Regional variation in the productivity of the English National Health Service

Draft date: 08/11/2011

CHRIS BOJKE (a), ADRIANA CASTELLI (a), ANDREW STREET (a), PADRAIC WARD (a) and MAURO LAUDICELLA (b, c)

(a) Centre for Health Economics, University of York, York, UK

(b) School of Health Science, City University London, UK

(c) Business School and Centre for Health Policy, Imperial College, London, UK

Abstract

We assess the productivity of Strategic Health Authorities (SHA) in England in 2007/8. We identify areas of the country where expenditure could be reduced without affecting the number of patients treated or the quality of their care. Productivity is calculated by comparing the total amount of health care output to total inputs for each SHA. The amount of healthcare output comprises the number and type of patients treated and the quality of the care received. Healthcare input includes National Health Service (NHS) and agency staff, supplies, equipment and buildings. Data about healthcare outputs are derived from the Hospital Episode Statistics and Reference Cost returns. Input data derive from the Workforce Census and financial returns made by NHS organisations. Productivity varies from 5% above to 6% below the national average. Productivity is highest in South West SHA and lowest in East Midlands, South Central and Yorkshire & the Humber SHAs. The relative positions of SHAs hold irrespective of the data source used to measure inputs. If all SHAs were as productive as South West the NHS could reduce its expenditure by £3.2bn each year. Further research should examine the reasons why the South West is more productive than elsewhere and to elicit best practice.

Keywords: health sector, productivity; geographical variation

1. Introduction

Regional variation in the provision of health care has long held a prominent role in English health policy. This recognises that moving toward an equitable distribution of health care resources (Department of Health, 2007, Department of Health Resource Allocation Working Party, 1976) needs to be coupled with information about the use to which these resources are put. But, historically, comparisons of regional provision have been based on a selective set of indicators, such as crude counts of hospital activity or day case rates (Carter et al., 1992, NHS Executive, 1992, Thomson et al., 1997) or the so-called 'postcode lottery' in prescribing (Patel et al., 2007). While individually of interest, the picture that these indicators provide is inevitably partial. Measurement of regional productivity, by contrast, is designed both to capture the totality of health care provision and to relate this provision explicitly to resource use.

For public accountability it is crucial to measure whether public spending in the NHS or by its various organisations has achieved 'value for money' (Atkinson, 2006). Indeed, the current economic reality facing the NHS provides an even greater imperative for measuring and understanding regional variations in productivity – identifying areas in which potential efficiency savings can be made and understanding the expected impact (both geographical and organisational) on outputs of changes to inputs across regions.

To our knowledge only a limited body of work exists that has specifically looked at regional variations in overall health care productivity (Schleiniger, 2008, Yu and Ariste, 2010). The majority of the literature has restricted itself to focussing on specific sectors or specialities of the health care system (for example, mental health (Madianos et al., 1999) or hip replacements (Fisher et al., 2010) and diabetes treatment (Kristensen et al., 2010)) or on variations in costs or spending (Fisher et al., 2009, Skinner and Fisher, 2010). This paper aims to address this gap in the literature by taking a whole output and input approach and analysing the relative productivity of different geographical areas of the NHS. We do this by combining large and complex but routinely collected datasets namely the Hospital Episode Statistics (HES), reference cost returns, financial returns and workforce census.

The paper is structured as follows. The first section describes the methods used. Data are described in section two. The third section outlines our results and the final section concludes with a discussion of the results and suggestions for future work.

2. Methods

This paper compares productivity across Strategic Health Authorities (SHAs) using data for a single year: 2007/08. This requires meaningful, comprehensive and accurate measures of the volumes of output and input and comparison of these in an informative manner. To this purpose we follow the approach adopted in the construction of the national productivity index developed in Castelli et al. (2008) and Street and Ward (2009), adapting this for use in a cross-sectional context. The main steps are discussed below.

