same LA/CCG in different PCTs.

We use calculating mortality of PCT based on mortality data of LAs as an example. The key problem is that not all LAs match to PCTs one on one. Some LAs might belong to a number of PCTs. Also some PCT might belong to a number of LAs. To solve the mapping problem we need to generate weights.

There are three steps involved in the mapping exercise.

Step 1: To break down each LA and each PCT to small geographic units, i.e. Lower Super Output Areas (LSOAs). LSOAs are small geographic areas in England. Each LA includes a number of LSOAs. And each PCT includes a number of LSOAs. We link LAs to PCTs by the common LSOAs they have. See Figure 3.

Figure 3. Mapping population between LA/Top tier LA/CCG to PCT



Step 2: From the population of each LSOA we can calculate the population in a LA and/or a PCT by adding the population from all relevant LSOAs together. More importantly we know the "structure" of the population for each PCT, i.e. how many people covered by a PCT comes from different LAs. The "structure" data are called the "weights". There are two types of weights that we could built up – depend on the types of outcome measures used.

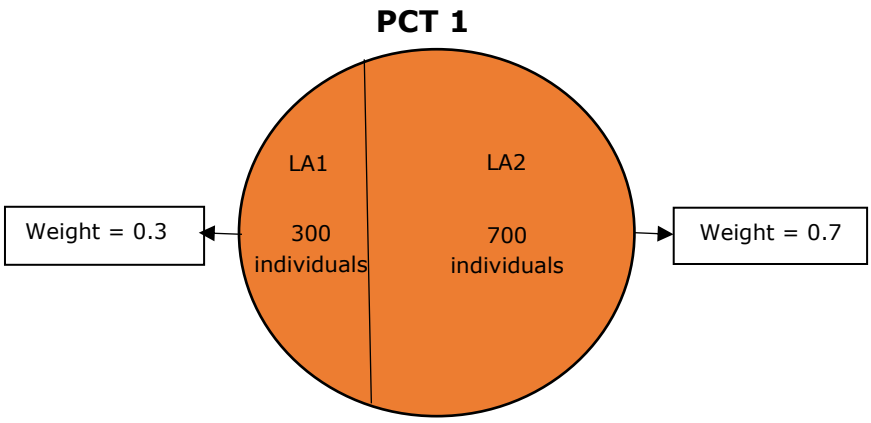
- Outcome measured by ratios: this is the format for most of the available outcome measures that we selected. For example, the mortality rate is recorded as a ratio

that represents the potential years of life lost per 10,000 population, assuming the population of England had the same population structure as the European Standard Population in 2013. The weights are measured by the proportion of the population in a PCT that comes from the different LAs.

See an example in

Figure **4**. The PCT1's population come from two LAs, i.e. LA1 with 300 people and LA2 with 700 people. The weight for LA1 is 0.3 (300/1000). The weight for LA2 is 0.7 (700/1000).

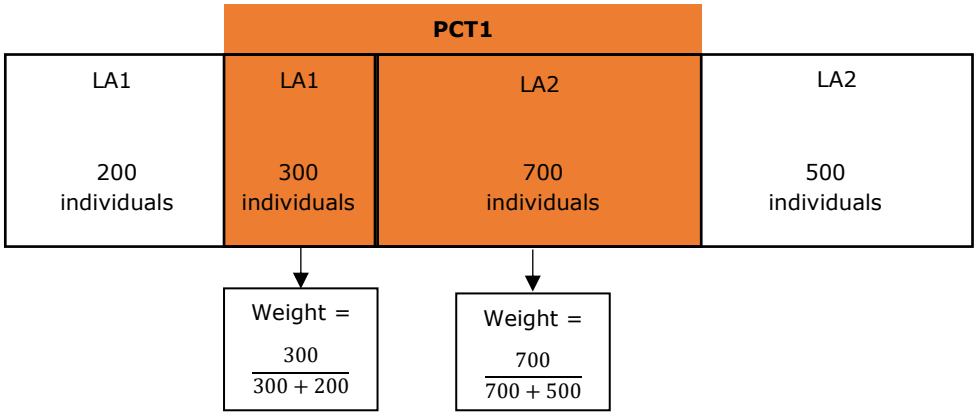
Figure 4. Weights for outcome measures in ratio



- Outcome measured by absolute numbers: this format appears for outcomes reflecting total number of patients affected. For example, the number of people died. The weights are measured by proportion of population from LAs that covered by a PCT.

  See an example in Figure 4. The PCT1's population come from two LAs, i.e. LA1 with 300 people and LA2 with 700 people. LA1 has 500 people in total and LA2 has 1200 people in total. The weight for LA1 is (0.6=300/500). The weight for LA2 is (0.58=700/1200).

Figure 5. Weights for outcome measures in absolute numbers



Step 3: We calculate the mortality of a PCT that has population from different LAs. The mortality of a PCT is calculated by the weighted average of the mortalities from those related LAs. The weights and populations are derived from Step 2. To continue the examples in

Figure 4 **and**

Figure 5.

- When the outcome is measured by ratio, e.g. mortality rate (
- Figure **4**)

Mortality rate for PCT1=0.3 X mortality rate for LA1 + 0.7 X mortality rate for LA2

When the outcome is measured by number of people who have died (

- Figure 5)

Number of people died in PCT1=0.6 X number of people died in LA1 + 0.58 X number of people died in LA2

3.1.2. Performance of the mapping

Our mapping from LAs to PCTs aggregates across population in common LSOAs which are smaller geographical units of about 1,500 people. However, since the original database we used to link across geographical units was based on smaller units (post codes), we found 37 LSOAs which each appear in two different PCTs. To avoid double counting the populations from those 37 LSOAs, we assume an equal split of the population from each of the 37 LSOAs to the two relevant PCTs.

We have compared the performance of our mapping methods from LA to PCT with results published by PHE (O'Conor, 2013). We obtain identical results when mapping LA population to PCT population except for small differences for 6 PCTs in the Birmingham area.

Also, we have mapped LAs to PCTs using the deprivation index IMD data from 2010. The average of original IMD data for 326 LAs gives a mean 19.15. The average of original IMD at PCT-level has mean 23.64 and standard deviation 8.41. Using our mapping weights to convert LAs relative outcomes to PCT outcomes, we obtain an identical mapped mean IMD at PCT level of 23.64, with a slightly smaller standard deviation of 8.35.

3.2. Selected variables per Programme Budget Category

Some statistics for the selected variables and the rationales for selection are summarised and explained below, depending on the method of analysis: DEA or QR.

3.2.1. Selection of health outcomes

The NHS priority domains defined in the Five Year Forward View have been organised in five dimensions which apply to different outcome frameworks: NHS Outcome Frameworks, CCG Outcome Frameworks, and Improvement Areas (NHS England, 2014; 2015). Most of the outcomes we use are included under these domains:

- Domain 1 - Preventing people from dying prematurely. This domain captures how successful the NHS is in reducing the number of avoidable deaths.

- Domain 2 - Enhancing quality of life for people with long-term conditions. This domain captures how successfully the NHS is supporting people with long-term conditions to live as normal a life as possible.
- Domain 3 - Helping people to recover from episodes of ill health or following injury. This domain captures how people recover from ill health or injury and wherever possible how it can be prevented.
- Domain 4 - Ensuring that people have a positive experience of care. This domain looks at the importance of providing a positive experience of care for patients, service users and carers.
- Domain 5 - Treating and caring for people in a safe environment and protecting them from avoidable harm. This domain explores patient safety and its importance in terms of quality of care to deliver better health outcomes.

Our choice of outcomes considers the availability of indicators under each of these domains, except Domain 5 which is not represented by the available outcomes.

Mortality data from NHS Compendium statistics (based on ONS registered deaths under 75 by local authority geographical area) are analysed for major disease categories: PBCs as defined by NHS England programme budgeting for estimating NHS expenditure across these programme categories covering the whole care pathway. Other mortality-related outcomes are also included such as excess mortality caused by mental health problems. Apart from mortality, DEA analyses other relevant outcomes in the clinical categories considered.

QR only considers standard mortality outcomes, as measured by mortality rates for each programme category. This allows to compare our results for different conditional quantile functions with linear regression (conditional mean function) as specified by the outcome models by Lomas et al. (2018). With this purpose, we select the same outcome variable: under 75 mortality measured as standardised years of life lost rate. These data are available at LA level from NHS Compendium mortality data and are presented pooled for three years: 2012, 2013, and 2014. Directly age-standardised rates are presented per 10,000 European Standard Population of 2013.

ONS Death Registry statistics are presented by ICD-10 classification of cause of death. However, the disaggregation by local authority provided by NHS Compendium statistics is not available for all causes of death in the ICD-10. We have selected 6 PBCs where the ICD-10 coverage is equal or cover large part of the diseases included in the PBC. For three PBCs, the coverage is almost the same, i.e. infectious diseases, cancer, and circulatory diseases. The PBC of respiratory diseases covers more diseases than the corresponding ICD-10 diseases with mortality statistics: asthma, COPD, and pneumonia. Claxton et al. (2015) calculate that these diseases only cover about 77% of mortality in the PBC of respiratory diseases. For the PBC of endocrine diseases, the only available mortality data is caused by diabetes and this covers about 63% of mortality in the PBC. Finally, the PBC of gastrointestinal diseases in represented by mortality caused by liver disease and ulcers, covering 57% of PBC mortality. Claxton et al. (2015) make an adjustment for this mismatch in the final calculation of the cost per life year by inflating mortality inversely to this coverage rate.

3.2.2. Health outcomes considered in the DEA models

Given the objectives and scope of this study, the three-stage procedure is applied to estimate the efficiency of the PCTs in seven out of the 23 PBCs.

The DEA analysis depends on the adequate selection of inputs and outcomes. As mentioned above, the health outcomes are selected considering the availability of indicators in four of the five NHS priority domains. In this section, the health outcomes included for each one of the seven PBCs included are defined.

Three main criteria were applied for the exclusion of health outcomes: (1) if there are an important number of missing data (over 20%); (2) if they are not considered final outcomes. All available health outcomes were analysed, but only those considered final outcomes of the healthcare sector were selected and included in the model. In addition, when two or more health outcomes showed coefficients of correlation higher than 0.50 only one of the health outcomes were included. The importance of each health outcome was tested by using Kolmogorov-Smirnov test. This test compares the efficiencies estimated when the health outcome *j* is included with the efficiency estimated when the outcome *j* is excluded. The selected health outcome corresponded to the one with the higher importance in the determination of the efficiency scores according to the Kolmogorov-Smirnov test. The rest of the section presents the included and excluded health outcomes by PBC.

3.2.2.1. Mental Health

- MH_Independently: Proportion of working age adults (18-69) who are receiving secondary mental health services and who are on the Care Programme Approach at the end of the month, who are recorded as living independently (with or without support) (%);
- MH_Employment: Proportion of working age adults (18-69) who are receiving secondary mental health services and who are on the Care Programme Approach at the end of the month who are recorded as being employed (%);
- MH_ExcessMort_2014_INV: Excess under 75 mortality rate in adults with serious mental illness. Standardised mortality ratio (SMR) expressed as a percentage based on general population and mental health population mortality rates, 95% confidence intervals (CI). In cases where the 2014 observation was missing for some reason, the most recent observation available was included[3].

The inverse of the variable is used and is estimated by applying:

$$\text{MH\_ExcessMort\_2014\_INV} = \frac{1}{\text{MH\_ExcessMort\_2014}} * 100 \qquad (3.1)$$

Higher values are "better" than smaller values in estimating the DEA model.

- MH_HRQoL_ 2014: Health-related quality of life for people with a long-term mental health condition. Directly standardised average health-status (EQ-5DTM)

[3] Mortality data for previous years were used for the following 17 PCTs for which 2014 data was not available: North Lincolnshire - 2011, Bassetlaw -2010, South Birmingham – 2010, Heart of Birmingham Teaching – 2011, Derbyshire County – 2010, North Lancashire – 2011, Central Lancashire – 2010, Western Cheshire – 2010, Hastings and Rother – 2010, North Staffordshire – 2010, South Staffordshire – 2011, Suffolk – 2010, West Essex – 2012, North East Essex – 2010, Mid Essex – 2011, Eastern and Coastal Kent – 2010, and Oxfordshire – 2010.

score for individuals reporting that they have a long-term mental health condition.

Two possible health outcomes were excluded, as these are considered intermediate outcomes:

- IAPTRec_2014: The percentage of referrals to Improving Access to Psychological Therapies (IAPT) services with a finished course of treatment who were initially at caseness which indicated a reliable improvement, presented with 95% confidence intervals (CI);
- IAPTImp_2014: The percentage of referrals to Improving Access to Psychological Therapies (IAPT) services with a finished course of treatment which indicated a reliable improvement, presented with 95% confidence intervals (CI).

3.2.2.2. Maternity

- NeonatalMort_2014_INV: Neonatal mortality and stillbirths. Directly age-standardised rates. The inverse of the variable is used, and is estimated similarly to equation 3.1;
- MAT01_Point_2012: Maternity Services Quality and Outcomes Framework (QOF) for April 2009 - March 2010.

One possible health outcome was excluded since 45% of the observations are missed:

- MaternalMort_2014: Maternal mortality (ICD10 O00-O99). Directly age-standardised rates per 100,000 European Standard population (15 to 44 years). Excluded**.**

3.2.2.3. Cancer

- SYLLR_Cancer_2014_INV: Years of life lost due to mortality from all cancers (ICD10 C00-C97). Directly age-standardised rates (DSR). The inverse of the variable is estimated as in equation 3.1;
- OneYSurv_2014: One-year net survival for adults diagnosed with cancer (aged 15 - 99 years), 95% confidence intervals. If the 2014 observation was missed, the most recent observation available was included.

Two possible health outcomes were excluded, the first one was considered an intermediate outcomes and the second one is highly correlated with SYLLR_Cancer_2014_INV:

- EarlyCancerDetection_2014: Percentage of cancers detected at stage 1 and 2. In cases when the 2014 observation was missing, the most recent observation available were included;
- DSR_Cancer_75_2014: Mortality from all cancers (ICD9 140-208 adjusted, ICD10 C00-C97). Directly age-standardised rates (DSR). Less than 75 years old.

3.2.2.4. Gastrointestinal

- SYLLR_Gastro_2014_INV: Years of life lost due to mortality from: (1) gastric, duodenal and peptic ulcers (ICD10 K25-K27); and (2) chronic liver disease including cirrhosis (ICD10 K70, K73-K74). Inverse of the Directly age-standardised rates (DSR) is estimated as in equation 3.1;

- AlcoholLiverEmerg_2014_INV: Directly age and sex standardised rate of emergency admissions for alcohol related liver disease in adults aged 19 years and older, per 100,000 registered patients, 95% confidence intervals (CI). The inverse of the variable is estimated as in equation 3.1. In cases when the 2014 observation was missing, the most recent observation available was included.

3.2.2.5. Cardiovascular

- SYLLR_CVD_2014_INV: Years of life lost due to mortality from all circulatory diseases (ICD10 I00-I99). Directly age-standardised rates (DSR). The inverse of the variable is estimated as in equation 3.1;
- CardiacRehab_2014: Proportion of referrals to a cardiac rehabilitation programme that were recorded as completed within 365 days of the start of an associated hospital admission, expressed as a percentage. In cases when the 2014 observation was missing, the most recent observation available was included;
- Stroke_discharge_2014: People with stroke who are discharged from hospital with a joint health and social care plan. In cases when the 2014 observation was missing, the most recent observation available was included.

Three possible health outcomes were excluded. These are considered intermediate outcomes:

- Stroke4hours_2014: People with stroke admitted to an acute stroke unit within 4 hours of arrival to hospital;
- Thrombolysis_2014: People who have had an acute stroke who receive thrombolysis;
- StrokeUnit_2014: People who have had an acute stroke who spend 90% or more of their stay on a stroke unit.

3.2.2.6. Respiratory

- SYLLR_Respiratory_2014_INV: Years of life lost due to mortality from: bronchitis, emphysema and other COPD (ICD10 J40-J44); asthma (ICD10 J45-J46); and pneumonia (ICD10 J12-J18). Directly age-standardised rates (DSR). The inverse of the variable is estimated as in equation 3.1;
- EmergencyRespiratoryChild_2014_INV: Directly age and sex standardised admission rate for emergency admissions for children aged 18 years and under with lower respiratory tract infections per 100,000 registered patients. The inverse of the variable is used. In cases when the 2014 observation was missing, the most recent observation available was included.

One possible health outcome was excluded, as it is considered an intermediate outcome:

- COPDRehab_2014: The percentage of people with Chronic Obstructive Pulmonary Disease (COPD) and Medical Research Council (MRC) Dyspnoea Scale >=3, identified on GP systems, referred to a pulmonary rehabilitation programme.

3.2.2.7. Endocrine

- SYLLR_Endocrine_2014_INV: Years of life lost due to mortality from diabetes (ICD10 E10-E14). Directly age-standardised rates (DSR). The inverse of the is estimated as in equation 3.1;

- DiabComplications_2014_INV: Indirectly age and sex standardised ratio of complications in people with diabetes. The inverse of the variable is used. In case that the 2014 observation was missed, the most recent observation available was included. According to NHS Digital Indicator Portal, participation rates vary widely between PCTs suggesting the need for caution on the interpretation of this variable.

Two possible health outcomes were excluded, as these are considered intermediate outcomes:

- Eigth_care_processes_2015: People with diabetes who have received nine care processes - Indicator values are currently based on eight care processes as data for eye screening are not available. Caution should be used when interpreting indicator values as GP practice participation rates vary widely and the indicator values for some CCGs are derived from a small number of GP practices;

StructEducation_2014: The percentage of people with diabetes diagnosed less than one year who are referred to structured education.

Table 1 shows the main descriptive statistics that describe the outcome variables included in the DEA for each PBC.

Table 1. Descriptive Statistics – Outcome measured included

	Median	Mean	Min.	1st Qu.	3rd Qu.	Max.	NA	Std Dev.
MentalHealth								
MH_Independently	63.8	60.6	1.6	50.9	73.5	92.6	0.0	19.3
MH_Employment	6.0	6.6	0.1	4.2	8.2	19.9	2.0	3.4
MH_ExcessMort_2014_INV	0.29	0.3	0.17	0.25	0.33	1.1	0.0	0.1
MH_HRQoL_2014	0.5	0.5	0.3	0.5	0.6	0.6	0.0	0.1
Maternity								
NeonatalMort_2014_INV	14.8	15.4	7.7	12.3	17.2	30.9	1.0	4.5
MAT01_Point_2012	288.0	319.0	72.0	210.0	390.0	870.0	0.0	153.0
Cancer								
SYLLR_Cancer_2014_INV	0.63	0.63	0.48	0.57	0.68	0.80	0.0	0.07
OneYSurv_2014	68.7	67.1	17.6	66.9	70.0	73.6	0.0	6.8
Gastrointestinal								
SYLLR_Gastro_2014_INV	4.5	4.7	1.6	3.3	6.0	9.8	0.0	1.7
AlcoholLiverEmerg_2014_INV	4.4	5.0	1.7	3.0	5.9	32.9	0.0	3.7
Cardiovascular								
SYLLR_CVD_2014_INV	1.2	1.2	0.7	1.0	1.3	2.0	0.0	0.3
CardiacRehab_2014	0.3	0.3	0.0	0.2	0.4	0.7	23.0	0.2
Stroke_discharge_2014	0.8	0.7	0.0	0.5	0.9	1.0	2.0	0.2
Respiratory								
SYLLR_Respiratory_2014_INV	4.2	4.4	1.7	3.2	5.3	8.1	0.0	1.4
Emergency Respiratory Child_2014_INV	0.3	0.3	0.2	0.2	0.4	1.3	1.0	0.2
Endocrine								
SYLLR_Endocrine_2014_INV	25.1	28.0	10.8	19.5	33.2	133.3	0.0	14.0

DiabComplications_2014_INV	1.0	1.1	0.6	0.9	1.2	6.6	0.0	0.5

Source: NHS Indicator Portal (NHS Digital).

3.2.3. Inputs included in the DEA models

Two assumptions are made in the selection of the inputs. First, the health outcomes considered are depending not only on the expenditures of one year, but on the expenditures of the past three years. Therefore, the per-capita expenditures of 2011, 2012 and 2013 in each particular PBC are considered are inputs.

Second, the model assumes that three categories of expenditures affect health outcomes in all other PBCs:

- Healthy individuals (PBC 21): Activities related to prevention and health promotion are included in this category (e.g. all weight management programmes).
- Social care needs (PBC22): Activities such as community care are included here as well as costs related to particular care needs, such as the cost of designated professionals (doctors and nurses) for safeguarding children.
- Other areas of spend (PBC 23): A number of miscellaneous which are not possible to include in any other PBC are classified here.

The sum of the per-capita expenditures per year in these three PBCs was calculated. Three years, 2011, 2012 and 2013, were considered are inputs.