2.1 Productivity

Productivity is measured by comparing the total amount of health care 'output' produced to the total amount of 'input' used to produce this output. Output consists of all health care provision to NHS residents in an SHA and inputs include the staff, intermediate goods and services, and capital resources that contribute to the production of health care for the residents of that SHA. For a single year the standardised productivity of each SHA *s* *(s=1…10)* is defined as follows:

$$Standardised\ Productivity\ of\ SHA\ s\ = \left\{\left[\left(\frac{X_s}{Z_s}\right) \Big/ \left(\frac{\bar{X}}{\bar{Z}}\right)\right] - 1\right\} \times 100$$

Eqn. 1

Where X_s is the volume of output produced and Z_s the amount of input used in SHA s. $\bar{X}$ and $\bar{Z}$ are, respectively, the volumes of outputs and inputs averaged across all SHAs. The standardised productivity of each SHA is then given by dividing the SHA specific output/input ratio by the national average output/input ratio and expressing this as a percentage difference. Thus if standardised productivity in SHA s is 10%, this means that productivity is 10% higher than the national average.

2.2 Measuring output

The volume of output includes all health care services provided to NHS residents in each SHA. To account for the great diversity in the types of services provided by the NHS measure incorporates over 6,500 different health care output categories. In order to aggregate these categories into a meaningful measure of total output, it is necessary to take such diversity into account; after all, a blood test should not carry the same weight as a heart transplant. We thus calculate output by weighting each type of health service by its national average cost to reflect the relative value of different health care services provided within and across different settings. This approach is consistent with the national accounts convention as outlined in the Eurostat Handbook (Eurostat, 2001).

Recognising that costs do not equate with 'value' (Castelli et al., 2007b), we account for variation in the quality of services across SHAs. We define differences in quality as being captured by differences in survival rates, health outcomes and inpatient waiting times for hospital care across SHAs, and outpatient waiting times for outpatient activity across SHAs. The effect of adjusting for quality will be to scale output up (down) if an SHA's quality measures are higher (lower) than the national average.

We thus define the total output (X) for each SHA (*s=1…10*) as:

$$X_s = \sum_{j=1}^{J} x_{js}\bar{c}_j\bar{q}_{js}$$

Eqn. 2

Where x_j is the amount of activity of health service type *j* with *j*=1…*J*. The cost weights in equation (2) are defined as $\bar{c}_j = c_j/\hat{c}_j$ where c_j represents the national average cost of activity *j* and $\hat{c}_j$ is an arbitrarily chosen benchmark cost (we set this to be equal to the average cost of hospital inpatient treatment, amounting to £1,167). Finally, $\bar{q}_{js} = q_{js}/\hat{q}_j$, q_{js} is the quality of output *j* in SHA *s* and $\hat{q}_j$ is the national average quality of output *j*.

The characterisation of quality varies across healthcare settings, partly because activities in different settings have different quality characteristics but also because availability of data differs across settings. The quality adjustment that applies to hospital care provided to elective and non-elective patients and to those admitted to hospital with mental health problems takes the form[1]:

$$\bar{q}_{js} = \left(\frac{a_{js} - k_j}{\widehat{a_J - k_J}}\right)\left[\frac{\frac{\left(1 - e^{-r_Q LE_{js}^{In}}\right)}{r_Q} - \frac{\left(e^{r_w W_{js}} - 1\right)}{r_w}}{\frac{\left(1 - e^{-r_Q \widehat{LE}_j^{In}}\right)}{r_Q} - \frac{\left(e^{r_w \widehat{W}_j} - 1\right)}{r_w}}\right]$$

[1] This variant of the quality adjustment for hospital care is an adaptation of the quality adjustment component of the output index developed by Dawson et al. (2005) for longitudinal analysis to its use with cross sectional data.

Eqn. 3

This quality adjustment aims to capture differences across SHAs in quality-adjusted life years (QALYs) and in the time patients wait prior to hospital admission. However, direct QALY calculations for each output are not possible as information on the QALYs gained from treatment are unavailable – neither is the change in each patient's health status measured nor is it known for how long this change is experienced. To address this information deficit, we create the equivalent of a QALY profile for each type of hospital output (Castelli et al., 2007a):

- Firstly, we account for whether or not the patient survives treatment by measuring the 30-day post discharge survival rates for each output in each SHA, a_{js}.
- Secondly, we account for the ratio of average health status (h^0) before and after (h^*) treatment, $k_j = {h_j^0}/{h_j^*}$. In the absence of output-specific information we assume that, on average, the ratio for elective patients is twice that for non-elective patients (Dawson et al., 2004).
- Thirdly, we capture the duration of treatment benefit by estimating the life expectancy associated with each output, LE_{js}^{In}, by considering the age and gender profiles of patients having each treatment in each SHA. r_Q is the discount rate applied to future life years.