The analysis uses per-capita expenditure adjusted by the differences in health care demand across PCTs using the Need Index and differences in price level using the Market Factor Index (Department of Health (DH), 2011). Expenditures per year was calculated by dividing the total expenditure on own population in a particular year by the PCT Unified Weighted population of that year. These are defined as:

- Expenditure on own population: Net expenditure adjusted to add back expenditure funded from sources outside of the NHS and to deduct expenditure on other PCT populations incurred through lead commissioning arrangements.
- Unified Weighted population: The PCT responsible population adjusted using the national weighted capitation formula, for the age structure of the population, its additional need over and above that accounted for by age (Need Index), and the unavoidable geographical variations in the costs of providing services (Market Factor Index).

In Table 2 descriptive statistics for the included inputs are presented.

Table 2. Per-Capita expenditures per PBC (adjusted by the differences in health care demand and differences in price level)

	Median	Mean	Min.	1st Qu.	3rd Qu.	Max.	Std Dev.
Mental Health							
Per-capita Exp. 2011	203	214	48	182	230	447	56.2
Per-capita Exp. 2012	207.0	215.0	121.0	184.0	232.0	409.0	47.9
Per-capita Exp. 2013	209.0	217.0	143.0	188.0	236.0	412.0	46.1
Maternity							
Per-capita Exp. 2011	64.1	69.9	32.2	54.3	79.8	167.9	22.1
Per-capita Exp. 2012	65.0	69.3	34.7	56.0	77.0	168.8	20.2

	Median	Mean	Min.	1st Qu.	3rd Qu.	Max.	Std Dev.
Per-capita Exp. 2013	62.1	66.7	35.1	54.1	76.0	162.8	20.1
Cancer*							
Per-capita Exp. 2011	103.7	104.3	60.6	89.7	116.4	193.3	20.6
Per-capita Exp. 2012	104.2	104.2	55.3	90.5	116.3	161.8	17.4
Per-capita Exp. 2013	105.7	106.4	49.4	91.2	118.6	165.5	20.1
Gastrointestinal							
Per-capita Exp. 2011	84.7	84.3	34.8	77.5	92.5	140.0	14.4
Per-capita Exp. 2012	86.6	87.3	56.2	80.5	94.3	118.7	11.1
Per-capita Exp. 2013	89.2	89.4	60.1	82.2	95.8	117.7	11.1
Cardiovascular*							
Per-capita Exp. 2011	130.9	132.4	87.8	119.7	146.0	215.2	20.5
Per-capita Exp. 2012	130.7	130.5	86.6	118.0	143.5	168.6	17.3
Per-capita Exp. 2013	126.9	128.3	82.5	115.2	140.8	175.0	18.3
Respiratory							
Per-capita Exp. 2011	82.4	82.5	48.9	75.6	88.7	123.0	11.7
Per-capita Exp. 2012	83.7	84.4	55.7	78.1	88.7	125.2	10.2
Per-capita Exp. 2013	89.4	89.1	55.7	81.9	94.8	121.9	10.7
Endocrine							
Per-capita Exp. 2011	53.2	54.3	38.4	48.6	59.1	86.1	8.4
Per-capita Exp. 2012	54.4	56.0	42.2	51.1	59.7	88.1	7.7
Per-capita Exp. 2013	56.9	57.7	40.7	52.5	62.1	79.6	7.2
Additional expenditures: Related to all clinical areas							
Healthy individuals (PBC 21) + Social care needs (PBC22) + Other areas of spend (PBC 23)							
Per-capita Exp. 2011	383.0	403.0	272.0	343.0	426.0	809.0	103.1
Per-capita Exp. 2012	399.0	408.0	290.0	360.0	436.0	783.0	77.2
Per-capita Exp. 2013	415.0	425.0	269.0	374.0	463.0	776.0	75.8
Healthy individuals (PBC 21) + Social care needs (PBC22)*							
Per-capita Exp. 2011	83.6	98.1	14.7	61.5	108.7	532.8	74.4
Per-capita Exp. 2012	92.1	100.1	1.6	67.3	119.9	469.0	60.4
Per-capita Exp. 2013	101.4	105.1	15.6	73.2	126.5	463.6	53.5
Healthy individuals (PBC 21)*							
Per-capita Exp. 2011	41.73	42.56	0.03	30.06	53.58	122.32	21.70
Per-capita Exp. 2012	39.56	39.46	0.17	28.92	49.85	88.42	16.11
Per-capita Exp. 2013	35.50	36.07	2.22	26.47	47.31	99.22	16.92

* The DEA estimated for Cancer includes as additional expenditures only the sum of the expenditures in PBC 21 and PBC 22.
** The DEA estimated for Cardiovascular included as additional expenditures only the sum of the expenditures in PBC 21.
Source: Based on data from the "Exposition book" elaborated by the Department of Health (DH, 2013).

3.2.4. Environmental variables included in the DEA models

In order to understand the influence of those factors that affect the ability of the PCTs to achieve higher levels of health outcomes, but that cannot be changed by the PCTs' managers, two main environmental variables were selected. First, the socioeconomic situation captured by the deprivation index, and second, the financial restriction faced by the PCTs reflected in the difference between the budget needed to fulfil health needs and the actual budget allocated by the NHS on each particular PCT.

- Deprivation Index: The English Index of Deprivation measure relative levels of deprivation in more than 32 000 LSOAs in England. Seven domains of deprivation were considered in the 2010 estimation of the Deprivation Index: (1) income deprivation, (2) employment deprivation, (3) health deprivation and disability domain, (4) education, skills and training deprivation, (5) barriers to housing and services, (6) crime domain, and (7) living environment deprivation. Two different representations of the deprivation index were included.
- IMD_2010_Standardized: Population weighted average of the combined scores for the LSOAs in a PCT. This measure is calculated by averaging the LSOA scores in each PCT after they have been population weighted. The standardized value of the original variable is used.

$$\text{IMD\_2010\_Standardized} = \frac{\text{IMD\_2010} - \text{Averague (IMD\_2010)}}{\sqrt{sd(\text{IMD\_2010})}} \quad (3.2)$$

- Extent_2010: Proportion of a PCT's population living in the most deprived LSOAs in the country. In this measure, 100% of the people living in the 10% most deprived LSOAs in England are captured in the numerator, plus a proportion of the population of those LSOAs in the next two deciles on a sliding scale – that is 95% of the population of the LSOA at the 11th percentile, and 5% of the population of the LSOA at the 29th percentile.

Given that IMD_2010 retains the fact that more deprived LSOAs may have more 'extreme' scores and is not affected by the number of inhabitants, this variable is the one included in most of the DEA models as an approximation of the socioeconomic level. Nevertheless, in those cases in which IMD_2010 is not statistically significant, Extent_2010 is included.

- TargetDistance_2010: The NHS classifies the PCTs according to the budget allocated. The budget that should be allocated to a PCT is estimated based on the Need Index which reflects the health care demand of the PCT. It can called the `required budget´. However, the allocated budget is only adjusted gradually every year toward the required budget. These adjustments can be positive or negative depending on whether the required budget is higher or lower than the budget allocated. Those locations in which the budget allocated is higher than the predicted budget required are grouped as 'over target' locations. Differences between the allocated budget and the required budget are considered in the analysis as an environmental variable and expressed in percentage terms.

Descriptive statistics are set out in Table 3.

Table 3. Descriptive Statistics – Environmental Variables

	Median	Mean	Min.	1st Qu.	3rd Qu.	Max.	Std Dev.
Deprivation Index							
Extent_2010	0.22	0.24	0.00	0.09	0.36	0.77	0.18
IMD_2010_ Standardized	-0.11	0.00	-5.12	-2.42	2.01	7.47	2.90
TargetDistance_2010-2011	-0.19	0.00	-1.15	-0.65	0.31	3.95	1.00

Source: Based on English indices of deprivation 2010 (https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010)

3.2.5. Quantile Regression variables

3.2.5.1. Dependent variables in Quantile Regression

The dependent variable is the mortality rate for the six selected PBCs. The mortality rate is measured by Standard Years of Life Lost rate average for the years 2012/13/14 (SYLLR 2012/13/14). The SYLLR for the pooled time period is the average of the individually calculated annual SYLLR. Since DEA analysis has also used this outcome variable for 5 PBCs, descriptive statistics are presented in Table 4, using a different metric, the inverse.

Table 4. Descriptive Statistics – Outcomes included in QR

	SYLLR 2012/13/14						
	Median	**Mean**	**Min.**	**1st Qu.**	**3rd Qu.**	**Max.**	**Std Dev.**
Infectious	5.57	6.37	2.41	4.45	8.1	14.45	2.57
Cancer	159.77	162.12	125.11	146.84	175.13	207.91	19.52
Endocrine	3.98	4.20	0.75	3	5.17	9.27	1.60
Cardiovascular	86.725	88.91	50.2	74.87	102.23	141.77	18.73
Respiratory	23.67	25.56	12.33	18.88	31.14	57.47	8.73
Gastrointestinal	22.43	24.40	10.19	16.55	30.11	61.8	9.54

3.2.5.2. Exogenous variables in Quantile regression

In the outcome model specifications for mortality outcomes, it is assumed that the only endogenous variable linked to mortality is PBC spend per head. Some models include as explanatory variable a measure of health needs which depend on age and on burden of disease, constructed by the DH in the Resource Allocation Weighted Capitation Formula used to distribute the health budget across PCTs. Arguably, this explanatory variable measuring burden of disease could be potentially endogenous, but health needs are only considered potentially endogenous by Claxton et al. (2015) and Lomas et al. (2018) in their expenditure models, where these needs are represented by mortality excluding that of the considered PBC. Therefore, following the specification of the preferred outcome models presented by Lomas et al. (2018), all explanatory variables used in our outcome models, except PBC spend per head, are considered exogenous.

3.2.5.3. Instrumental variables in Quantile Regression

The endogeneity of the PBC spend explanatory variable produces a bias in the estimated coefficient due to the correlation of PBC spend and the random error term in the outcome equation. For the case of the English health spending data, Martin et al. (2008) discuss the endogeneity of health expenditure due to the influence of past health outcomes beside correlations with unobservable heterogeneity and measurements errors of the mortality outcome. To correct for the bias caused by this endogeneity, we use an instrumental variable (IV) estimation method which requires some exogenous variables (instruments) which only affect mortality through their effect on health expenditure so that the instruments are uncorrelated with the error term in the outcome equation (exclusion restriction). To validate the exclusion restriction, the system requires more than one instrument to test the overidentification restriction (Hansen-Sargan test). Moreover, the instruments must be good predictors (not weak instruments) of health expenditure.

We have tried different instruments from the set of instruments proposed by Claxton et al. (2015) in Table 92, page 347, but we measure them in the 2011 Population Census, not the 2001 Census. For the index of multiple deprivation, we take 2010 data. Each outcome model is estimated with at least two excluded instruments to guarantee overidentification and test their exclusion validity.

The explanatory variables included in the different outcome models for the six PBCs are set out in Table 5, and the descriptive statistics included in the QR are in Table 6.

Table 5. Explanatory variables/ instruments included in the different outcome models for the six PBCs

Health needs	
CARANneed	Combining Age-Related and Additional Needs for acute services 2011/12
CARANneed1213	Combining Age-Related and Additional Needs for acute services 2012/13 Lomas et al. (2018)
HIVneedprev	HIV need for prevention 2011/12
HIVneed	HIV need for prevention 2012/13 (used by Lomas, Martin and Claxton (2018))
Deprivation variables	
IMD2010	Average deprivation
IncomeScale	income scale in the IMD 2010 index
EmploymentScale	employment scale in the IMD 2010 index
Socioeconomic variables from Population Census 2011	
BORNEXEU	Residents born outside the EU divided by all residents
WHITEEG	Population in white ethnic group divided by total population
POPALLTI	Proportion of population with LTI/disability
POP16_64LTI	Proportion of population of working age with LTI/disability aged 16–74 years
POPPUCAR	Proportion of population providing unpaid care
POPPUCA1	Proportion of population providing unpaid care for 1–19 hours a week
POPPUCA2	Proportion of population providing unpaid care for 20–49 hours per week
POPPUCA3	Proportion of population providing unpaid care for > 50 hours a week
NQUALall	Proportion population years with no qualifications
FTSTUDEN	Proportion of population aged 16–74 years that are full-time students
HHNOCAR	Proportion of households without a car
OWNOCC	Proportion of households that are owner occupied
LAHARENT	Proportion of households that are rented from LA or HA
PRIVRENT	Proportion of households that are rented from private landlords
LONEPENH	Proportion of households that are one pensioner households
LONE65andover	Proportion of households that are one person 65 and over households
LONEPARH	Proportion of households that are lone parent households with dependent children
PC74LTUN	Proportion of those aged 16–74 years that are long-term unemployed
PROFOCCU	Proportion of those aged 16–74 years in managerial and professional occupations

Table 6. Descriptive statistics – Explanatory variables/instruments included in the QR

Variable	Mean	Std. Dev	Min	Max
Spend per head 212/13 in each PBC adjusted by need (see also Table 2)				
g1_1213net	30.847	17.508	12.153	99.147
g2_1213net	106.373	20.122	49.362	165.459
g4_1213net	57.670	7.228	40.675	79.614
g10_1213net	128.256	18.288	82.509	174.967
g11_1213net	89.076	10.665	55.743	121.891
g13_1213net	89.412	11.121	60.088	117.656
Deprivation variables				
IncomeScale	49791.7	22864.0	14110.0	122060.0
EmploymentScale	19902.7	9332.73	5000.0	54350.0
IMD2010	23.640	8.408	8.809	45.310
Health Need variables				
CARANneed	1.025	0.129	0.727	1.354
HIV prev Index	1.080	0.666	0.564	4.098
Socioeconomic variables				
OWNOCC	0.615	0.116	0.242	0.754
LAHRENT	0.188	0.073	0.081	0.437
PRIVRENT	0.162	0.060	0.084	0.376
NQUALall	0.230	0.051	0.101	0.352
PROFOCCU	0.305	0.069	0.181	0.547
FTSTUDENTS	0.092	0.036	0.054	0.222
LONE65andover	0.122	0.021	0.060	0.167
LONEPARH	0.075	0.017	0.047	0.144
POPPUCAR	0.102	0.014	0.065	0.126
POPPUCAR1	0.063	0.009	0.043	0.081
POPPUCAR2	0.014	0.003	0.009	0.022
POPPUCAR3	0.024	0.006	0.012	0.040
POPALLLTI	0.180	0.032	0.112	0.256
POP16_64LTI	0.133	0.027	0.076	0.206
HHNOCAR	0.284	0.118	0.126	0.648
BORNEXEU	0.103	0.100	0.012	0.424
WHITEEG	0.837	0.166	0.290	0.985
PC74LTUN	0.019	0.006	0.010	0.037
FTSTUDEN	0.094	0.037	0.056	0.226

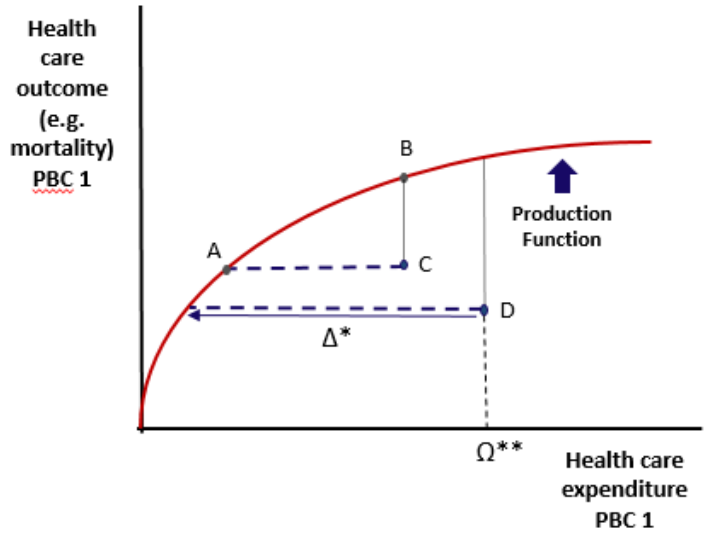
4. RESULTS

4.1. DEA

The results derived from the DEA analysis can be consider from two points of view: (1) by analysing the efficiency scores; or (2) by analysing the decrease in expenditure per year that could be possible without affecting health outcomes. The latter indicates how much could in principle be spent on new treatments without reducing current health outcomes achieved. We illustrate these approaches in Figure 6 where A and B represent efficient PCTs. It is not possible for these PCTs to spend less without reducing the health outcome level. PCTs C and D are inefficient, for them it would be possible, by reorganising their production of health, to achieve the same health outcome level with lower expenditure. For PCTs C and D the opportunity cost of funding a new health technology, in terms of the health outcome, could be zero in terms of its impact on current health outcomes, if they improve efficiency. In the case of the PCT D, Δ* indicates the proportional reduction of inputs needed by an inefficient PCTs to be efficient while its outcomes are held constant. In other words, Δ* represents the decrease in expenditure per year that could be possible without affecting the health outcome level. Since efficient PCTs cannot change expenditures without affecting outcomes, the efficiency scores define the decrease in expenditure per year that could be possible without affecting health outcomes.

By dividing Δ* by Ω**, we obtain the ***percentage*** decrease in expenditures per year that could be possible without affecting health outcomes. This measure will be analysed in this chapter.

Figure 6. DEA and the estimation of the possible decrease in expenditures per year that could be possible without affecting health outcomes



*Δ = decrease in the health care expenditures per year of PCT D in PCT 1 that could be possible without affecting health outcomes in PBC 1. **Ω = total annual health expenditure of PCT D in PBC 1

Additionally, we present a comparison of the results estimated in the first stage of the DEA, where the inputs included are not adjusted by the effect of the environmental variables (see Appendix 3), with the results of the third stage, where the inputs are adjusted by the effect of the two selected exogenous variables. This allows us to compare two possibilities. First, when the environmental variables are not included we

are assuming that all PCTs have similar conditions, meaning homogeneous environments, in which to achieve the maximum level of health outcomes given health expenditures. Second, we consider environmental variables that can affect health outcomes and over which PCT managers have no control. In this case, we are acknowledging that PCTs are operating under different conditions, and performance is influenced by variables beyond managerial control. Therefore, the inclusion of the environmental variables allows us to "leveling the playing field" such that two PCTs, one with better conditions than the other (in terms of the environmental variables selected) could be classified as efficient even if with the same level of expenditure one produce worse health outcomes than the other.

The Deprivation Index is the first of two environmental variables considered. The health outcomes of PCTs with more economically deprived populations are expected to be consistently poorer in comparison with those of PCTs in less deprived areas. A number of factors are considered to be possible causes of inequalities. These factors can be grouped into: supply, demand and environmental factors. Environmental factors are factors outside of the range on which health policies could have an effect and which can be expected therefore to impact efficiency (Baade et al., 2016; Haynes, Pearce and Barnett, 2008). Recent literature has tested the role of the demand factors in explaining inequalities finding that socioeconomic differences are translated into differences in health outcomes (Macinko et al., 2003; Pickett and Wilkinson, 2015). Evidence suggests that socioeconomic inequalities exist in incidence, prevalence and survival rates (Brenner et al., 1991; Jansen et al., 2014). In recent years, a number of studies have been conducted showing an association between regional deprivation and incidence and mortality (Kuznetsov et al., 2012).

The second environmental variable is Target Distance, which indicates whether a PCT is underfunded or overfunded depending on the budget allocated by the NHS in comparison to the `required budget´ estimated based on the health needs of the population as defined by the NHS allocation formula. The allocated budget is only adjusted gradually every year toward the required budget, this means that a PCT whose actual budget is lower than the required budget (an underfunding PCT) will face a deficit of resources for more than one period. This could be limiting its capacity to improve the health outcomes of the population. At the other extreme, an overfunded PCT will have resources over those required to fulfil the health need of the population. For this group there will be a surplus of resources (difference between actual budget and required budget) that can be designated to, for example additional healthcare programs (e.g. preventive programs). This could lead to improvements in health outcomes in the overfunded PCT that are not related to its efficiency in using heath care resources. The distance between the actual budget and the required budget is considered outside of the control of the PCT's managers since it is defined by the NHS. An underfunded PCT could be operated efficiently but not have the required resources to achieve the same level of health outcomes as another overfunded PCT which could be operating inefficiently.

A number of health outcomes are considered when PCT leaders are deciding on the allocation of their resources. The DEA method allows us to account for expenditure achieving multiple health outcomes. One our objectives in using DEA is to test the feasibility of including more than one health outcome by using the DEA. For seven out of the eight included PBCs for which mortality data was available, we identified at least one additional health outcome whose information is publicly available and is included among the NHS priority domains. The exception was infectious diseases for which the only

indicator identified is years of life lost due to mortality from infectious and parasitic disease (ICD10 A00-B99). Therefore, for the PBC infectious diseases the DEA was not estimated[4].

The main objective of applying the DEA analysis in this study is to analyse whether the PCTs have some leeway to decrease expenditures in certain PBCs without affecting health outcomes. This would imply that funding a new health technology by reallocating expenditures from a particular PBC does not always require a displacement of activity with a related loss of health. If displacement is not the only option, an opportunity cost estimated on the basis that it is would be greater than the actual opportunity cost. The results of the DEA show that, in each of the PBCs we examined there exists a group of PCTs with the possibility to fund new health treatments through an improvement in efficiency rather than displacing other activity.

The results suggest that there are differences across PCTs in the level of efficiency in each PBC. It is not generally the case that PCTs are "efficient" or "inefficient" in all PBCs. Most are efficient in some and less efficient in others. This is likely to influence the mix of services PCTs choose to invest or disinvest in at the margin when they look at funding a new technology. Appendix 3 shows in which PBCs particular PCTs are fully efficient (efficiency score = 1). For most PCTs the efficiency score varies across PBCs. This indicates that there are health areas in which the PCT has a higher margin for increasing efficiency, i.e. there are particular PBCs for which there is leeway to adjust expenditures without affecting outcomes. This could affect the decision regarding the reallocation of resources to respond to a mandatory inclusion of a new health treatment.

Out of the 151 PCTs included, the estimation of the efficiency scores for all seven PBCs was possible for 101 PCTs, given the information available (see Table 12, Appendix 3). Two out of the 101 were fully efficient in all PBCs analysed: North Yorkshire and York, and South East Essex. In addition, four PCTs were fully efficient in six of the seven PBCs estimated: Bury, Brighton and Hove City, Suffolk and Hertfordshire. For the remaining 50 PCTs, the estimation of the efficiency score of at least one PBC was not possible because of missing data. Among this group Bolton (six efficiency scores estimated) and Hampshire (five efficiency scores estimated) were in every case fully efficient.

The DEA shows differences in opportunity cost across PCTs and across PBCs. First, the results indicate that in each PBC there is a group of efficient PCTs located in the increasing returns to scale part of the productivity curve and that this group varies by PBC. This means that for a particular PBC there will be a subset of PCTs (the increasing return to scale group) for which a decrease of 1% in expenditures will negatively affect health outcomes by more than 1%. The findings also show that an additional group of efficient PCTs (the decreasing or constant return to scale group) for which a decrease of expenditures by 1% will have a negative effect on health outcomes that is lower or equal to 1%. Third, the analysis also identifies a group of PCTs (the inefficient group) for which a decrease in resource will have a zero or 0% effect on health outcomes. The PCTs that comprise each one of these three groups varies across PBCs[5]. These differences in opportunity cost across PCTs and PBCs mean that estimating a threshold reflecting the

[4] It was included in the quantile regression for which only mortality data was needed.

[5] In a previous report Hernandez-Villafuerte, Zamora and Towse, (2018) presented the results of a cluster analysis suggesting that the PCTs can be divided into eight groups which appear to have different production functions. The results of the cluster analysis were estimated based only on the distribution of health expenditures among PBCs. We recommend further analysis to re-estimate the cluster distribution of the English health areas (PCTs or CCGs) considering not only health care expenditures, but also health care outcomes.

average displacement across all PCTs cannot take account of the likelihood that in a national health system there are variations in budget and different technical constraints between health locations. Moreover, if the policy makers are considering opportunity cost in the allocation of resources, it might be expected that PCTs will displace services on those clinical areas with lower health losses, meaning areas with zero (inefficient PBCs) or lower opportunity cost (PBCs with decreasing or constant return to scale).

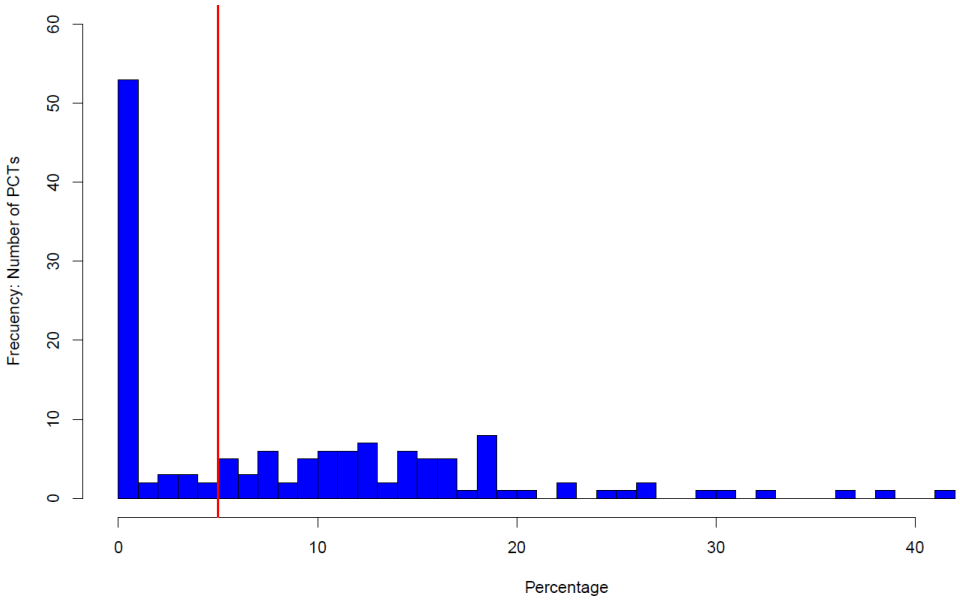
4.1.1. Mental Health

Figure 7 shows the results without taking account of the impact of the environmental variables. Figure 8 shows the results when an adjustment is made for the influence of environmental variables. In both cases, although a high number of PCTs are efficient or close to being efficient, there is the possibility to decrease expenditure without affecting outcomes.

When the impact of the environmental variables is not taken account of (Figure 7) there are PCTs that can decrease their expenditures by over 30% without affecting outcomes. When adjustment is made, i.e. if we consider that PCTs do not influence the socioeconomic level or the resource allocation to them, the maximum adjustment that can be made is of around 25% (Figure 8). Adjusting for these environmental factors means that the number of PCTs which can adjust their expenditures by more than 5% decreases, although the number of fully efficient PCTs only increases to 52

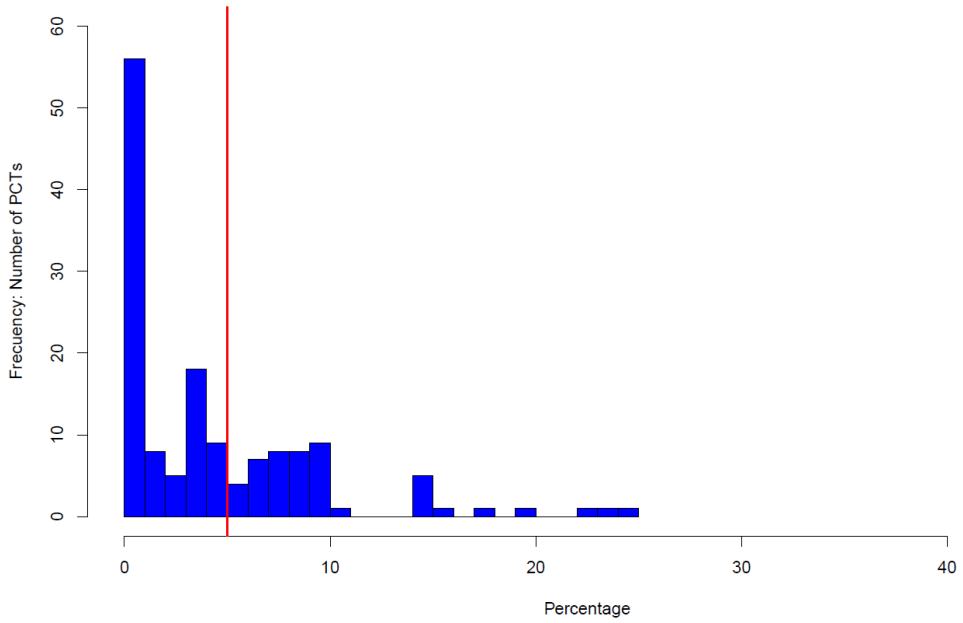
Figure 8 from 51 in Figure 7. This indicates that although the effect of the environmental variables limits the freedom of the PCTs to make savings that could be used for funding new health technologies, this capacity still exists, although to a lesser extent.

Figure 7. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Mental Health (no adjustment for environmental variables)



Two PCTs have missing data in the variable MH_Employment (5F7 and 5QT).
Outliers: Five outliers (5A3, 5HQ, 5NW, 5PY and 5QC) out of 149 observations.
Efficiency: 51 (35.4%) efficient PCTs out of 144.
Inefficiency: 81 (56.2%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

Figure 8. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Mental Health (with adjustment for environmental variables)



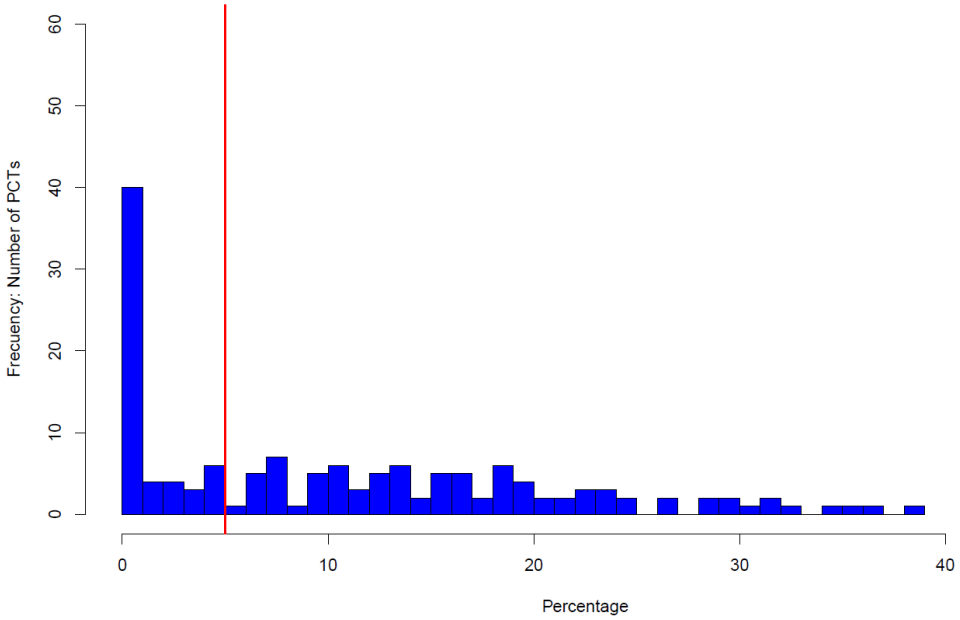
Two missing data in the variable MH_Employment (5F7 and 5QT)
Outliers: Five outlier (PCTs 5A3, 5HQ, 5NW, 5PY and 5QC) out of 149 observations.
Extent_2010 included as environmental variable instead of IMD_2010
Efficiency: 52 (36.1%) efficient PCTs out of 144.
Inefficiency: 48 (33.3%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

Additionally, the results suggest that only 13 out of the 52 efficient PCTs are located in the increasing returns to scale part of the productivity curve: 5ET, 5H8, 5JX, 5K3, 5LF, 5LH, 5LQ, 5N1, 5NP, 5NV, 5P1, 5PW and 5QP (see Appendix 3). This means that for the majority of the efficient PCTs (39 out of 52) a decrease in expenditures would affect outcomes in less than a proportional way.

4.1.2. Maternity

In comparison with mental health, PCTs appear to be less efficient in producing health outcomes for the PBC maternity, since there is a lower number of efficient PCTs. However, there are around a third of the PCTs for which the possible adjustment in expenditures is higher than 5% (Figure 10), which is similar to what was observed in Figure 8 for mental health.

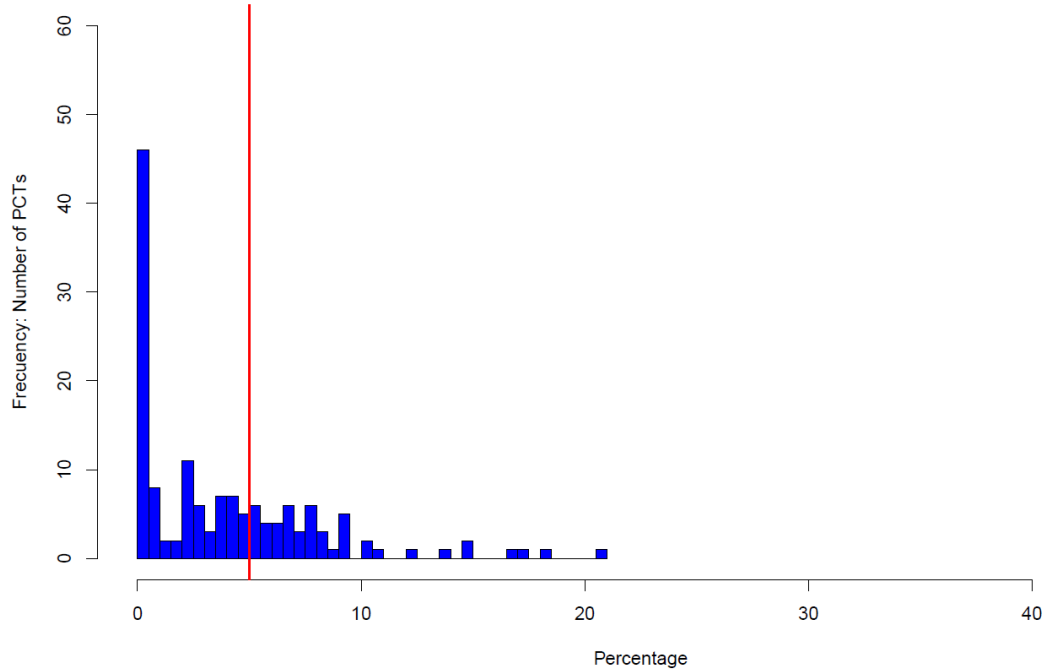
Figure 9. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Maternity (no adjustment for environmental variables)



One missing data in the variable NeonatalMort_2014_INV (5D7).
Outliers: Four outliers (5F7, 5M2, 5NW and 5QC) out of 150 observations.
Efficiency: 38 (26%) efficient PCTs out of 146.
Inefficiency: 89 (60.9%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

Similar to mental health, when comparing Figure 9 and Figure 10, it is possible to observe that the average level of efficiency increases when the environmental variables are adjusted for. This is to be expected, since to adjust the inputs for the effect of the environmental variables (see Appendix 2) is leveling the field by adapting for the conditions over which the PCTs do not have any control. Two variables were selected, first, the Deprivation Index whose inclusion responds to the literature suggesting that socioeconomic differences are translated into health outcomes differences (Pickett and Wilkinson, 2015; Macinko et al., 2003). Second, whether the PCT is underfunded or overfunded depending on their health needs as defined by the NHS allocation formula. A PCT could be efficient but not have the required resources to achieve the same level of health outcomes that another PCT. The latter could be inefficient but have a budget that is higher than the one required to satisfy its health needs. A deprived area and underfunded PCT producing as its maximum capacity would be classified as inefficient when the environment is not considered, since its relative disadvantage in terms of socioeconomic condition and funding level do not play a role in the estimation. The same PCT could be classified as efficient when scores are estimated based on a hypothetical situation in which a deprived area and underfunded PCT has the same conditions as a less deprived and overfunded PCT. A reduction in the estimated inefficiency is expected and observed for all PBCs when the inputs have been adjusted for the effect of the environmental variables.

Figure 10. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Maternity (with adjustment for environmental variables)



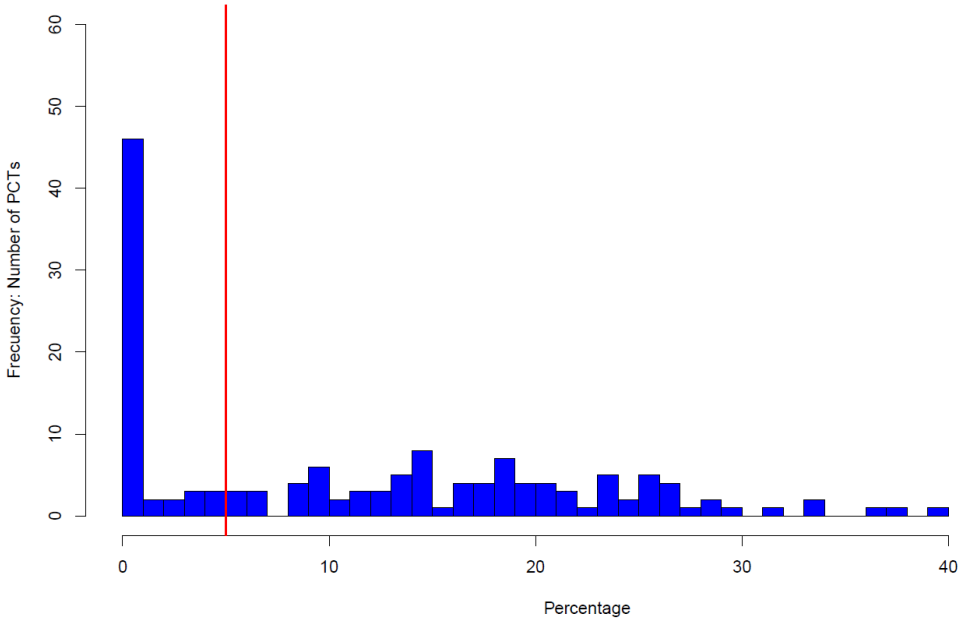
One missing data in the variable NeonatalMort_2014_INV (5D7).
Outliers: Four outliers (5F7, 5M2, 5NW and 5QC) out of 150 observations.
Efficiency: 41 (28.1%) efficient PCTs out of 146.
Inefficiency: 49 (33.6%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

In the Maternity PBC the results suggest that seven out of the 41 efficient PCTs are located in the increasing returns to scale part of the productivity curve: 5A3, 5D8, 5L1, 5ND, 5NF, 5PX and TAL.

4.1.3. Cancer

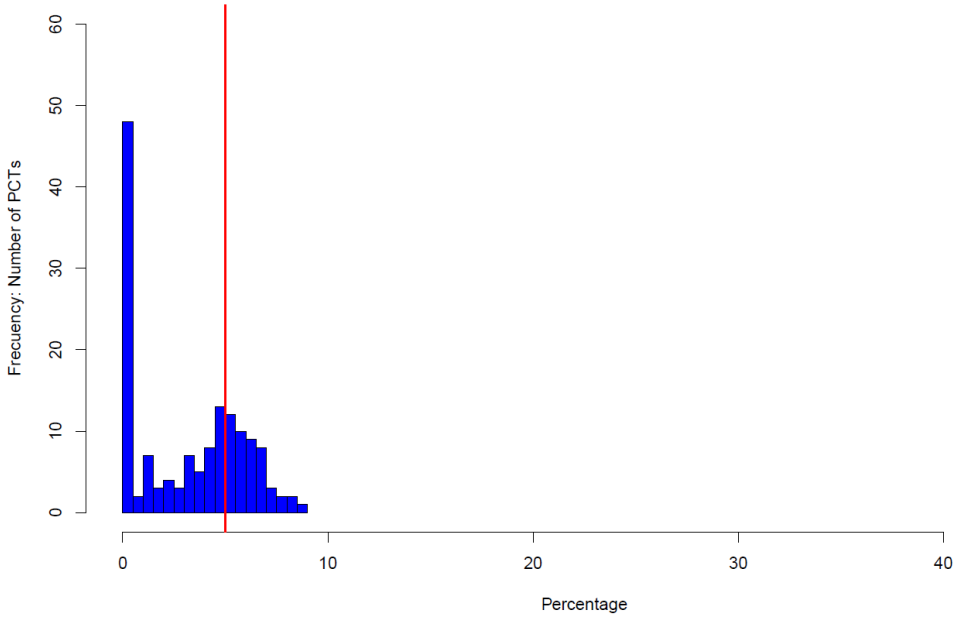
The DEA model estimated for Cancer considers only the sum of the per-capita expenditure in PBCs 21 and 22 but not PBC 23. When the sum of total expenditure in the three PBCs is included, the SFA suggests that the model is not well specified, and the error term cannot be split between statistical noise and inefficiency (see Appendix 2). The correct estimation of the SFA is a key step in the Fried et al. (2002) three steps methodology, therefore, expenditures in PBC 23, miscellaneous, are excluded from the additional expenditures (see section 3.2.3).