The final term in the above equation 3 captures changes in waiting times for each output, w_{js}, in recognition of the welfare loss associated with not being treated immediately. This formulation implies that the marginal disutility of waiting increases as the delay increases, with the disutility captured by the discount rate r_w (Dawson et al., 2005). Waiting time is measured at the 80th percentile of the waiting time distribution for each type of treatment. This recognises that reductions in relatively long waiting times confer benefits on all patients by reducing the risk of having to face a very long wait.

Similarly, the longer a patient has to wait for an outpatient appointment, the greater the disutility experienced. In recognition that shorter waiting times imply higher quality, outpatient activity is scaled up in SHAs where waiting times are lower than the national average.

Thus, in summary, total output will be higher than the national average in SHAs that have: higher volumes of activity; more complex or costly activities; higher rates of hospital survival; and lower inpatient and outpatient waiting times.

2.3 Measuring inputs

Inputs into the health care system consist of labour (doctors, nurses, technicians and managers), intermediate goods and services (drugs and clinical supplies) and capital (buildings and equipment with an asset life of more than a year). The use of these 'factors of production' can be calculated directly or indirectly.

A 'direct' measure of input can be calculated when data on the volume and price of inputs are available. Thus, for SHAs the total volume of inputs used, Z_s^D, is given by

$$Z_s^D = Z_s^L + Z_s^A + Z_s^M + Z_s^K$$

Eqn. 4

Where Z_s^L measures NHS labour inputs, Z_s^A captures the input of non-NHS (e.g. agency) staff, Z_s^M measures the use of intermediate inputs, and Z_s^K measures the use capital inputs. For example the total input of NHS staff in each SHA, Z_s^L, amounts to:

$$Z_s^L = \sum_{p=1}^{P} \sum_{n=1}^{N} z_{nps}^L \ \overline{\omega}_n$$

Eqn. 5

Where p=1…P represents all the organisations within the SHA, namely hospital (and foundation) trusts, community and mental health trusts, ambulance trusts and PCTs; z_{nps}^L is the volume of NHS staff of type n and $\overline{\omega}_n$ is an index of wages, with $\overline{\omega}_n = \omega_n / \widehat{\omega}$ where w_n is the national average wage for staff of type n and $\widehat{\omega}$ is an arbitrary benchmark wage. We have chosen £76,000 as the benchmark, this corresponding to the average earnings of hospital-based doctors.

Information on the volume of inputs is not always available nor is it comprehensive. No data are collected about the physical amount of intermediate or capital resources used by NHS organisations. Even the Workforce Census data used to measure the volume and type of NHS staff underestimate their full contribution, as the measure of a Full Time Equivalent does not allow for overtime work (Street and Ward, 2009).

All hospital organisations do, however, report their expenditure comprehensively through their financial returns. These expenditure data can be used to calculate input use, by employing ‘indirect’ measurement. The indirect measure of total input costs in an SHA is defined as:

$$Z_s^{IND} = E_s^L + E_s^A + E_s^M + E_s^K$$

Eqn. 6

Where Z_s^{IND} is an aggregation of expenditure on NHS labour (E_s^L), agency staff (E_s^A), intermediate goods and services (E_s^M) and capital (E_s^K). Again taking NHS staff as the example E_s^L can be defined as follows:

$$E_s^L = \sum_{p=1}^{P} \sum_{n=1}^{N} E_{nps}^L$$

Eqn. 7

Where E_{nps}^L is the total expenditure in the NHS staff category *n* as reported in the financial returns of organisation *p* in SHA*s*.

Each organisation's expenditure is the product of the volume and price of its inputs. But the prices of labour, buildings and land vary across SHAs for reasons that are beyond organisational control. We remove these exogenous price effects by applying the sub-indices of the Department of Health's Market Forces Factor (MFF) to expenditure on labour and capital inputs (see Appendix for further details). There are not considered to be exogenous geographical influences on the prices for intermediate inputs.