Figure 11. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Cancer (no adjustment for environmental variables)



No missing data. Outliers: Four outliers (5HY, 5LE, 5NA and 5PQ) out of 151 observations.
Efficiency: 43 (29%) efficient PCTs out of 147.
Inefficiency: 91 (61.9%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

Figure 12. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Cancer (with adjustment for environmental variables)



No missing data. Outliers: Four outliers (5HY, 5LE, 5NA and 5PQ) out of 151 observations.
Efficiency: 44 (29.9%) efficient PCTs out of 147.
Inefficiency: 47 (31.9%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

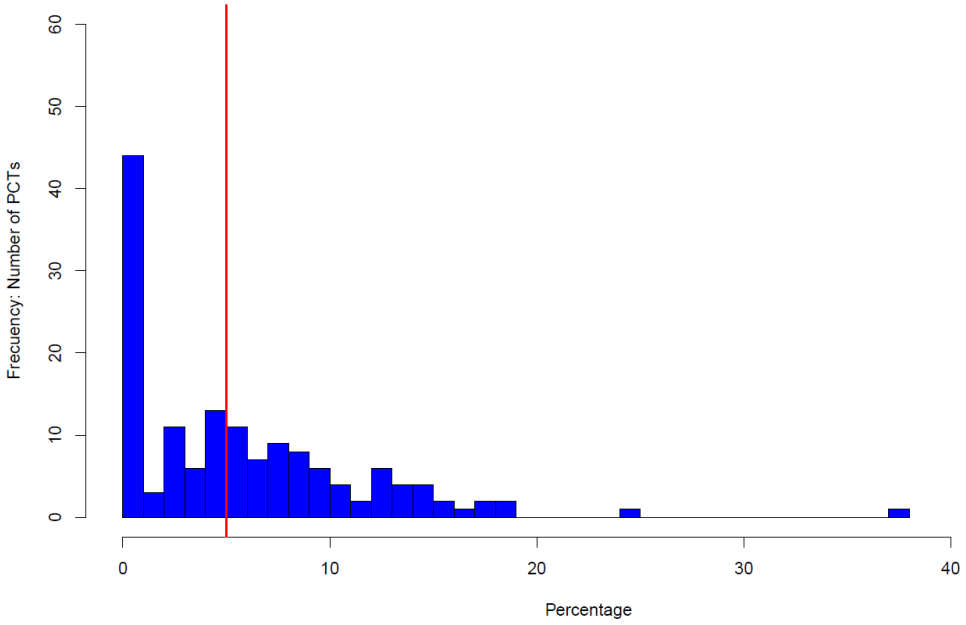
Eight efficient PCTs (5C1, 5C5, 5CN, 5N7, 5NC, 5NW, 5PC are 5PG) are located in the increasing returns to scale part of the productivity curve.

4.1.4. Gastrointestinal

Similar to cancer, adjusting for environmental variables has a decisive effect on the production of health outcomes in the gastrointestinal PBC (Figure 13 as compared to Figure 14). The number of PCTs that can adjust expenditures by more than 5% goes from being 47.6% in Figure 13 to 6.8% in Figure 14.

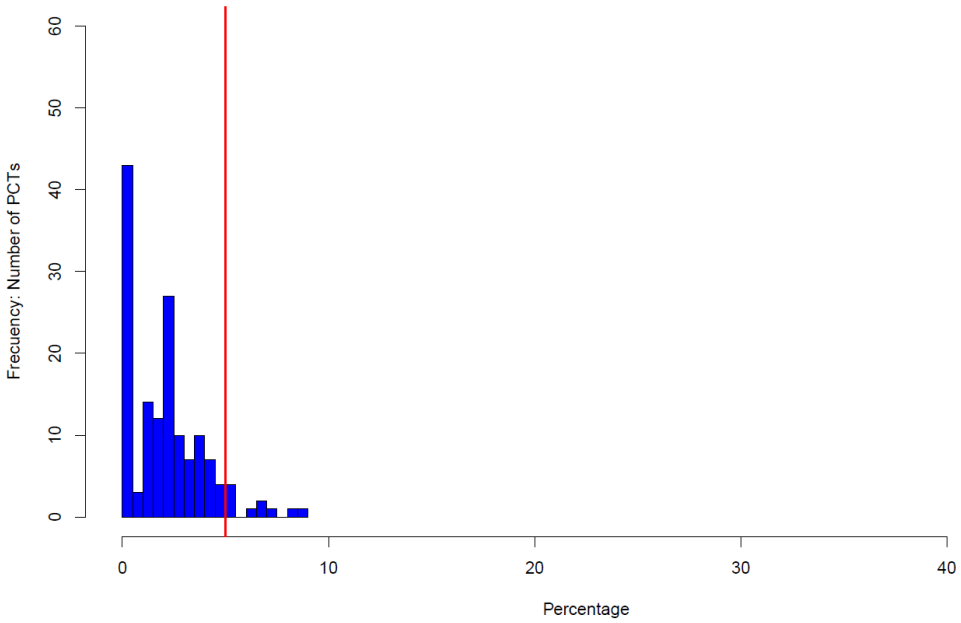
None of the 39 efficient PCTs are located in the increasing returns to scale part of the productivity curve. This suggests that, in spite of lower opportunity to make savings by increasing efficiency, the fact that the effect on health outcomes of a decrease in expenditures is less than proportional in all cases suggests a lower opportunity cost for the efficient group of PCTs in this are as in comparison with their performance in other PBCs.

Figure 13. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Gastrointestinal (no adjustment for environmental variables)



No missing data.
Outliers: Four outliers (5C3, 5MX, 5PL and 5QF) out of 151 observations.
Efficiency: 39 (26.5%) efficient PCTs out of 147.
Inefficiency: 70 (47.6%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

Figure 14. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Gastrointestinal (with adjustment for environmental variables)



No missing data.
Outliers: Four outliers (5C3, 5MX, 5PL and 5QF) out of 151 observations.
Efficiency: 39 (26.5 %) efficient PCTs out of 147.
Inefficiency: 10 (6.8 %) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

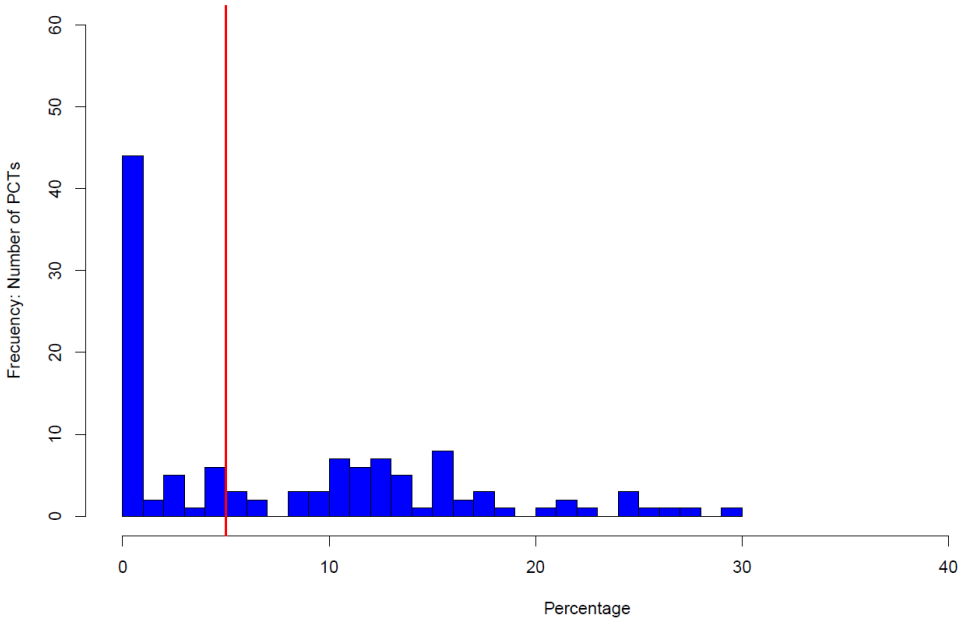
4.1.5. Cardiovascular

The DEA model estimated for cardiovascular considers as additional expenditure only the sum of the per-capita expenditure in PBCs 21 instead of the sum of the expenditures in PBCs 21, 22 and 23 (see section 3.2.3). This is because when the sum of the total expenditures of PBCs 22 and 23 were included, the SFA suggests that the model is not well specified, and the error term cannot be split between statistical noise and inefficiency (see Appendix 2).

The environmental variable TargetDistance_2010 was statistically insignificant in explaining the variations in the inefficiencies estimated in the first step of the Fried et al. (2002) methodology (Appendix 2). Therefore, it is excluded from the analysis. This is the only PBC for which one of the two selected environmental variables were not statistically significant.

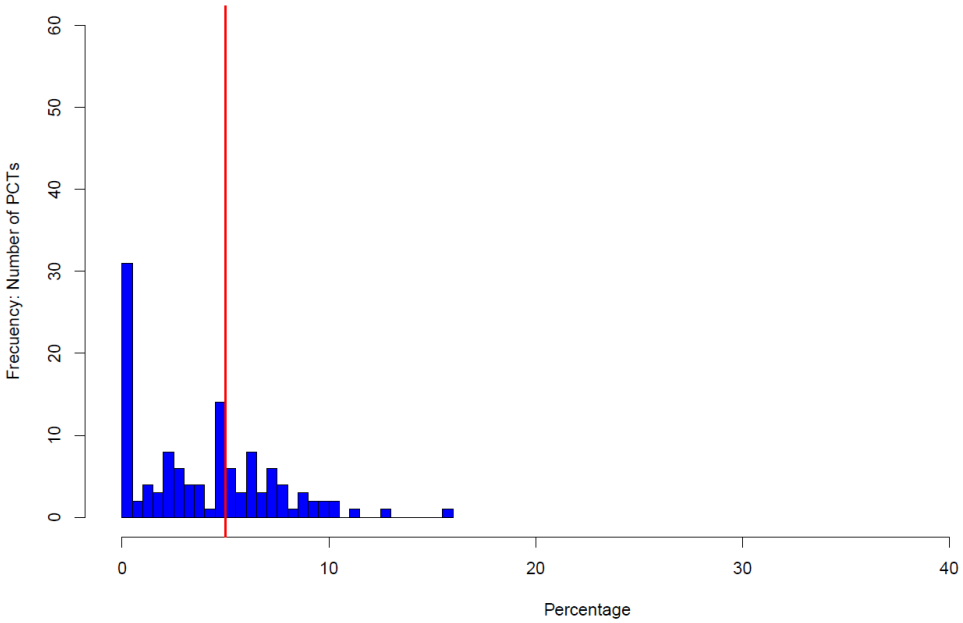
The cardiovascular PBC shows a level of efficiency higher than those observed in mental health and maternity, but lower than that observed in gastrointestinal.

Figure 15. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Cardiovascular (no adjustment for environmental variables)



23 missing data in the variable CardiacRehab_2014 and one additional missed data in the variable Stroke_discharge_2014. Outliers: Seven outliers (5EM, 5J6, 5LD, 5LE, 5N7, 5PX and 5QF) out of 127 observations.
Efficiency: 53 (44.2%) efficient PCTs out of 120. Inefficiency: 62 (51.7%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

Figure 16. Percentage expenditure decrease per year possible without affecting health outcomes: Cardiovascular (with adjustment for environmental variables)



23 missing data in the variable CardiacRehab_2014 and one additional missed data in the variable Stroke_discharge_2014. The environmental variable TargetDistance_2010 was excluded from the analysis. Outliers: Seven outliers (5EM, 5J6, 5LD, 5LE, 5N7, 5PX and 5QF) out of 127 observations. Efficiency: 52 (43.2%) efficient PCTs out of 120. Inefficiency: 43 (35.8%) PCTs can decrease per-capita expenditures per

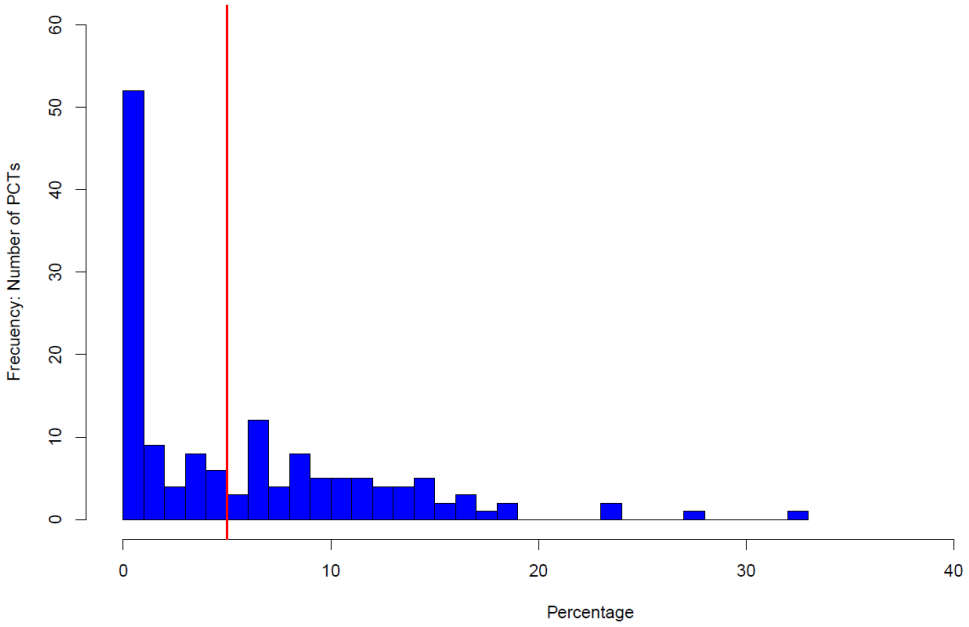
year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

One out of the 52 efficient PCTs is located in the increasing returns to scale part of the productivity curve: 5PC.

4.1.6. Respiratory

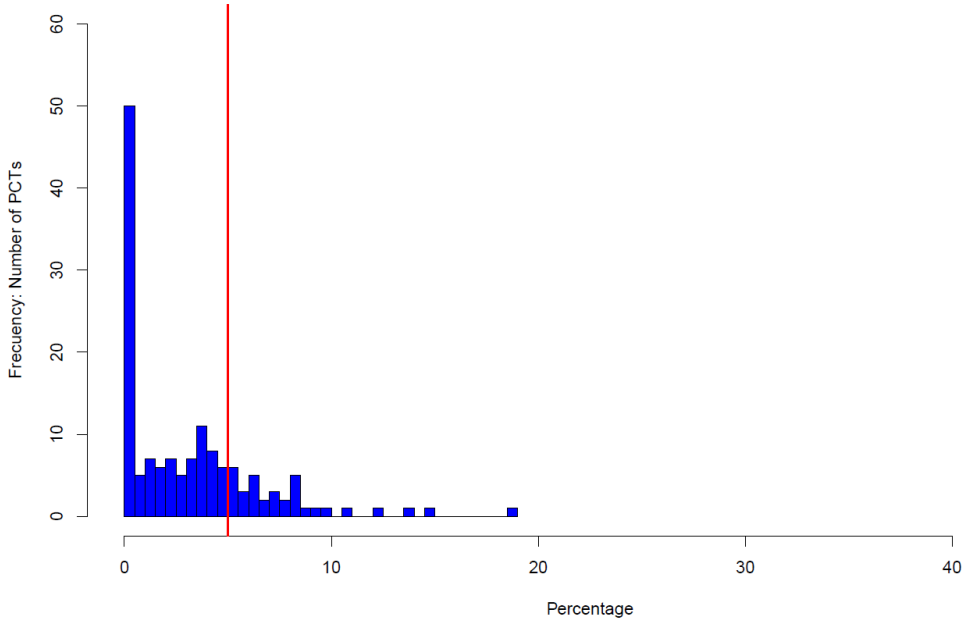
Figure 17 and Figure 18 show the results for respiratory.

Figure 17. Percentage expenditure decrease per year possible without affecting health outcomes: Respiratory (no adjustment for environmental variables)



One missing data in the variable EmergencyRespiratoryChild_2014_INV Outliers: Four outliers (5C3, 5C9, 5N7 and 5QP) out of 150 observations. Efficiency: 48 (32.9%) efficient PCTs out of 146. Inefficiency: 67 (45.9%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

Figure 18. Percentage expenditure decrease per year possible without affecting health outcomes: Respiratory (with adjustment for environmental variables)



One missing data in the variable EmergencyRespiratoryChild_2014_INV Outliers: Four outliers (5C3, 5C9, 5N7 and 5QP) out of 150 observations. Extent_2010 included as environmental variable instead of IMD_2010 Efficiency: 48 (32.9%) efficient PCTs out of 146. Inefficiency: 34 (23.3%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

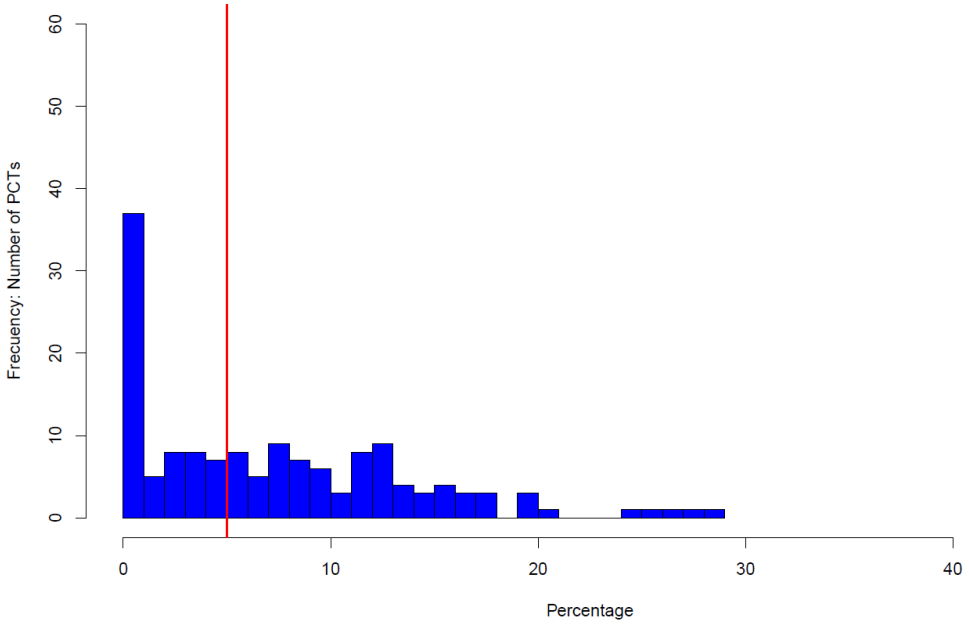
The number of PCTs that can decrease expenditures by over 5% without an effect on outcomes is smaller than in most of the other PBCs (23.3%), except for gastrointestinal (6.8%). However, the number of fully efficient PCTs is similar.

Seventeen out of the 48 efficient PCTs are located in the increasing returns to scale part of the productivity curve: 5CN, 5ET, 5JX, 5K8, 5LA, 5LF, 5LQ, 5M1, 5M2, 5M8, 5NP, 5PK, 5PL, 5PT, 5PW, 5QA and 5QN.

4.1.7. Endocrine

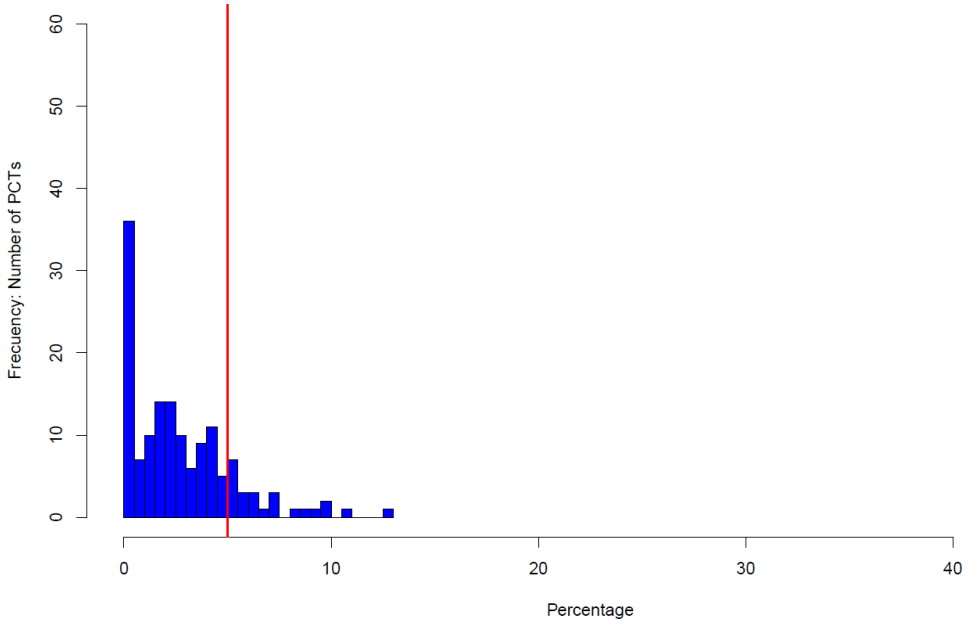
Figure 19 and Figure 20 display the results for the PBC endocrine.

Figure 19. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Endocrine (no adjustment for environmental variables)



No missing data.
Outliers: Five outliers (5D9, 5F7, 5H8, 5L1 and 5QQ) out of 151 observations.
Efficiency: 34 (23.3%) efficient PCTs out of 146. Inefficiency: 81 (55.5%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

Figure 20. Percentage decrease in expenditure per year that could be possible without affecting health outcomes: Endocrine (with adjustment for environmental variables)



No missing data.
Outliers: Five outliers (5D9, 5F7, 5H8, 5L1 and 5QQ) out of 151 observations.
Efficiency: 34 (23.3%) efficient PCTs out of 146. Inefficiency: 24 (16.44%) PCTs can decrease per-capita expenditures per year in at least **5%** without affecting health outcomes. These correspond to all observations located after the **red line**.

Thirty two out of the 34 efficient PCTs are located in the increasing returns to scale part of the productivity curve: 5A3, 5A4, 5A8, 5EM, 5F1, 5FE, 5HQ, 5JX, 5KL, 5LH, 5LQ, 5M2, 5M8, 5N5, 5N7, 5NE, 5NP, 5NV, 5NW, 5P1, 5PL, 5PQ, 5PT, 5PW, 5QC, 5QF, 5QJ, 5QM, 5QP, 5QT, 5QV and TAC. This suggest that endocrine is the PBC for which the opportunity cost of the group of efficient PCTs is the highest.

4.1.8. Robustness test to the choice of final outcomes

We present non-parametric robustness tests in Table 7 to assess to what extent the results are a consequence of the outcomes choices. We conduct a DEA with only mortality results, and analyse the contribution of other of other outcomes, one at each step.

Two non-parametric statistical test procedures are used to compare the distribution of DEA scores for different outcomes portfolio: (1) a Kolmogorov-Smirnov test of the equality of the distributions, and (2) a Kruskal–Wallis test of the hypothesis that several samples are from the same population.

Each PBC in Table 7 shows the DEA outcomes in rows with the value of the tests indicating whether the corresponding outcome does or does not have a significant importance for the change of the DEA efficiency frontier.

Table 7. Robustness test to the choice of final outcomes

	Kolmogorov-Smirnov Test		Kruskal-Wallis test	
DEA Final Outcomes	**Value of the test**	**p-value**	**Value of the test**	**p-value**
Mental Health				
MH_Independently	0.16*	0.05	5.75**	0.02
MH_Employment	0.05	1.00	0.34	0.56
MH_ExcessMort_INV_2014	0.03	1.00	0.14	0.71
MH_HRQoL_2014	0.09	0.60	1.27	0.26
Maternity				
NeonatalMort_2014	0.07	0.88	0.65	0.42
MAT01_Point_2012	0.13	0.17	4.75**	0.03
Cancer				
SYLLR_CancerPCT_2014_INV	0.07	0.81	0.97	0.32
Last_OneYSurv_PCT_2014	0.12	0.28	1.59	0.21
Gastrointestinal				
SYLLR_GastroPCT_2014_INV	0.16**	0.04	3.58*	0.06
Last_AlcoholLiverEmerg_2014_INV	0.07	0.89	0.44	0.51
Cardiovascular				
SYLLR_CVDPCT_2014_INV	0.24***	0.00	8.93***	0.00
Last_CardiacRehab_2014	0.14	0.18	3.28*	0.07
Stroke_discharge_2014	0.13	0.31	3.41*	0.06
Respiratory				
SYLLR_RespiratoryPCT_2014_INV	0.10	0.51	1.38	0.24
EmergencyRespiratoryChild_2014_INV	0.11	0.34	3.36*	0.07
Endocrine				
SYLLR_EndocrinePCT_INV	0.05	1.00	0.44	0.51
Last_DiabComplications_2014_INV	0.10	0.51	0.94	0.33

Notes: Significance levels: *** p<0.01, ** p<0.05, * p<0.1

In Endocrine, Cancer, and Respiratory the distribution of DEA efficiency scores is not sensitive to the choice of outcomes at 5% significance level. The PBC most affected by the choice of outcomes is Cardiovascular, where the three DEA outcomes have a significant importance. That is, they do not represent the same underlying efficiency frontier according to the Kruskal-Wallis test. For Mental Health, the outcome on independent living has a significant importance. For Maternity, it is the score of the QOF primary care practice that has a significant impact. For Gastrointestinal, a SYLLR has significant importance.

4.1.9. Summary of the DEA results

Table 8 shows a summary of the results presented in the previous figures. Gastrointestinal and endocrine present the lowest percentage of efficient PCTs. However, these PBCs are also the ones with the lowest percentage of PCTs that can decrease expenditures by more than 5%. This indicates that in these two PBCs, although the PCTs managers have a margin of freedom for funding new health technologies by increasing efficiency, this margin could be quickly depleted. On the opposite side, cardiovascular and mental health shows the highest percentages of fully efficient PCTs, but at the same time the highest percentages of inefficient PCTs with a margin of adjustment over the 5%. In this case, there are fewer PCTs managers with a margin for squeezing efficiency out of their PCT, but those that can, have a larger amount of inefficiency to exploit. These results indicate complexity in understanding the opportunity cost of adopting a new technology.

Table 8. Ranking of PBCs

PBC	% of fully efficient PCTs	Ranking efficient	% of PCTs that can decrease expenditures in more than 5%	Ranking Inefficient
Cardiovascular	43.2	1	35.8	7
Mental health	36.1	2	33.3	5
Respiratory	32.9	3	23.3	3
Cancer	29.9	4	31.9	4
Maternity	28.1	5	33.6	6
Gastrointestinal	26.5	6	6.8	1
Endocrine	23.3	7	16.4	2

According to the results presented in Table 8, respiratory, cancer and, to a lesser extent, maternity, can be considered PBCs with a middle level of efficiency.

In Table 9 the efficient PCTs that are in the increasing returns to scale group of PCTs in more than one PBC are displayed. These is the group of PCTs with the highest opportunity cost. Four PCTs Bury (5JX), Brighton and Hove City (5LQ), Central and Eastern Cheshire (5NP) and North East Essex (5PW) are in the increasing returns to scale group in the PBCs Mental, Respiratory and Endocrine. This means that these four PCTs have highest opportunity costs in more PBCs in comparison to the remaining PCTs.

Table 9. Fully Efficient PCTs located in the increasing returns to scale part of the productivity curve in more than one PBC

Maternity and Endocrine	Cardiovascular and Cancer	Respiratory and Cancer	Endocrine and Cancer	Mental and Respiratory	Mental and Endocrine	Respiratory and Endocrine	Mental, Respiratory and Endocrine
5A3	5PC	5CN	5N7	5ET	5LH	5M2	5JX
			5NW	5LF	5NV	5M8	5LQ
					5P1	5PL	5NP
					5QP	5PT	5PW

Out of the 151 PCTs included, 77 are included in the increasing returns to scale group in at least one PBCs. Table 9 shows the 25% of the 77 PCTs that are in the group with the

highest opportunity cost in more than one PBC. The fact that 75% of the mentioned 77 PCTs are in the increasing return to scale group in only one PBC once again highlights the variation on opportunity cost across PCTs and PBCs.

4.2. Quantile Regression

As noted above we have used a quantile regression model with joint estimation of the variance and covariance matrix of the quantiles: $\tau_1 = 0.1$, $\tau_2 = 0.25$, $\tau_3 = 0.5, \tau_4 = 0.75$, $\tau_5 = 0.9$, to allow for tests comparing effects across quantiles. Given that our purpose is to complement the results of the regression methods presented by Lomas et al. (2018), we also use a linear regression model which estimates the covariate effects at the conditional mean. We focus our analysis on the outcome models, that is, models which analyse the effect of PBC expenditure on mortality for diseases in the same PBC. For these outcome models, we adopt the outcome specification preferred by Lomas et al. (2018) and the estimation method using instrumental variables to account for the endogeneity of the explanatory variable spend per head.

The results are presented below in Figure 21 to Figure 26. Details of the estimations are presented in the Appendix 4, including details of the specification of outcome models and of the instrumental variables used. Figure 21Figure 21 to Figure 26 show in the black horizontal lines the estimate of the average effect and 95% confidence interval as obtained from linear regression methods. We use IV techniques instead of OLS to estimate the linear regression models because expenditure is endogenous and, in the presence of an endogenous regressor, OLS will provide both a biased and an inconsistent estimator of the returns to spend. Nonetheless, when the test of endogeneity results in non-rejection of the exogeneity assumption for expenditure (for the PBCs Infectious and Endocrine), we present OLS estimates. The linear regression model can be compared with the quantile regression estimated presented by the blue line, also with 95% confidence intervals.

For PBC 1 (Infectious Diseases) Figure 21, PCTs are more efficient in reducing mortality at the lower tail of the mortality distribution, i.e. when the mortality rate is low. This return to spend is significantly larger than the average effect estimated by least squares: an elasticity of -0.7 at quantile 10% versus -0.3 at the mean. The effect is nil for PCTs with high rates of mortality caused by infectious diseases (e.g. the estimated elasticity at quantile 0.75 is -0.15 and not statistically significant). The finding of marginal returns to spend decreasing when mortality level increases is a distinctive characteristic of infectious diseases and is not easy to understand. It may reflect the contagious nature of the disease such that preventive measures in low and mid risk populations are more effective in preventing mortality and that little can readily be done to avoid mortality in high risk and affected populations.

For all cancer mortality (Figure 22), the average return to spend in the Cancer PBC 2 reflected in the conditional mean elasticity (-0.35) is only representative of the middle part of the QR distribution, at quantiles 0.25 and at the median 0.50. At low mortality rates, cancer mortality is not much reduced by increasing expenditure. In contrast, at the upper tail represented by the conditional effect on quantiles 0.75 and 0.90, mortality elasticity to expenditure is significantly larger: for those PCTs with the largest SYLL mortality rates, the return to spend is about a 0.7% reduction in mortality for a 1% increase in spend per head, around double the return of 0.35% at the mean and median.

For mortality caused by Diabetes (Figure 23), our results align with Lomas et al. (2018) pointing at no significant effect, either at the mean or at different quantiles, although the

estimate at the median is significant at 5% level and represents a reduction of 0.43% in SYLL mortality rate for a 1% increase in spend per head allocated to the PBC of Endocrine Diseases.

The effect of PBC spend per head on mortality as measured by SYLL rate is mostly stable along the distribution of mortality for PCBs 10, 11, and 13 (Circulatory, Respiratory and Gastrointestinal diseases), set out in Figures 24, 25 and 26. This average effect represents about 1.5% decreases in SYLL mortality rate for a 1% increase in PBC spend per head. The estimates obtained from the linear regression are close to the quantile regression estimates for the median and are good measures of the effect for all 151 PCTs along the mortality distribution for these three PBCs. Nonetheless, the effects are more precisely estimated by quantile regression in the Circulatory diseases PBC and we can see significant differences to the outcome elasticity to spend obtained as between quantiles in the lower and upper tails of the mortality distribution, with significantly larger reductions in mortality for PCTs with high mortality rates caused by CVD.