While direct measurement is the preferred method of measuring the resources used by SHAs (Atkinson, 2005), data are available only about the volume of NHS labour input. Hence, we construct a third measure of input use, labelled the 'mixed' index Z_s^{MX}. This index sums the direct measure of NHS labour, Z_s^L, and the indirect measures for agency staff, intermediate goods and services and capital. Thus, the mixed measure of SHA input becomes:

$$Z_s^{MX} = Z_s^L + E_s^A + E_s^M + E_s^K$$

Eqn. 8

Finally, as the overall intention is to compare the services provided to each SHA's resident population with the resources used to provide this care, we need to recognise that patients are not always treated in their region of residence. To ensure that the 'money follows the patient' we take account of the fact that resources in each SHA are used both to treat residents of the SHA and residents of other SHAs.

There is considerable patient mobility in relation to hospital care across regions. Figure 1 illustrates the amount of immigration (patients treated in an SHA residing outside that SHA) and emigration (patients resident in an SHA who move to other SHAs for treatment) relative to total activity in percentage terms. This means that hospitals in London, in particular, are devoting relatively more of their resources to the treatment of patients from outside their SHA. So that inputs and outputs relate to the resident population, we calculate a 'migration factor' that measures the number of patients coming to the SHA for hospital treatment net of those living in the SHA who are treated elsewhere as a proportion of the total number of SHA residents treated in hospital. Hence, the input measure is adjusted downwards (upwards) in those SHAs that are net importers (exporters) of hospital inpatients. Technical details of all input calculations appear in the Appendix.

[Figure 1 about here]

Thus, total input will be lower than the national average in SHAs that: employ fewer staff, whether NHS or agency; employ relatively fewer staff in higher pay bands; spend less on intermediate goods and services; and have lower levels of capital expenditure.

3. Data Sources

3.1 Outputs

The volume of output includes all health care services provided to NHS patients resident in each SHA[2] as captured by Hospital Episode Statistics (HES) and Reference Cost data for the financial year 2007/08. HES is the prime source of data for the provision of hospital (inpatient and day case) services to NHS patients, covering all medical and surgical specialties and including patients treated in the private sector but publicly funded, although there are some quality issues regarding private sector data (Healthcare Commission, 2007, Street et al., 2010). The dataset consists of over 15 million Finished Consultant Episodes (FCEs) each representing the time a patient spends under the care of a single consultant (Clarke and McKee, 1992). We construct continuous inpatient spells (CIPS) from these FCEs which track patients across consultants and/or hospitals as part of their period of care (Castelli et al., 2008, Lakhani et al., 2005). We then count the number of CIPS in each Healthcare Resource Group (HRG) for each SHA. The cost of each CIPS is calculated on the basis of the most expensive FCE within the CIPS, with costs for each HRG derived from the Reference Costs data. We then calculate the national average cost per patient in each HRG.

The Reference Costs capture data about activities conducted in mental health and community care settings, outpatient and accident and emergency departments, and diagnostic facilities. These activity data are reported in various ways, including attendances,

[2] With the exception of primary care services about which reliable data at regional level are unavailable.

contacts, bed days and number of tests. We refer to these diverse activities collectively as Non-Admitted activity.

3.2 Inputs

To assess the inputs used in producing health services in each region, we analyse the workforce census and financial data submitted by all NHS providers.

Data on the number of NHS staff employed are taken from NHS Workforce Census data, and these are used to calculate z_{nps}^{L}. These data show headcounts and full time equivalents (FTEs) of staff employed in the NHS on 30th September 2007. We use FTEs in our calculations of NHS labour input. There are 417 different types of staffing categories.

The NHS Workforce Census data do not include information on earnings. Thus, to calculate the index of wages $\overline{\omega}_n$ we use the iView database, which contains earnings data by occupation for both medical and non medical staff employed in the NHS. The data are disaggregated by occupation code, and report national average figures for each occupation.

The financial returns used to measure the sub-components of Z_s^{IND} detail expenditure on both NHS and non-NHS staff by broad categories of labour input (i.e. medical staff, healthcare assistants, maintenance and works staff, administrative and clerical staff, managers); on intermediate inputs such as drugs and gases, clinical supplies, catering, premises costs and purchases of health care from non-NHS bodies; and on capital equipment (current outlays on equipment and past expenditure reported as depreciation on assets[3]).

[3] Further details on how the use of capital resources is assigned to each time period are found in Street and Ward (2009).