Figure 21. QR - SYLL Infectious Diseases elasticity to PBC spend

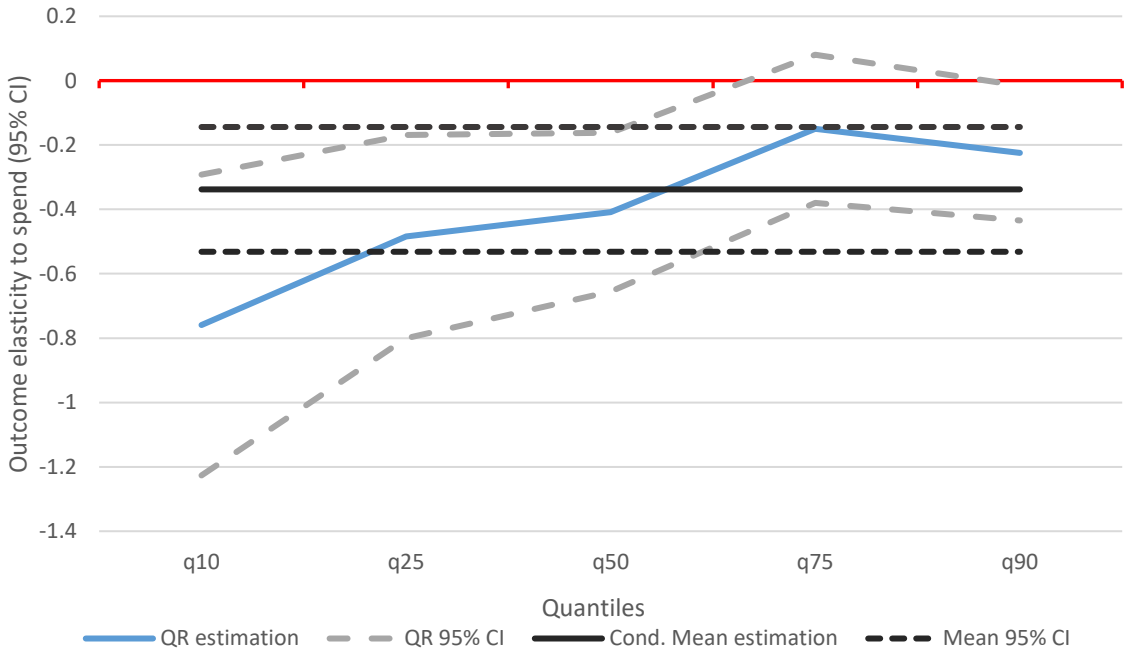


Figure 22. QR - SYLL Cancer elasticity to PBC spend

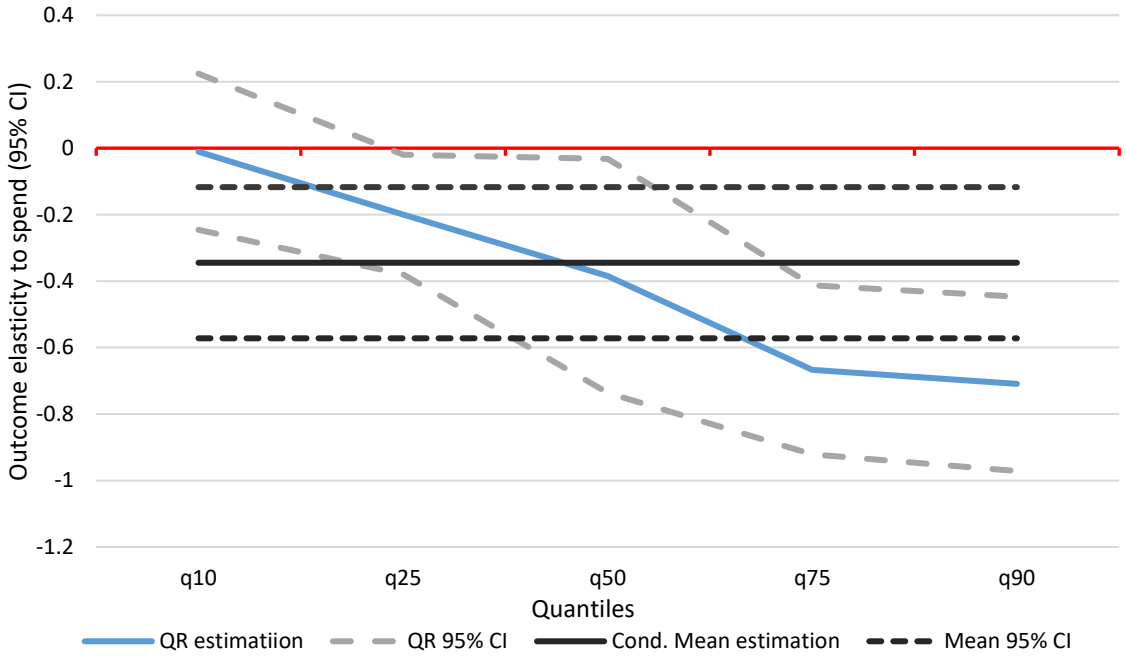


Figure 23. QR - SYLL Endocrine Diseases elasticity to PBC spend

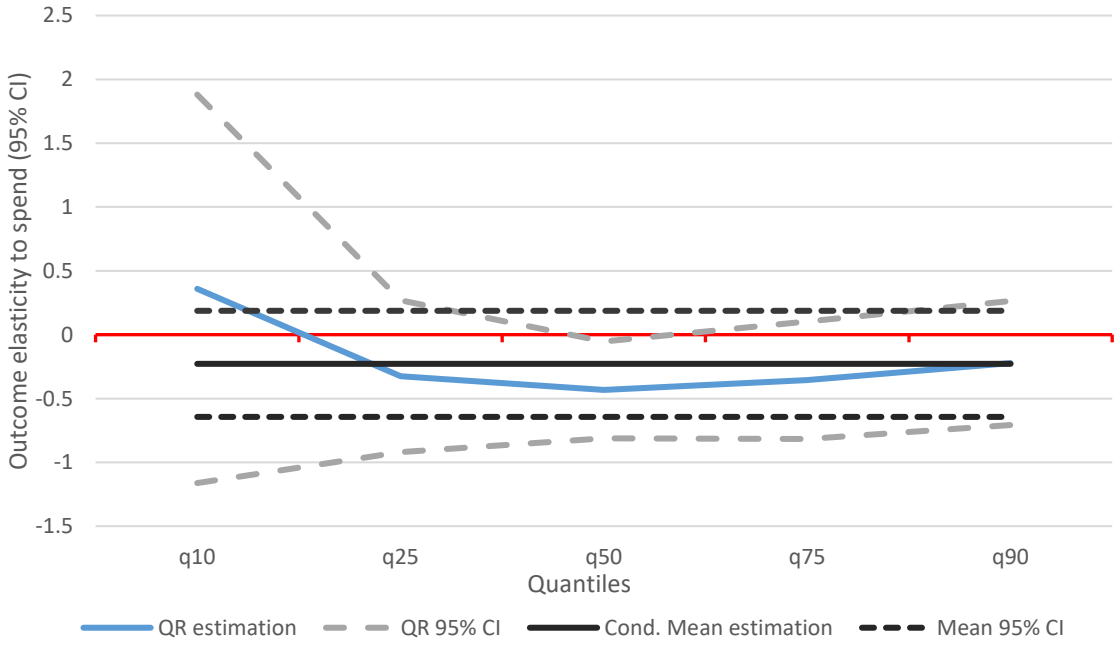


Figure 24. QR - SYLL Circulatory Diseases elasticity to PBC spend

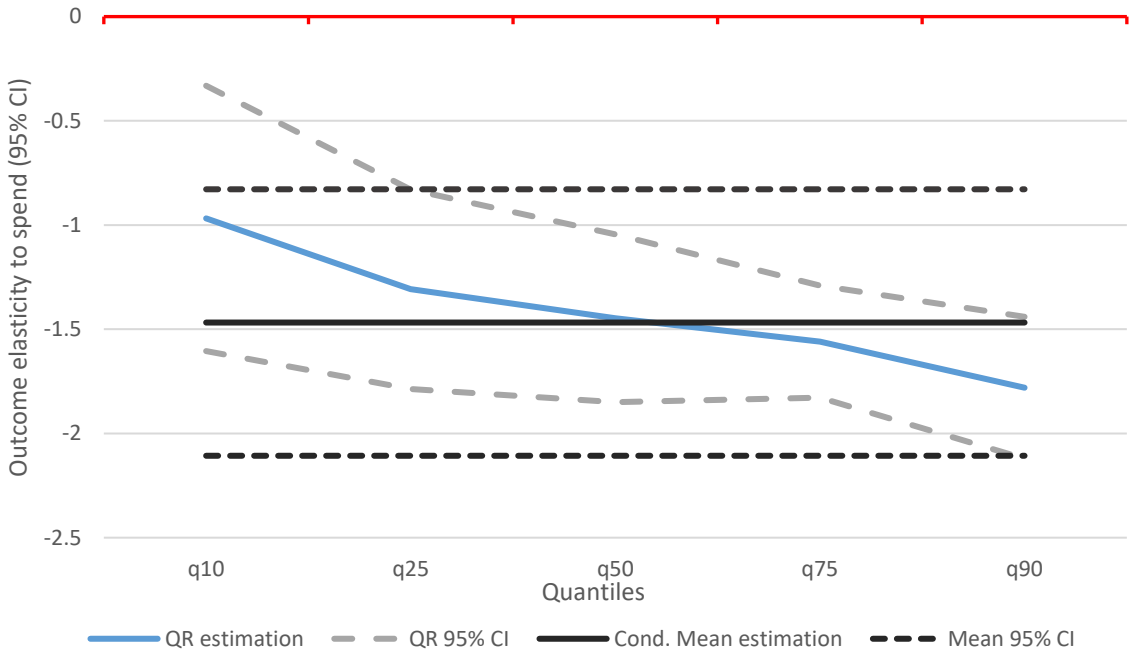


Figure 25. QR - SYLL Respiratory Diseases elasticity to PBC spend

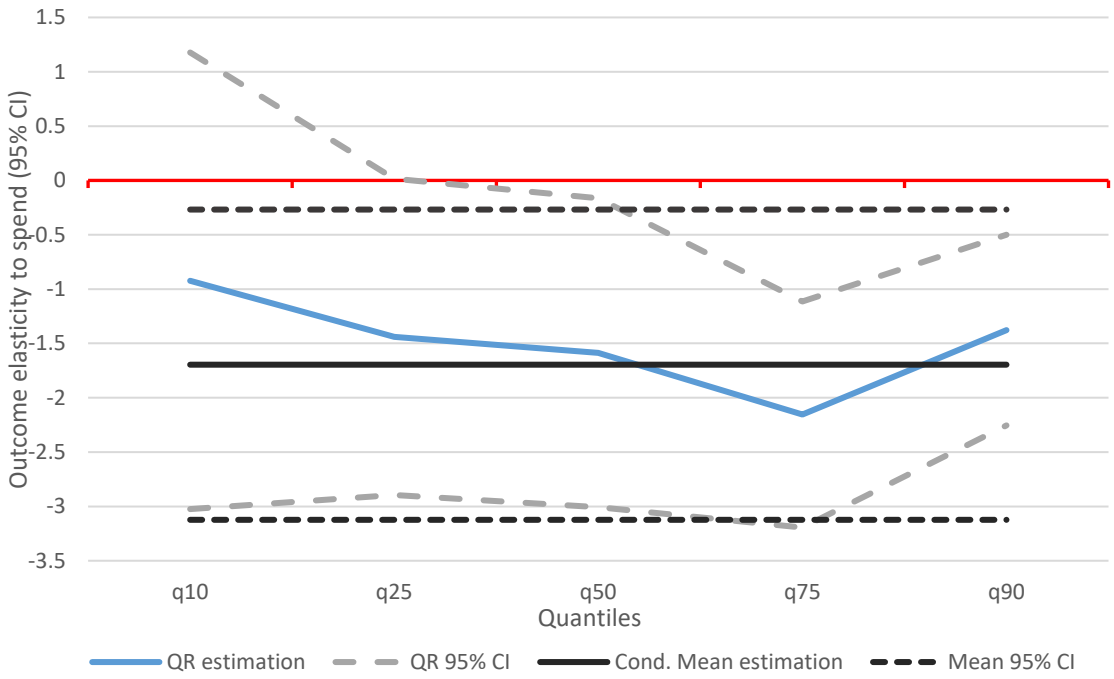
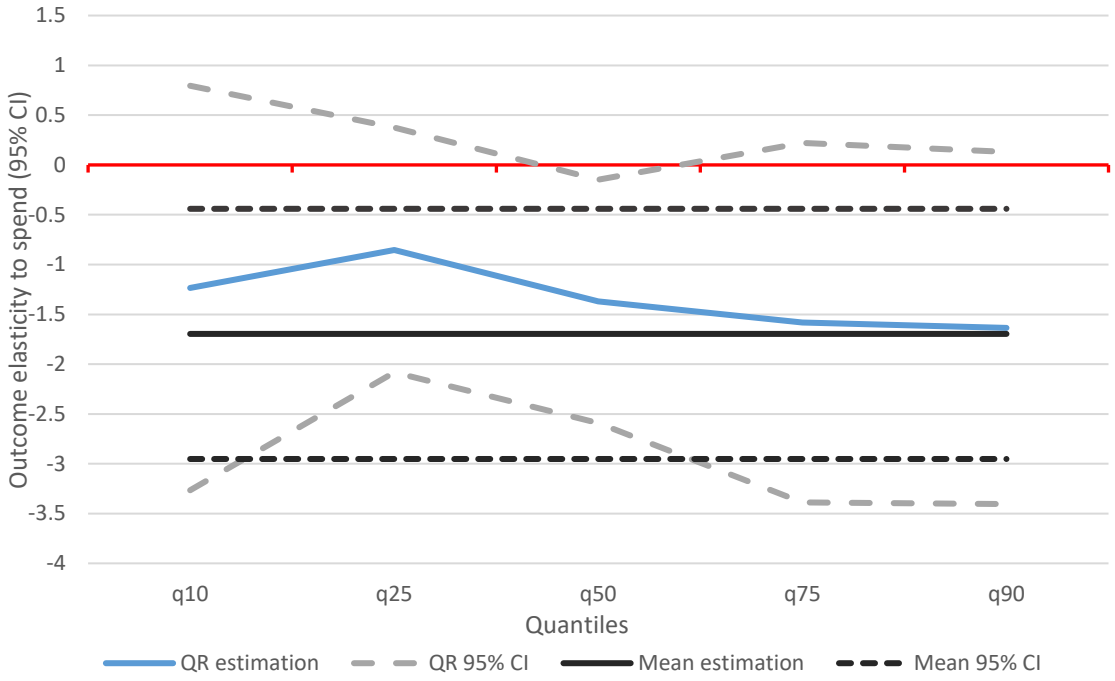


Figure 26. QR - SYLL Gastrointestinal Diseases elasticity to PBC spend



Results from the QR suggest that the effect of increasing health expenditure per head on the reduction of mortality rate can be different as between different groups of PCTs according to their mortality rates. These differences are observed across PBCs.

These two findings might have important implications for health resource allocation in England. Instead of assuming the same elasticity for a PBC across PCTs, the heterogeneity across PCTs and PBCs regarding how efficient providers convert health resources to health outcomes should be recognised by decision makers

5. COMPARISON OF DEA AND QUANTILE REGRESSION RESULTS

Although the objectives and methods of DEA and QR are different, they both explore the efficiency of the health system. Two questions arise: first, whether there is any consistency in the findings from the two approaches; and secondly, what they tell us about the efficiency of expenditure at the margin.

DEA constructs a measurement at the PCT level of technical efficiency based on the distance between (composite) inputs and (composite) outputs. It identifies the most efficient PCTs, those that achieve the highest level of health outcomes given the fixed level of expenditure, which form the production frontier. The scale elasticity of the PCTs on the frontier is then calculated, identifying whether they are achieving increasing, constant, or decreasing returns to scale. The performance of PCTs who are not on the frontier are compared with those who are to give a measure of their performance in a particular programme area.

QR estimates the effect of health spending on the mortality rate, pooling all PCTs. The slope coefficient measures the elasticity of incremental effects on the mortality rate associated with incremental changes in health expenditure[6]. The unique feature of the QR is that it allows estimation of a slope coefficient for each quantile by introducing different weights at different points of the mortality distribution. We choose the 151 quantiles representing the mortality rate of each PCT.

There is a systematic negative and significant correlation between PCTs' DEA efficiency scores and the QR outcome elasticity estimates in each PBC. Correlation coefficients range between -0.15 and -0.5.

The absolute value of QR elasticities is positively correlated with the mortality level for the five PBCs also analysed in DEA. Correlation coefficients range between 0.5 and 0.9 for cancer and circulatory diseases respectively.

Rank correlation analysis was used to compare the DEA efficiency scores and the QR outcome elasticities. Two efficiency score rankings were calculated, estimated from models with and without environmental variables, with higher values representing greater efficiency. The outcome elasticity rankings were also ranked from highest to lowest.

Table 10 shows the results using Spearman rank correlation coefficients in the five clinical areas that are common between the DEA and QR analyses. The correlations are all negative, and significant except for Endocrine disease. Results are similar when considering DEA efficiency scores obtained with and without environmental variables. They imply that more efficient PCTs (as measured in the DEA analysis) tend to have a lower absolute value of mortality elasticity to spend (as measured by the QR analysis); in other words, they obtain lower reductions in mortality for a marginal increase in health expenditure.

The last three columns of Table 10 also show a comparison of the mean mortality elasticity to spend in efficient and other PCTs, using estimates including the environmental variables. The results are consistent with those obtained from the ranking correlations and show that an increase in 1% of spend in each PBC results in a reduction in mortality about 1% lower in efficient PCTs than in others. The exception, again, is for Endocrine diseases, where the difference is not statistically significant.

Table 10. Comparison of DEA and QR estimations

	Spearman rank correlation		**Mean-comparison test**		
	No EV	**With EV**	**Fully efficient μ1**	**Non-fully efficient μ0**	**p-value H0: μ0=μ1 H1: μ0≠μ1**
PBC 2: Cancer	-0.211**	-0.214***	0.346	0.478	0.003
PBC 4: Endocrine	-0.107	-0.078	0.134	0.248	0.099
PBC 10: Cardiovascular	-0.267***	-0.322***	1.314	1.493	0.007
PBC 11: Respiratory	-0.169**	-0.194**	1.264	1.546	0.009
PBC 13: Gastrointestinal	-0.230***	-0.292***	1.084	1.366	0.002

Notes: Significance levels: *** p<0.01, ** p<0.05, * p<0.1

[6] Because the DEA analysis is based on multiple outcomes, the estimated scale elasticities are not directly comparable with the QR outcome elasticities. The DEA estimates assume that with a given increase in inputs, outputs increase in the same proportion as each other.

A plausible explanation for these results is that PCTs operating efficiently in a PBC tend to have lower rates of mortality, and for most disease areas, the lower the mortality, the harder it is to achieve additional reductions.

6. DISCUSSION AND POLICY IMPLICATIONS

Estimation of an opportunity cost-based cost-effectiveness threshold using a health production function approach involves many assumptions about the behaviour of the implied function. These are compounded by the nature of the programme budgeting data that are used for estimation, both because PBCs aggregate activity over conditions that are highly heterogeneous with respect to health outcomes and cost, and also because the expenditure data are based on NHS accounting rules designed for a different purpose than to measure with any degree of precision actual expenditure on care within PBCs . This study uncovers further problems with these assumptions that may undermine attempts to obtain a simple singular system-wide threshold estimate.

The existence of production inefficiency, that is the inability of some PCTs to achieve the best practice performance found in others, means that estimates of the opportunity cost of introducing new technologies based on average performance could be (i) biased and (ii) subject to far greater variation than normally assumed. Moreover, the PCTs who are found to be inefficient vary between PBCs, confounding further the plausibility of estimates based on averages. There is evidence for some PBCs that some apparent inefficiencies result from adoption of a different underlying production function technology, casting further doubts on the assumption of a common production function for all that underlies a common threshold.

The direction of bias is unclear *a priori*. The presence of inefficiency means that a reduction in a PCTs effective budget due to additional expenditure required to fund a new technology could, at least in part, be met without reductions in outcomes by improving efficiency, suggesting a lower opportunity cost and therefore a higher cost-effectiveness threshold than implied by the average performance of PCTs. However, the negative relationship between efficiency and outcome elasticity implies that less efficient PCTs have more opportunities to improve outcomes for low cost than more efficient ones, who are more likely to face decreasing returns to scale. If the response to a reduction in the effective budget is to improve efficiency, the resources released should be spent on technologies with low ICERs that have already been adopted by more efficient PCTs; if those resources are spent instead on the new technology, the opportunity cost may be greater than the average, implying a lower, not a higher, cost-effectiveness threshold is relevant for inefficient PCTs. Similarly, if inefficient PCTs do not improve their efficiency then this implies that a reduction in expenditure may have a greater effect on outcomes than implied by the average.

This issue highlights a fundamental problem with the attempt to estimate a cost-effectiveness threshold from opportunity cost information. It is apparent that the true opportunity costs of reductions in expenditure depend not only on estimates of an aggregate 'health production function', but also on the real-world responses of health decision makers to the local constraints that they face and to changes in them. Levels of inefficiency are only one of the variables that affect how national average estimates of notional opportunity costs impact on opportunity costs in terms of locally provided services.

The QR results suggest another route into estimating an "average" threshold, weighting by the absolute levels of mortality in a particular disease area. Again, it is not clear *a priori* how this would vary compared to the average estimated using linear regression. It depends on the attributes of the underlying production function to deliver health in each disease area, as well as the degree of inefficiency, both of which will impact on the observed QR results.

The implications of this for setting a cost-effectiveness criterion for NICE and other NHS bodies are therefore not straightforward. However, they suggest that the direct translation from estimated levels of historic opportunity cost to cost-effectiveness thresholds for future investment is not justified. The average estimates generated by current research use a very large number of empirical and theoretical assumptions that are equally hard to justify, particularly when different approaches and assumptions produce radically different estimates. Although the criterion is labelled 'cost-effectiveness', its use by the NHS is in support of its equity goal; in the case of NICE, its aim of ensuring geographical equity. A possible conclusion from our analyses might be that different cost-effectiveness criteria should be set for different PBCs or PCTs, but both of those options might conflict with stated system wide equity goals. The equity implications of the current NHS-wide cost-effectiveness criterion in practice should be examined in the light of our findings, since they suggest that its impact differs between different clinical areas (PBCs) and geographical areas (PCTs) and therefore the patient populations served by each. This is contrary to stated NHS and NICE aims.

One way to approach this problem is to accept that there are multiple sources of information relevant to the setting of cost-effectiveness criteria and that these may not be capable of being synthesised using scientific methods alone, but involve political judgements. A possible source of information would be an NHS mandated, targeted and supported survey of opportunity costs in terms of services at the local level, to generate routine data on this issue; *ad hoc* academic studies such as Appleby et al. (2009) are not adequate for that purpose. Moreover, estimating a value for the criterion based on current practice is not enough; evidence should also be incorporated on the likely effects of any criteria that are set. It may be that it is already the case that NICE, the NHS and the DHSC do exactly that, but if so, this is not transparent. An alternative would be an independent public body specifically tasked with assessing the evidence, commissioning evidence where it is lacking, and publishing the evidence and the body's deliberations and conclusions (Appleby, Devlin and Parkin, 2007).

7. REFERENCES

Appleby, J., Devlin, N. and Parkin, D., 2007. NICE cost effectiveness threshold. How high should it be. *BMJ,* 335 358-359.

Appleby, J., Devlin, N., Parkin, D., Buxton, M. and Chalkidou, K., 2009. Searching for cost effectiveness thresholds in the NHS. *Health Policy,* 91 (3), 239-245.

Baade, P. D., Dasgupta, P., Youl, P. H., Pyke, C. and Aitken, J. F., 2016. Geographical inequalities in surgical treatment for localized female breast cancer, Queensland, Australia 1997–2011: improvements over time but inequalities remain. *International Journal of Environmental Research and Public Health,* 13 (7), 729.

Bogetoft, P. and Otto, L., 2011. Benchmarking with DEA, SFA, and R. Springer.

Brenner, H., Mielck, A., Klein, R. and Ziegler, H., 1991. The role of socioeconomic factors in the survival of patients with colorectal cancer in Saarland/Germany. *Journal of Clinical Epidemiology,* 44 (8), 807-815.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P. C. and Sculpher, M., 2015. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technology Assessment (Winchester, England),* 19 (14), 1.

Cylus, J., Papanicolas, I. and Smith, P. C., 2015. Using data envelopment analysis to address the challenges of comparing health system efficiency. *Global Policy,* n/a-n/a.

Department of Health (DH), 2011. *Resource Allocation: Weighted Capitation Formula (Seventh Edition)*. London. Department of Health (DH).

Emrouznejad, A., Parker, B. R. and Tavares, G., 2008. Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences,* 42 (3), 151-157.

Fortin, N., Lemieux, T. and Firpo, S., 2011. Decomposition methods in economics. *Handbook of Labor Economics.* Elsevier.

Fried, H. O., Lovell, C. K., Schmidt, S. S. and Yaisawarng, S., 2002. Accounting for environmental effects and statistical noise in data envelopment analysis. *Journal of Productivity Analysis,* 17 (1-2), 157-174.

Haynes, R., Pearce, J. and Barnett, R., 2008. Cancer survival in New Zealand: ethnic, social and geographical inequalities. *Social science & medicine,* 67 (6), 928-937.

Hernandez-Villafuerte, K., Zamora, B. and Towse, A., 2018. *Issues surrounding the estimation of the opportunity cost of adopting a new healthcare technology*. Office of Health Economics.

Hollingsworth, B., 2008. The measurement of efficiency and productivity of health care delivery. *Health Economics,* 17 (10), 1107-1128.

Jansen, L., Eberle, A., Emrich, K., Gondos, A., Holleczek, B., Kajüter, H., Maier, W., Nennecke, A., Pritzkuleit, R. and Brenner, H., 2014. Socioeconomic deprivation and cancer survival in Germany: an ecological analysis in 200 districts in Germany. *International Journal of Cancer,* 134 (12), 2951-2960.

Jones, A. M., 2011. Models for Health Care. *In:* Clements, M. & Hendry, D. (eds.) *The Oxford Handbook of Economic Forecasting.* Oxford University Press.

Koenker, R. and Bassett, G., 1978. Regression quantiles. *Econometrica,* 46 (1), 33-50.

Kuznetsov, L., Maier, W., Hunger, M., Meyer, M. and Mielck, A., 2012. Regional deprivation in Bavaria, Germany: linking a new deprivation score with registry data for lung and colorectal cancer. *International Journal of Public Health,* 57 (5), 827-835.

Lomas, J., Martin, S. and Claxton, K., 2018. *Estimating the marginal productivity of the English National Health Service from 2003/04 to 2012/13*. Centre for Health Economics, University of York.

Macinko, J. A., Shi, L., Starfield, B. and Wulu Jr, J. T., 2003. Income inequality and health: A critical review of the literature. *Medical Care Research and Review,* 60 (4), 407-452.

Martin, S., Rice, N. and Smith, P. C., 2008. Does health care spending improve health outcomes? Evidence from English programme budgeting data. *Journal of Health Economics,* 27 (4), 826-842.

Martin, S., Rice, N. and Smith, P. C., 2012. Comparing costs and outcomes across programmes of health care. *Health Economics,* 21 (3), 316-337.

National Audit Office, 2015. *Mapping between different commissioners of healthcare* [Online]. National Audit Office. Available: www.nao.org.uk/audit-insights [Accessed 06/10].

NHS England, 2014. *Five year forward view* [Online]. Available: https://www.england.nhs.uk/five-year-forward-view/ [Accessed 01/10].

NHS England, 2015. *CCG Outcomes Indicator Set 2015/16: Technical Guidance*. NHS England.

O'Conor, R., 2013. *Mortality statistics for England 2005-2011: by a range of commissioning organisations. Used in the preparation of SPOT 2013 (Spend and Outcome Tool)*. Public health Engalnd Knowledge and Intelligence Team (Northern and Yorkshire) Available: https://www.gov.uk/government/publications/spend-and-outcome-tool-spot.

Pelone, F., Kringos, D. S., Romaniello, A., Archibugi, M., Salsiri, C. and Ricciardi, W., 2014. Primary care efficiency measurement using data envelopment analysis: A systematic review. *Journal of Medical Systems,* 39 (1), 1-14.

Pickett, K. E. and Wilkinson, R. G., 2015. Income inequality and health: a causal review. *Social Science & Medicine,* 128 316-326.

Simar, L. and Wilson, P. W., 2002. Non-parametric tests of returns to scale. *European Journal of Operational Research,* 139 (1), 115-132.

Simar, L. and Wilson, P. W., 2011. Two-stage DEA: caveat emptor. *Journal of Productivity Analysis,* 36 (2), 205.

8. APPENDIX 1. PROGRAMME BUDGET CATEGORIES

The English Department of Health hast established the national Programme Budgeting Project in order to organized information regarding the allocation of the health care resources by specific disease areas. PCTs provided a breakdown of their expenditure on specific healthcare conditions. This condition are called Programme Budget Categories. There categories were based on the World Health Organisation International Classification of Diseases.

Table 11. Programme Budget Categories (PBCs)

PBC CODE	PBC NAME	PBC CODE	PBC NAME
PBC 1	**Infectious diseases**	**PBC 11**	**Problems of the respiratory system**
01a	HIV and AIDS	11a	Obstructive airways disease
01x	Infectious diseases (Other)	11b	Asthma
PBC 2	**Cancers and tumours**	11x	Problems of the respiratory system (other)
02a	Head or neck cancers	**PBC 12**	**Dental problems**
02b	Upper gastro intestinal cancers	**PBC 13**	**Problems of the gastro intestinal system**
02c	Lower gastro intestinal cancers	13a	Upper gastro intestinal system problems
02d	Lung cancers	13b	Lower gastro intestinal system problems
02e	Skin cancers	13c	Hepatobiliary problems
02f	Breast cancers	13x	Problems of the gastro intestinal system (other)
02g	Gynaecological cancers	**PBC 14**	**Problems of the skin**
02h	Urological cancers	14a	Burns
02i	Haematological cancers	14x	Problems of the skin (other)
02x	Cancers and tumours (other)	**PBC 15**	**Problems of the musculoskeletal system**
PBC 3	**Disorders of blood**	**PBC 16**	**Problems due to trauma and injuries**
PBC 4	**Endocrine, nutritional and metabolic problems**	**PBC 17**	**Problems of the genito urinary system**
04a	Diabetes	17a	Genital tract problems
04b	Endocrine	17b	Renal problems
04x	Endocrine, nutritional and metabolic problems (other)	17c	Sexually transmitted infections
PBC 5	**Mental health disorders**	17x	Problems of genito urinary system (other)
05a	Substance misuse	**PBC 18**	**Maternity and reproductive health**
05b	Organic mental disorders	**PBC 19**	**Conditions of neonates**
05c	Psychotic disorders	**PBC 20**	**Adverse effects and poisoning**
05d	Child and adolescent mental health disorders	20a	Unintended consequences of treatment
05x	Mental health disorders (other)	20b	Poisoning
PBC 6	**Problems of learning disability**	20c	Violence
PBC 7	**Neurological**	20x	Adverse effects and poisoning (other)
07a	Chronic pain	**PBC 21**	**Healthy individuals**
07x	Neurological (other)	**PBC 22**	**Social care needs**
PBC 8	**Problems of vision**	**PBC 23**	**Other**
PBC 9	**Problems of hearing**	23a	GMS/PMS
PBC 10	**Problems of circulation**	23x	Miscellaneous
10a	Coronary heart disease		
10b	Cerebrovascular disease		
10c	Problems of rhythm		
10x	Problems of circulation (other)		

Source: NHS England, 2018. Programme Budgeting. Accessed 27/06/2018. https://www.england.nhs.uk/resources/resources-for-ccgs/prog-budgeting/

9. APPENDIX 2. METHODOLOGY FOR THE DEA: FRIED ET AL. (2002) THREE STEPS PROCEDURE

9.1. First Step: Initial DEA

In the first step, the DEA is applied by including all selected health outcomes and inputs, but excluding the environmental variables. In this step initial measures of the PCTs efficiency scores are obtained. The variable returns to scale DEA used here can be expressed as the following linear programming problem:

$$min_{\theta,\lambda} \quad \theta$$

subject to

$$\theta x_i^q \geq \sum_{k=1}^{K} \lambda^k x_i^k$$

$$\sum_{k=1}^{K} \lambda^k y_j^k \geq y_j^q$$

$$\lambda^k \geq 0$$

$$\sum_{k=1}^{K} \lambda^k = 1 \qquad (9.1)$$

Where $x_i^k \geq 0$ (k =1,…,K and i=1,…,I) is the input i used by PCT k, and $y_j^k \geq 0$ is the health outcome j (j=1,…,J) produce by PCT k. λ^k is the weight given to PCT k in its comparison with PCT q and θ is the efficiency score of PCT q. A $\theta = 0$ means that PCT q is fully efficient. In the first step of the analysis an efficiency score is estimated for each PCT.

During this step, the DEA is estimated twice. First, all PCTs are included and a set of θ^k (k= 1,…,K) are estimated. Second, after the identification and exclusion of the outliers (see Section 2.1.2), the DEA expressed in equation (9.1) is re-estimated to obtain a new set of efficiencies scores that are not affected by the outliers.

9.2. Second Step: Stochastic frontier analysis (SFA)

Fried et al. (2002) assumed that inefficiencies can be obtained from the first stage by observing the so called slacks. They defined inputs slack as:

$$s_{qi} = \left[x_i^q - \sum_{k=1}^{K} \lambda^k x_i^k\right] \qquad (9.2)$$

According to Fried et al. (2002), these first estimated inefficiencies can be split into three elements: (1) environmental influences, (2) pure managerial inefficiencies, and (3) statistical noise associated to measurement errors in inputs and/or health outcomes used to generate the first stage slacks. By using (9.2) a set of slacks can be estimated for each input included. This allows the split of these three effects for each input slack, and so, to estimate the effect that environmental variables in each input slack.

If we assume that the slacks estimated in (9.2) are explained in part by the effect of two environmental variables, is it possible to estimate one equation for each input i (i=1,…,I) where environmental variables and error terms varies across K PCTs (k=1,….,K):

$$s_{ki} = \beta_{i0} + \beta_{i1} DepIndex_k + \beta_{i2} TargetDistance_k + \varepsilon_{ki} \qquad (9.3)$$

Where s_{ki} is the slack of PCT k *(k= 1,...,K)* for input i (i=1,…,I).

In addition to the effect of the environmental variables, Fried et al. (2002) suggest to take advantage of the particularities of the SFA to be able of splitting the error term ε_{ki} into two elements: the statistical noise and the inefficiency. SFA is a parametric approach used to estimate production or cost functions, while explicitly accounting for the presence of producers' inefficiency. The SFA assume that in case of inefficiency in

the production, the error term ε_i estimated in (9.3) is actually a composed error that reflect two elements v_{ki} and u_{ki} .

The first element, v_{ki}, reflects statistical noise and is distributed as $v_{ki} \sim N(0, \sigma_{vi}^2)$; while the second one, u_{ki}, reflect managerial inefficiency and is distributed as $u_{ki} \sim N_+(0, \sigma_{ui}^2)$. If we assume that v_{ki} and u_{ki} are distributed independently of each other and of the environmental variables, we can estimated I regressions (i=1 …,I) using maximum likelihood techniques. In the analysis presenting in this analysis, two environmental variables are considered, therefore the equations to be estimated are:

$$s_{ki} = \beta_{i0} + \beta_{i1} DepIndex_k + \beta_{i2} TargetDistance_k + v_{ki} - u_{ki} \quad (9.4)$$

The minimum slack that can be achieve in a noise environment, characterized by variables $(DepIndex_k, TargetDistance_k, v_{ki})$ and parameters $(\beta_{i0}, \beta_{i1}, \beta_{i2}, \sigma_{vi}^2)$, corresponds to the first part of equation (9.4): $(\beta_{i1} DepIndex_k + \beta_{i2} TargetDistance_k + v_{ki})$. Any slacks in excess are attributed to managerial inefficiencies, and captured by the nonnegative error component u_{ki}, with parameters (μ_i, σ_{ui}^2), that reflects the variability of managerial inefficiencies across producers and inputs.

A first estimation of the SFAs expressed in (9.4) was done. Tests for the normally of the errors, heteroscedasticity and multicollinearity were considered. The Breusch-Pagan test and the Goldfeld-Quandt test indicated that statistically significant heteroscedasticity was presented in each one of the SFA estimated. Some author suggests that when heteroscedasticity exists correcting for it leads not only to a substantial improvement of the statistical properties of estimators but also to improved efficiency and ranking measures. Therefore, instead of the original input slacks, a log transformation of s_{ki} is used to estimate the SFA.

$$\ln(s_{ki}) = \beta_{i0} + \beta_{i1} DepIndex_k + \beta_{i2} TargetDistance_k + v_{ki} - u_{ki} \quad (9.5)$$

Once equation 7.5 is estimated by applying the SFA, the results are used to adjust PCTs´ inputs, such that it allows for "levelling the playing field" in which the PCTs efficiencies are estimated. In this way, those PCTs that are benefited from the environmental conditions will not, for this reason, appears as having higher levels of efficiency.

It is important to highlight that original inputs slacks can be expressed as:

$$s_{ki} = e^{\beta_{i0}} * e^{\beta_{i1} DepIndex_k} * e^{\beta_{i2} TargetDistance_k} * e^{v_{ki}} * e^{-u_{ki}} \quad (9.6)$$

Those PCTs that are in an advantaged position because of a relatively favourable environment or because a relatively better luck (represented by the statistical noise), will have their inputs adjusted upward in a proportion that represent their level of advantage. In order to estimate this proportion, the PCT that is in the worst situation is identified. The differences between the most disadvantage PCT and all others PCTs are estimated for each element:

$$ADDepIndex_{ik} = max_k \left[e^{\beta_{i1} DepIndex_k}\right] - e^{\beta_{i1} DepIndex_k} \quad (9.7)$$

$$ADTargetDistance_{ik} = max_k \left[e^{\beta_{i2} TargetDistance_k}\right] - e^{\beta_{i2} TargetDistance_k} \quad (9.8)$$

$$ADRandomNoise_{ik} = max_k \left[e^{v_{ki}}\right] - e^{v_{ki}} \quad (9.9)$$

with k = 1,…,K and i = 1,…,I

In order to estimate equation (9.9) is necessary to separate ε_{ki} of equation (9.3) in the two composed elements: statistical noise and managerial inefficiencies. From the

conditional estimators for managerial inefficiency, it is possible to obtain the statistical noise residual by considering:

$$\hat{E}[v_{ki} \mid v_{ki} - u_{ki}] = \ln(s_{ki}) - \beta_{i0} - \beta_{i1} DepIndex_k - \beta_{i2} TargetDistance_k - \hat{E}[u_{ki} \mid v_{ki} - u_{ki}] \quad (9.10)$$

This provide conditional (on $v_{ki} - u_{ki}$) estimators of v_{ki}. According to Bogetoft and Otto (2011), it is possible to estimate:

$$\hat{E}[u_{ki} \mid v_{ki} - u_{ki}] = \mu_* + \sigma_* \frac{\phi(\mu_*/\sigma_*)}{\Phi(\mu_*/\sigma_*)}$$

Where $\delta_i = \sqrt{\frac{\sigma_{ui}^2}{\sigma_{vi}^2}}$

$$\mu_* = -\varepsilon_{ki} \frac{\delta_i{}^2}{(1+\delta_i{}^2)}$$

$$\sigma_* = \frac{\delta_i}{(1+\delta_i{}^2)} \sigma^2 \quad (9.11)$$

With ε_{ki} distributed as $\varepsilon_{ki} \sim N(0, \sigma_i^2)$.

Equations (9.7), (9.8) and (9.9) shows the differences between the PCTs in terms of environment and "lucky" conditions that need to be adjusted in order to have the field levelled as the level of the most disadvantage PCT. With this in mind, adjusted inputs are estimated as follow:

$$xAD_i^k = x_i^k + ADDepIndex_{ik} + ADTargetDistance_{ik} + ADRandomNoise_{ik} \quad (9.12)$$

9.3. Third Step: Adjusted DEA

In the last step, xAD_i^k are used to estimate a new set of efficiencies scores (θAD_k), using the following linear programming problem:

$$min_{\theta AD, \lambda AD} \quad \theta$$

subject to

$$\theta AD x_i^q \geq \sum_{k=1}^{K} \lambda AD^k xAD_i^k$$

$$\sum_{k=1}^{K} \lambda AD^k y_j^k \geq y_j^q$$

$$\lambda Ad^k \geq 0$$

$$\sum_{k=1}^{K} \lambda AD^k = 1 \quad (9.13)$$

10. APPENDIX 3. DEA EFFICIENCY SCORES

Table 12. Efficiency Scores - DEA

PCT code	PCTname	# Efficient PBCs	Mental Health	Maternity	Cancer	Gastrointestinal	Cardiovascular	Respiratory	Endocrine
5A3	South Gloucestershire PCT	5		**1.00**	1.00	**1.00**	**1.00**	**1.00**	**1.00**
5A4	Havering PCT	2	**1.00**	0.98	0.94	0.98	0.96	0.96	**1.00**
5A5	Kingston PCT	2	**1.00**	1.00	**1.00**	0.97	0.97	0.95	0.94
5A7	Bromley PCT	2	0.96	0.95	0.95	0.97	**1.00**	**1.00**	0.93
5A8	Greenwich Teaching PCT	3	0.91	0.85	**1.00**	1.00		**1.00**	**1.00**
5A9	Barnet PCT	5	0.92	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.95
5AT	Hillingdon PCT	2	**1.00**	0.97	0.97	0.98	**1.00**	1.00	0.97
5C1	Enfield PCT	4	0.92	0.99	**1.00**	**1.00**	**1.00**	**1.00**	0.98
5C2	Barking and Dagenham PCT	1	0.97	0.91	**1.00**	0.96	0.95	0.94	0.96
5C3	City and Hackney Teaching PCT	3	**1.00**	0.90	**1.00**		**1.00**		0.95
5C4	Tower Hamlets PCT	2	**1.00**	0.82	0.97	0.93		**1.00**	0.89
5C5	Newham PCT	5	**1.00**	0.83	**1.00**	**1.00**	**1.00**	**1.00**	0.90
5C9	Haringey Teaching PCT	3	0.83	0.89	**1.00**	**1.00**	**1.00**		0.96
5CN	Herefordshire PCT	3	**1.00**	0.97	**1.00**	0.98		**1.00**	0.98
5CQ	Milton Keynes PCT	0	0.97	0.99	0.95	0.98	0.96	0.96	0.98
5D7	Newcastle PCT	0	0.91		0.95	0.96		0.94	0.99
5D8	North Tyneside PCT	1	0.98	**1.00**	0.96	0.98		0.96	0.98
5D9	Hartlepool PCT	1	1.00	**1.00**	0.98	0.97		0.93	
5E1	Stockton-on-Tees Teaching PCT	1	**1.00**	0.97	0.92	0.97			0.99
5EF	North Lincolnshire PCT	1	0.98	0.99	0.94	0.98	**1.00**	0.97	0.98
5EM	Nottingham City PCT	3	0.96	0.92	0.93	**1.00**		**1.00**	**1.00**
5ET	Bassetlaw PCT	2	**1.00**	0.99	0.96	0.98	1.00	**1.00**	0.98
5F1	Plymouth Teaching PCT	1	0.96	0.96	0.95	0.99	0.96	0.98	**1.00**
5F5	Salford PCT	0	0.91	0.94	0.95	0.95	0.97	0.93	0.96
5F7	Stockport PCT	2			0.99	**1.00**	0.96	**1.00**	
5FE	Portsmouth City Teaching PCT	1	0.95	0.94	0.99	0.97	0.95	0.96	**1.00**
5FL	Bath and North East Somerset PCT	3	0.97	**1.00**	0.95	0.98	**1.00**	**1.00**	0.96
5GC	Luton Teaching PCT	1	0.92	0.93	**1.00**	0.97	0.95	0.95	0.97
5H1	Hammersmith and Fulham PCT	1	0.98	0.94	0.99	0.97	**1.00**	0.95	0.97
5H8	Rotherham PCT	1	**1.00**	0.96	0.95	1.00	0.96	0.98	
5HG	Ashton, Leigh and Wigan PCT	0	0.99	0.97	0.94	0.96	0.95	0.96	0.97
5HP	Blackpool PCT	0	0.86	0.96	0.97	0.98	0.94	0.89	0.96
5HQ	Bolton PCT	6		**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
5HX	Ealing PCT	2	0.98	0.94	**1.00**	0.96	**1.00**	0.98	0.95
5HY	Hounslow PCT	1	0.97	0.94		0.99	**1.00**	0.99	0.96
5J2	Warrington PCT	0	0.95	0.96	0.96	0.98	0.95	0.96	0.97
5J4	Knowsley PCT	1	0.91	0.93	**1.00**	0.93	0.95	0.86	0.93
5J5	Oldham PCT	0	0.93	0.94	0.96	0.98	0.94	0.90	0.95
5J6	Calderdale PCT	0	0.93	0.93	0.95	0.96		0.93	0.97
5J9	Darlington PCT	0	0.90	0.93	0.97	0.96		0.92	0.98
5JE	Barnsley PCT	0	0.93	0.98	0.93	0.97		0.95	0.99
5JX	Bury PCT	6	**1.00**	**1.00**	**1.00**	**1.00**	0.96	**1.00**	**1.00**
5K3	Swindon PCT	5	**1.00**	**1.00**	0.98	**1.00**	**1.00**	**1.00**	0.99

5K5	Brent Teaching PCT	0	0.95	0.90	0.96	0.96	0.94	0.96	0.92
5K6	Harrow PCT	1	0.96	0.95	**1.00**	0.98	0.96	0.97	0.95
5K7	Camden PCT	1	0.85	0.93	0.98	0.96	**1.00**	0.99	0.94
5K8	Islington PCT	3	0.75	0.88	**1.00**	0.99	**1.00**	**1.00**	0.91
5K9	Croydon PCT	3	0.94	0.92	**1.00**	0.96	**1.00**	**1.00**	0.98
5KF	Gateshead PCT	1	0.95	0.95	**1.00**	0.96		0.94	0.98
5KG	South Tyneside PCT	0	0.98	0.99	0.95	0.99	0.96	0.97	1.00
5KL	Sunderland Teaching PCT	2	0.99	0.98	**1.00**	1.00		0.98	**1.00**
5KM	Middlesbrough PCT	0	0.91	0.91	0.96	0.96		0.92	0.97
5L1	Southampton City PCT	4	0.96	**1.00**	**1.00**	**1.00**	**1.00**	0.96	
5L3	Medway Teaching PCT	0	0.95	0.95	0.95	0.97		0.97	0.96
5LA	Kensington and Chelsea PCT	4	0.78	0.86	**1.00**	**1.00**	**1.00**	**1.00**	0.91
5LC	Westminster PCT	2	0.85	0.85	**1.00**	0.97	0.95	**1.00**	0.90
5LD	Lambeth PCT	1	0.76	0.79	**1.00**	0.93		0.98	0.93
5LE	Southwark PCT	0	0.90	0.91		0.97		0.99	0.98
5LF	Lewisham PCT	4	**1.00**	0.99	0.99	**1.00**	**1.00**	**1.00**	0.99
5LG	Wandsworth PCT	2	0.98	0.92	**1.00**	0.99	**1.00**	0.98	0.94
5LH	Tameside and Glossop PCT	5	**1.00**	**1.00**	0.95	**1.00**	0.95	**1.00**	**1.00**
5LQ	Brighton and Hove City PCT	6	**1.00**	**1.00**	0.97	**1.00**	**1.00**	**1.00**	**1.00**
5M1	South Birmingham PCT	2	0.93	0.96	0.96	**1.00**	0.98	**1.00**	0.98
5M2	Shropshire County PCT	4	**1.00**		0.95	**1.00**	0.96	**1.00**	**1.00**
5M3	Walsall Teaching PCT	0	0.93	0.92	0.94	0.96	0.96	0.92	0.95
5M6	Richmond and Twickenham PCT	5	**1.00**	0.91	**1.00**	0.98	**1.00**	**1.00**	**1.00**
5M7	Sutton and Merton PCT	3	**1.00**	0.96	**1.00**	0.98	**1.00**	0.97	0.99
5M8	North Somerset PCT	3	**1.00**	0.98	0.94	0.99	0.96	**1.00**	**1.00**
5MD	Coventry Teaching PCT	1	0.92	0.92	0.93	0.98	**1.00**	0.98	0.96
5MK	Telford and Wrekin PCT	2	0.95	0.95	0.93	0.97	**1.00**	**1.00**	0.98
5MV	Wolverhampton City PCT	0	0.97	0.94	0.94	0.99	0.99	0.99	0.96
5MX	Heart of Birmingham Teaching PCT	0	0.80	0.83	0.97		0.94	0.97	0.87
5N1	Leeds PCT	3	**1.00**	**1.00**	0.94	0.99		**1.00**	0.99
5N2	Kirklees PCT	0	0.97	0.95	0.96	0.96	0.95	0.95	0.96
5N3	Wakefield District PCT	0	0.93	0.95	0.93	0.96	0.95	0.92	0.97
5N4	Sheffield PCT	0	0.95	0.98	0.94	0.97		0.97	0.99
5N5	Doncaster PCT	1	0.90	0.96	0.94	0.98		0.94	**1.00**
5N6	Derbyshire County PCT	2	**1.00**	1.00	0.92	0.98	**1.00**	0.96	0.99
5N7	Derby City PCT	4	**1.00**	0.99	**1.00**	**1.00**			**1.00**
5N8	Nottinghamshire County PCT	1	0.98	**1.00**	0.94	0.98	0.98	0.98	0.98
5N9	Lincolnshire PCT	1	0.96	**1.00**	0.91	0.98	0.95	0.96	0.99
5NA	Redbridge PCT	3	**1.00**	1.00		**1.00**	0.95	**1.00**	0.99
5NC	Waltham Forest PCT	3	**1.00**	0.97	**1.00**	0.99	**1.00**	0.98	0.97
5ND	County Durham PCT	2	0.96	**1.00**	1.00	0.98	**1.00**	0.96	0.99
5NE	Cumbria PCT	2	0.97	**1.00**	0.95	0.98	0.96	0.96	**1.00**
5NF	North Lancashire PCT	3	**1.00**	**1.00**	0.93	**1.00**	0.96	0.98	0.98
5NG	Central Lancashire PCT	0	0.98	0.97	0.97	0.98	0.96	0.96	0.98
5NH	East Lancashire PCT	0	0.91	0.93	0.95	0.96	0.96	0.91	0.95
5NJ	Sefton PCT	1	0.95	**1.00**	0.93	0.95	0.95	0.93	0.95
5NK	Wirral PCT	0	0.92	0.94	0.95	0.95	0.95	0.94	0.94
5NL	Liverpool PCT	1	0.91	**1.00**	0.97	0.95	0.95	0.91	0.93
5NM	Halton and St Helens PCT	0	0.92	0.95	1.00	0.95	0.98	0.88	0.96