4. Results

4.1 Outputs

Table I reports the volume of outputs by SHA. Three sets of figures are presented for (1) Hospital patients, (2) Non-Admitted patients and (3) All patients. In each case 'unadjusted activity' shows a simple count of volumes of activity. These counts show that the North West has the highest volume of activity in both hospital and non-admitted settings while the North East has the lowest volume of activity in both settings.

The 'cost-weighted' columns show the impact of cost-weighting the output by the cost index $\bar{c}_j$. In using as our benchmark the average cost of hospital inpatient treatment (amounting to £1,167), cost-weighted hospital activity is unchanged nationally. Nevertheless, there are changes across SHAs reflecting variation in case-mix. For example, South East Coast output increases by 6% whereas Yorkshire and the Humber output decreases by 4%, implying that hospital case-mix is relatively more complex in the former than the latter SHA.

In calculating the cost-weighted activity for non-admitted patients the choice of the benchmark $(\hat{c}_j)$ has a marked impact. In effect, the cost-weighted figures are scaled downward by an average of 96% because such activity is relatively much cheaper than the benchmark. Again, different case-mix activity leads to differential impacts across SHAs: London's non-admitted output is rescaled to 4.5% of its unweighted activity whereas activity in South Central is rescaled to 3%.

[Table I about here]

In terms of differences in quality, Figure 2 shows SHAs' deviations from the national average in the main quality measures used in the calculation of the output index: 30-day survival rates, separately for elective and non-elective hospital activity, 80th percentile waiting times for elective hospital activity and average waiting times for outpatient visits.

The top left quadrant of Figure 2 shows 30-day survival rates from the national average for patients admitted on an elective basis, with survival rates being higher in the North West and London. Thus, all else equal, quality-adjusted output in these SHAs will appear higher than cost-weighted output. The deviation in survival rates for non-elective patients is shown in the top right quadrant with rates in London and South Central better than elsewhere[4]. Allowing for the survival effects, therefore, will raise the amount of non-elective output above cost-weighted counts of output for these two SHAs.

The remaining two quadrants show 80th percentile waiting times for hospital elective activity (bottom left) and for outpatient visits (bottom right). Hospital waiting times are higher than the national average in the South East Coast SHA so, all else equal, the output of this SHA will appear lower if waiting times are accounted for than if output were merely a count of activity. South East Coast, London and West Midlands have lower average waiting times for outpatient visits than the national average, so outpatient output in these SHAs will be scaled up when outpatient waiting times are taken into consideration.

[Figure 2 about here]

[4] The scale of 30-day survival rate for elective hospital activity differs from that for non-elective hospital activity.

The quality-adjusted columns in Table I show the effect of accounting for quality, the impact of which is most clearly seen in the hospital sector: at the extremes, London's cost-weighted output increases by 5% when quality is included in the output measure whereas South East Coast's is reduced by 4%. For non-admitted patients, allowing for quality has a minor impact as it applies only to waiting times for outpatient attendances and mental healthcare services. Regional differences in quality raise cost-weighted output for non-admitted patients by 0.04% in London and reduce it by 0.03% in East of England.

Across SHAs the combined relative impact of cost-weighting and quality-adjusting output is greater in the hospital sector: Yorkshire and the Humber SHA's output is adjusted downwards by 3.5% due to a less resource intensive output, whereas South Central and South East Coast output is increased by 2.25%. In the case of South Central this is entirely due to case-mix whereas for South East Coast a cost-weighted increase of 6% is reduced by a lower than average quality adjustment. Overall rescaled output is still highest in the North West while non-hospital output remains highest in London.

4.2 Inputs

Table II reports a breakdown of expenditure by all NHS organisations in each SHA by input type. Two sets of figures are presented for NHS staff, one derived from the Workforce Census, the other from financial returns.

On average, labour (NHS staff and agency) accounts for around 60% of total expenditure with 27% devoted to intermediate inputs and 13% to capital. These proportions vary across SHAs. Expenditure on labour as a proportion of total spend is highest in the North East at 62% and lowest in South Central at 57%. In terms of intermediate inputs, expenditure is highest in London at 34% and lowest in East Midlands, North West and Yorkshire & the

Humber each at 26%. Proportionate spending on capital ranges from 15% in both the North West and West Midlands to a low of 8% in London.