5NN	Western Cheshire PCT	0	0.93	0.99	0.94	0.97	0.95	0.96	0.96
5NP	Central and Eastern Cheshire PCT	5	**1.00**	**1.00**	**1.00**	0.99	0.99	**1.00**	**1.00**
5NQ	Heywood, Middleton and Rochdale PCT	0	0.91	0.93	0.95	0.98	0.95	0.96	0.95
5NR	Trafford PCT	0	0.98	0.95	0.95	0.97	0.97	0.95	0.96
5NT	Manchester PCT	1	0.91	0.96	0.99	0.96	**1.00**	0.94	0.95
5NV	North Yorkshire and York PCT	7	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
5NW	East Riding of Yorkshire PCT	4			**1.00**	**1.00**	0.96	**1.00**	**1.00**
5NX	Hull PCT	0	0.93	0.92	0.94	0.95	0.97	0.92	0.97
5NY	Bradford and Airedale PCT	0	0.91	0.91	0.94	0.94		0.94	0.94
5P1	South East Essex PCT	7	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
5P2	Bedfordshire PCT	3	**1.00**	**1.00**	0.95	0.98	**1.00**	0.98	0.98
5P5	Surrey PCT	1	0.96	**1.00**	0.99	0.98	0.98	0.97	0.96
5P6	West Sussex PCT	1	0.96	0.98	0.98	0.98	**1.00**	0.96	0.96
5P7	East Sussex Downs and Weald PCT	4	0.96	**1.00**	0.93	**1.00**	**1.00**	0.95	**1.00**
5P8	Hastings and Rother PCT	1	0.95	**1.00**	0.93	0.95	0.98	0.93	0.95
5P9	West Kent PCT	3	**1.00**	**1.00**	**1.00**	0.99	0.97	0.98	0.98
5PA	Leicestershire County and Rutland PCT	4	**1.00**	**1.00**	0.99	**1.00**	0.97	**1.00**	0.99
5PC	Leicester City PCT	4	0.95	0.93	**1.00**	**1.00**	**1.00**	**1.00**	0.99
5PD	Northamptonshire PCT	2	0.93	**1.00**	0.98	0.98	**1.00**	0.95	0.97
5PE	Dudley PCT	1	**1.00**	0.96	0.92	0.99	0.96	0.98	0.98
5PF	Sandwell PCT	0	0.95	0.95	0.95	0.98	0.97	0.96	0.96
5PG	Birmingham East and North PCT	2	0.91	0.92	**1.00**	**1.00**		0.99	0.95
5PH	North Staffordshire PCT	2	**1.00**	0.97	0.93	0.98	**1.00**	0.95	0.97
5PJ	Stoke on Trent PCT	2	**1.00**	0.93	0.96	0.99	**1.00**	0.92	0.97
5PK	South Staffordshire PCT	3	**1.00**	**1.00**	0.92	0.98	0.97	**1.00**	0.98
5PL	Worcestershire PCT	3	**1.00**	1.00	0.96		0.98	**1.00**	**1.00**
5PM	Warwickshire PCT	3	**1.00**	**1.00**	**1.00**	0.98	0.98	0.99	0.98
5PN	Peterborough PCT	1	0.85	0.91	**1.00**	0.96		0.92	0.94
5PP	Cambridgeshire PCT	1	0.99	0.98	0.99	**1.00**	0.97	0.99	0.98
5PQ	Norfolk PCT	4	0.96	**1.00**		0.99	**1.00**	**1.00**	**1.00**
5PR	Great Yarmouth and Waveney PCT	1	**1.00**	0.97	0.94	0.97		0.95	0.99
5PT	Suffolk PCT	6	**1.00**	**1.00**	0.95	**1.00**	**1.00**	**1.00**	**1.00**
5PV	West Essex PCT	0	0.97	0.95	0.95	0.98	0.96	0.96	0.96
5PW	North East Essex PCT	5	**1.00**	1.00	0.96	**1.00**	**1.00**	**1.00**	**1.00**
5PX	Mid Essex PCT	3	**1.00**	**1.00**	0.95	**1.00**		1.00	0.98
5PY	South West Essex PCT	0		0.96	0.96	0.98	0.95	0.96	0.96
5QA	Eastern and Coastal Kent PCT	4	0.92	**1.00**	**1.00**	**1.00**	0.99	**1.00**	0.96
5QC	Hampshire PCT	5			**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
5QD	Buckinghamshire PCT	5	**1.00**	0.98	**1.00**	**1.00**	**1.00**	**1.00**	0.98
5QE	Oxfordshire PCT	3	**1.00**	0.98	**1.00**	0.99	**1.00**	0.99	1.00
5QF	Berkshire West PCT	3	**1.00**	0.98	0.97			**1.00**	**1.00**
5QG	Berkshire East PCT	2	**1.00**	0.96	0.99	**1.00**	0.98	0.97	0.97
5QH	Gloucestershire PCT	1	**1.00**	0.98	0.96	0.98	0.96	0.95	0.98
5QJ	Bristol PCT	4	0.94	0.96	**1.00**	1.00	**1.00**	**1.00**	**1.00**
5QK	Wiltshire PCT	3	**1.00**	0.98	0.95	0.98	**1.00**	**1.00**	0.98
5QL	Somerset PCT	1	**1.00**	0.97	0.93	0.98	0.98	0.97	0.98
5QM	Dorset PCT	3	1.00	**1.00**	0.95	0.98	**1.00**	0.99	**1.00**
5QN	Bournemouth and Poole PCT	1	0.97	0.96	0.98	0.99	0.95	**1.00**	0.99

5QP	Cornwall and Isles of Scilly PCT	5	**1.00**	**1.00**	1.00	**1.00**	**1.00**		**1.00**
5QQ	Devon PCT	0	0.97	0.98	0.94	0.99	0.99	0.99	
5QR	Redcar and Cleveland PCT	0	1.00	0.96	0.94	0.97		0.95	0.97
5QT	Isle of Wight Healthcare PCT	4		**1.00**	0.94	**1.00**	**1.00**	0.99	**1.00**
5QV	Hertfordshire PCT	6	**1.00**	**1.00**	**1.00**	**1.00**	0.98	**1.00**	**1.00**
TAC	Northumberland Care Trust	3	**1.00**	**1.00**	0.94	0.98		0.97	**1.00**
TAK	Bexley Care Trust	1	0.97	0.99	**1.00**	0.98		0.99	0.98
TAL	Torbay Care Trust	2	0.90	**1.00**	0.98	0.95	**1.00**	0.94	0.98
TAM	Solihull PCT	2	**1.00**	0.99	0.94	**1.00**		0.97	0.99
TAN	North East Lincolnshire PCT	0	0.84	0.92	0.97	0.92	0.95	0.85	0.96
TAP	Blackburn with Darwen Teaching Care Trust Plus	0	0.85	0.92	0.95	0.91	0.95	0.81	0.97

11. APPENDIX 4. QUANTILE REGRESSIONS

The specification of the model estimated on the QR are taken from the preferred outcome specification in Lomas et al. (2018) for the 6 selected PBCs.

The dependent variable is the three-year average SYLLR as presented by ONS Compendium Statistics for years 2012-2013-2014, These same data are used by Lomas et al. (2018) to define their dependent variable, the only difference is the mapping used and the final geographical area chosen to represent mortality outcomes. Lomas et al. (2018) uses original data at top-tier local authority (LA) for 152 LAs (unitary authority, metropolitan district, London borough, counties). They also use these mortality data mapped to PCT-level area. We use original data for 326 LAs (Local authority districts, unitary authority, metropolitan district, London borough) and map these to PCT-level according to the mapping method described based on 2011 Census population.

Regarding the explanatory variables, we use for each of the 6 PBCs the same explanatory variables as used by Lomas et al. (2018), except some differences in the measurement year for CARAN and HIV needs. These explanatory variables and their descriptive statistics have been presented in Table 5 and Table 6.

The final instruments used by Lomas, Martin and Claxton (2018) are not available, so that we have chosen several specifications from the instruments presented in Table 5 (deprivation and socioeconomic variables) to achieve overidentification, meeting the overidentification test which indicates the validity of the instruments. These final instruments used are detailed in results presented in this Appendix.

The following tables (Table 13 – Table 18 inclusive) detail the estimations presented in Figure 1 to Figure 26. Results published by Lomas et al. (2018) are presented in the first column for the sake of comparison. Estimates of the conditional mean model are presented unweighted to use as benchmark of comparison with QR estimates, since QR estimation does not allow weights. Mean estimates are also weighted to consider different PCT size as Lomas, Martin and Claxton (2018) do. Most of the explanatory variables in out model coincide with those selected in York team' preferred specification model, although our variables on health needs (HIV need and CARAN need refer to year 2011/12 instead to 2012/13 in Lomas et al.'s estimations. Figures 19 to 24 represent an horizontal line for the conditional mean estimate (unweighted) of the outcome elasticity to PBC spend (coefficient of variable lg`PBC'_1213netpoppheadOHP) in each one of the six PBCs. The blue line links the QR estimates of this outcome elasticity in each table. Figures also show the 95% CI corresponding to the estimated standard deviation (in brackets).

Table 13 QR Estimates for PBC 1 Infectious Disease

Lomas, Martin and Claxton (2018)		**ESTIMATES**						
PBC 1 Infectious		**PBC 1 Infectious**						
2012/13 spend		2012/13 spend						
SYLLR 2012/13/14		SYLL 2012/13/14						
Weighted		unweighted	weighted					
OLS mean		OLS mean		Quantile Regression				
LA- level		PCT-level						
	mean		mean	mean	q50	q10	q25	q75	q90
lLAg1_1213netpoppheadOHP	-0.362***	lg1_1213netpoppheadOHP	-0.3379***	-0.3218***	-0.4083**	-0.7593**	-0.4844**	-0.1497	-0.2248*
	[0.089]		[0.0981]	[0.0861]	[0.1249]	[0.2365]	[0.1598]	[0.1165]	[0.1061]
lLAHIVneedph	0.276***	lHIVneedprev	0.6851***	0.6667***	0.6809***	1.1703***	0.8145***	0.4355**	0.4858***
	[0.045]		[0.1107]	[0.1036]	[0.1159]	[0.2256]	[0.1839]	[0.1310]	[0.1100]
lIMD2010	0.649***	lIMD2010	0.4513***	0.4616***	0.4545***	0.3343***	0.3659**	0.5455***	0.6310***
	[0.064]		[0.0620]	[0.0526]	[0.0607]	[0.0886]	[0.1202]	[0.0682]	[0.0799]
lLONEPENH	-0.177	lLONE65andover	-0.0182	-0.0262	-0.2368	0.3559	-0.253	-0.0235	0.0615
	[0.183]		[0.1847]	[0.1775]	[0.1923]	[0.2193]	[0.3960]	[0.1618]	[0.1250]
_cons	0.698	_cons	1.4885**	1.3989**	1.2447*	3.7004***	1.6047	0.7468	1.0102*
	[0.437]		[0.5385]	[0.5065]	[0.5008]	[0.9784]	[1.0515]	[0.4750]	[0.4959]
N. Observations	147	N. Observations	151	151	151	151	151	151	151
R-squared	0.582	R-squared/Pseudo R-squared	0.587	0.612	0.394	0.291	0.333	0.427	0.406
		Test H0: spend coef=q50						*	
		Test H0 spend coef=q10						*	*
		Test H0 spend coef=q25						*	
		Test H0 spend coef=q75							

Notes:

Significance levels: * for p<.05, ** for p<.01, and *** for p<.001.

Table 14 QR Estimates for PBC 2 Cancer

Lomas, Martin and Claxton (2018) **PBC 2 Cancer** 2012/13 spend SYLLR 2012/13/14 instrument spend weighted IV second stage (GMM) LA-level		**ESTIMATES** **PBC 2 Cancer** 2012/13 spend SYLLR 2012/13/14 instrument spend PCT-level	unweighted	weighted IV second stage (GMM)	Quantile Regression				
	mean		mean	mean	q50	q10	q25	q75	q90
lLAg2_1213pheadOHP	-0.361**	lg2_1213netpoppheadOHP	-0.3447**	-0.4693*	-0.3845*	-0.0107	-0.1999*	-0.6669***	-0.7091***
	[0.149]		[0.1160]	[0.2076]	[0.1782]	[0.1189]	[0.0910]	[0.1288]	[0.1327]
lLACARANneed1213	1.023***	lCARANneed	0.6779***	0.6313***	0.6818***	0.7722***	0.6882***	0.6370***	0.6421***
	[0.134]		[0.0610]	[0.0920]	[0.0754]	[0.1203]	[0.0815]	[0.0460]	[0.0476]
_cons	6.744***	_cons	6.6738***	7.2532***	6.8596***	5.0199***	5.9565***	8.2265***	8.4428***
	[0.691]		[0.5408]	[0.9640]	[0.8374]	[0.5541]	[0.4306]	[0.6002]	[0.6251]
N. Observations	149	N. Observations	151	151	151	151	151	151	151
R-squared		Pseudo R-squared			0.44	0.38	0.39	0.49	0.51
Endogeneity test	8.48	Endogeneity test statistic	8.26	16.22					
Endogeneity p-value	0.004	Endogeneity p-value	0.00	0.00					
		Hansen-Sargan test	3.23	0.51					
		Hansen-Sargan p-value	0.07	0.47					
		Test H0: spend coef=q50				*			
		Test H0 spend coef=q10						***	***
		Test H0 spend coef=q25						***	***
		Test H0 spend coef=q75							

Notes:
Significance levels: * for p<.05, ** for p<.01, and *** for p<.001.
Instruments: lIMD2010, lLONEPARH

Table 15 QR Estimates for PBC 4 Endocrine

Lomas, Martin and Claxton (2018)			**ESTIMATES**						
PBC 4 Endocrine			**PBC 4 Endocrine**						
2012/13 spend			2012/13 spend						
SYLLR 2012/13/14			SYLLR 2012/13/14						
instrument spend									
weighted			unweighted	weighted					
IV second stage (GMM)			OLS		Quantile Regression				
LA-level		PCT-level							
	mean		mean	mean	q50	q10	q25	q75	q90
lLAg4_1213pheadOHP	-0.499	lg4_1213netpoppheadOHP	-0.2284	-0.2898	-0.4328*	0.3593	-0.3253	-0.3562	-0.2218
	[0.349]		[0.2103]	[0.1686]	[0.1920]	[0.7700]	[0.3003]	[0.2326]	[0.2459]
lLAIMD2010	0.579***	lIMD2010	0.5157***	0.4493***	0.5935***	0.6990*	0.4058**	0.4111***	0.4565***
	[0.116]		[0.0936]	[0.0896]	[0.0974]	[0.2951]	[0.1484]	[0.0669]	[0.1339]
LPROFOCCU	-0.409**	lPROFOCCU	-0.3104	-0.2662	-0.121	-0.3511	-0.4845*	-0.3435*	-0.6857**
	[0.165]		[0.1711]	[0.1516]	[0.1258]	[0.3964]	[0.2439]	[0.1330]	[0.2374]
_cons	1.118	_cons	0.311	0.8372	1.1304	-3.0991	0.6686	1.3434	0.3924
	[1.164]		[0.8606]	[0.7499]	[0.6593]	[3.3679]	[1.2368]	[0.9707]	[1.1931]
N. Observations	149	N. Observations	151	151	151	151	151	151	151
		R-/ Pseudo R-squared	0.37	0.37	0.22	0.17	0.21	0.22	0.22
		Test H0: spend coef=q50							
		Test H0 spend coef=q10							
		Test H0 spend coef=q25							
		Test H0 spend coef=q75							

Notes:

Significance levels: * for p<.05, ** for p<.01, and *** for p<.001.

Table 16 QR Estimates for PBC 10 Circulatory Disease

Lomas, Martin and Claxton (2018) **PBC 10 Circulatory** 2012/13 spend SYLLR 2012/13/14 instrument spend weighted IV second stage (GMM) LA-level		**ESTIMATES** **PBC 10 Circulatory** 2012/13 spend SYLLR 2012/13/14 instrument spend unweighted weighted IV second stage (GMM) PCT-level			Quantile Regression				
	mean		mean	mean	q50	q10	q25	q75	q90
lLAg10_1213pheadOHP	-1.464***	lg10_1213netpoppheadOHP	-1.4678***	-1.4941***	-1.4475***	-0.9682**	-1.3082***	-1.5593***	-1.7806***
	[0.268]		[0.3261]	[0.2906]	[0.2039]	[0.3221]	[0.2423]	[0.1360]	[0.1722]
lLACARANneed1213	2.304***	lCARANneed	1.0093***	0.8137***	0.9480***	1.1132***	0.9531***	1.0172***	0.8720***
	[0.234]		[0.1596]	[0.1811]	[0.0873]	[0.1572]	[0.0828]	[0.0856]	[0.1098]
_cons	11.541***	_cons	11.5613***	11.6850***	11.4676***	9.0168***	10.7230***	12.0688***	13.1948***
	[1.302]		[1.5853]	[1.4111]	[0.9852]	[1.5736]	[1.1873]	[0.6653]	[0.8405]
N. Observations	149	N. Observations	151	151	151	151	151	151	151
		Pseudo R2			0.52	0.43	0.49	0.55	0.51
		Endogeneity test	35.81	33.28					
		Endogeneity p-value	0.00	0.00					
		Hansen J test	1.26	0.99					
		Hansen J p-value	0.26	0.32					
		Test H0: spend coef=q50							*
		Test H0 spend coef=q10						*	*
		Test H0 spend coef=q25							
		Test H0 spend coef=q75							

Notes:
Significance levels: * for p<.05, ** for p<.01, and *** for p<.001.
Instruments: lIncomeScale, lIMD2010

Table 17 QR Estimates for PBC 11 Respiratory Disease

Lomas, Martin and Claxton (2018)			**ESTIMATES**						
PBC 11 Respiratory			**PBC 11 Respiratory**						
2012/13 spend			2012/13 spend						
SYLLR 2012/13/14			SYLLR 2012/13/14						
instrument spend			instrument spend						
weighted			unweighted	weighted					
IV second stage (GMM)			IV second stage (GMM)			Quantile Regression			
LA-level		PCT-level							
	mean		mean	mean	q50	q10	q25	q75	q90
lLAg11_1213pheadOHP	-1.704***	lg11_1213netpoppheadOHP	-1.6957*	-2.1179*	-1.5856*	-0.9237	-1.439	-2.1543***	-1.3774**
	[0.459]		[0.7284]	[0.8318]	[0.7186]	[1.0632]	[0.7364]	[0.5265]	[0.4438]
LPERMSICK11	6.265***	lPERMDISAB	6.2950***	6.8531***	5.6388**	5.3822*	5.3409**	7.1804***	5.5687***
	[1.189]		[1.6839]	[1.8566]	[1.8999]	[2.5876]	[1.6805]	[1.9052]	[1.0904]
LPERMSICK11SQ	0.742***	lPERMDISABSQ	0.8408**	0.9298**	0.7391*	0.7091	0.6885**	0.9847**	0.7289***
	[0.166]		[0.2586]	[0.2864]	[0.2963]	[0.3917]	[0.2599]	[0.2996]	[0.1672]
_cons	23.203***	_cons	22.2196***	24.9720***	20.6811***	16.9491	19.4580**	25.7621***	19.8558***
	[3.903]		[5.6845]	[6.2601]	[5.5990]	[8.9492]	[5.8788]	[4.6862]	[3.4653]
N. Observations	149	N. Observations	151	151	151	151	151	151	151
		Pseudo R2			0.45	0.4	0.4	0.45	0.49
		Endogeneity test	10.59	12.32					
		Endogeneity p-value	0.001	0.00					
		Hansen J test	4.49	4.13					
		Hansen J p-value	0.11	0.13					
		Test H0: spend coef=q50							
		Test H0 spend coef=q10							
		Test H0 spend coef=q25							
		Test H0 spend coef=q75							

Notes:

Significance levels: * for p<.05, ** for p<.01, and *** for p<.001.

Instruments: lIncomeScale, lFTSTUDEN, lOWNOCC

Table 18 QR Estimates for PBC 13 Gastrointestinal Disease

Lomas, Martin and Claxton (2018) **PBC 13 Gastrointestinal** 2012/13 spend SYLLR 2012/13/14 instrument spend weighted IV second stage (GMM) LA-level		**ESTIMATES** **PBC 13 Gastrointestinal** 2012/13 spend SYLLR 2012/13/14 instrument spend unweighted weighted IV second stage (GMM) PCT-level			Quantile Regression				
	mean		mean	mean	q50	q10	q25	q75	q90
lLAg13_1213pheadOHP	-1.904**	lg13_1213netpoppheadOHP	-1.6963**	-2.4014**	-1.3696*	-1.2347	-0.8539	-1.5829	-1.6361
	[0.897]		[0.6408]	[0.8588]	[0.6180]	[1.0271]	[0.6219]	[0.9133]	[0.8948]
lLACARANneed1213	3.878***	lCARANneed	2.0598***	1.8804***	2.1290***	1.7594***	2.0812***	2.2484***	1.9830***
	[0.832]		[0.2092]	[0.2634]	[0.1862]	[0.3444]	[0.2632]	[0.1834]	[0.2228]
lLACARANneed1213SQ	3.735***	lCARANneedSQ	3.8617**	5.1474***	4.4194**	3.2343	2.9991	4.1301*	1.5045
	[1.352]		[1.2252]	[1.4761]	[1.4028]	[1.9086]	[1.8306]	[1.6899]	[1.7174]
_cons	11.547***	_cons	10.6408***	13.7727***	9.1884**	8.2634	6.6781*	10.2801*	10.7062**
	[4.024]		[2.8755]	[3.8569]	[2.7585]	[4.6343]	[2.7959]	[4.0889]	[4.0108]
N. Observations	149	N. Observations	151	151	151	151	151	151	151
		Pseudo R2			0.40	0.28	0.33	0.40	0.40
		Endogeneity test	12.00	19.01					
		Endogeneity p-value	0.00	0.00					
		Hansen J test	4.88	1.13					
		Hansen J p-value	0.18	0.77					
		Test H0: spend coef=q50							
		Test H0 spend coef=q10							
		Test H0 spend coef=q25							
		Test H0 spend coef=q75							

Notes:
Significance levels: * for p<.05, ** for p<.01, and *** for p<.001.
Instruments: lFTSTUDEN, lPOPAllLTI, lPOP16_64LTI, lIMD2010