The only difference between our direct and mixed method is in how our NHS staff estimates are calculated. Looking at Table II we can see that there is variation of up to 6.9% in our staffing estimates depending on the data source used, with the discrepancy being greatest in the South West.

[Table II about here]

4.3 Productivity

We calculate two measures of standardised productivity as defined in Equation 1, in order to assess the sensitivity of the estimates to assumptions about the construction of the measure of input. These are presented in Figure 3. Both measures use Equation 2 to calculate the output index. The 'mixed' productivity measure uses Equation 8 to calculate the input index. The resulting measures of standardised productivity show that productivity is highest in South West SHA, at 5.3% above the national average and lowest in the East Midlands where it is 6.6% below the national average.

The 'indirect' productivity measure substitutes the input index calculated using Equation 8 with that calculated using Equation 6. This variant of the productivity measure has a favourable impact on the estimates for London, South Central and South West SHAs, implying that – after accounting for MFF – these are paying relatively less than the national average per member of staff. The opposite is the case for North East, West Midlands, and Yorkshire & the Humber SHAs. For the other SHAs, productivity estimates are not particularly sensitive to the choice of how to measure inputs. The sensitivity of productivity estimates to the choice of input index is probably due to the fact that organisations receiving above average MFF allocations are constrained by national wage bargaining in the wages they offer. In effect, therefore, these organisations are using the additional monies received through MFF not merely to pay higher wages but also to recruit more staff.

[Figure 3 about here]

By placing SHAs into groups with comparable population sizes and comparing outputs and inputs per capita, Figure 4 gives additional insight into what are driving the differences in productivity ratios. The Figure places the output per head of population and input per head of population for each SHA side by side on axes chosen such that if the SHA has a better than average productivity then the output column will exceed that of the input column. Conversely, if the output column is lower than that of the input column then this means that the SHA will have a lower than average productivity ratio. Expressed in this way, the gap between the columns is indicative of the extent of the variation in productivity. The South West and East Midlands - with the biggest gaps between outputs and inputs - are clearly identified as the two SHAS with the best and worst productivity ratios.

If we consider the group of SHAs with populations of around four million, consisting of South Central, South East Coast and East Midlands, we find that all three have input columns higher than output columns, which explains their lower than national average productivity. In addition, the reason why East Midlands has the worst ratio becomes clearer: although it has a similar output to its comparator SHAs, it uses relatively more inputs.

Similarly, if we consider the group of SHAs with populations of around five million consisting of South West, Yorkshire & the Humber, West Midlands and East of England, we can see that the SHA with best productivity ratio, South West, has the same output as Yorkshire & the Humber and the West Midlands but uses relatively lower inputs. For its population size, East of England is an interesting outlier: it has low output but uses a similarly relatively low amount of input to achieve that output; indeed the columns would suggest a population size of four million rather than five million. Nevertheless, because its relatively low output is more than offset by it relatively lower input use, East of England has above average productivity.

For the North East SHA, although serving the smallest population, output per head is relatively high with a correspondingly high use of inputs. As indicated in Figure 3 whether the SHA is viewed as having above or below average productivity depends on the choice of input measure. For the largest SHAs of the North West and London both have a similar positive difference between outputs and inputs and hence have similar productivity ratios.

[Figure 4 about here]

5. Conclusions

By linking together large-scale routinely collected datasets we produce and compare productivity estimates across the ten Strategic Health Authorities in England in 2007/08. We analyse data from Hospital Episode Statistics, the Reference Costs, Financial Returns, and workforce census. Data about patients seen in primary care are not available. Other than primary care, the data cover all patients treated by all organisations in each SHA. To our knowledge such an analysis of productivity variation using such detailed patient and organisational data has not been carried out anywhere before now.

We have measured productivity for each SHA by comparing the total amount of health care 'output' provided for the SHA's resident population to the total amount of 'input' used to produce this output. Output consists of all health care services provided to NHS patients in the hospital and community care sectors. The output measure also takes account of quality by measuring regional differences in hospital survival rates, and inpatient and outpatient waiting times. Inputs include the staff, intermediate goods and services, and capital resources that contribute to the production of health care. Inputs are adjusted for the market forces factor and we account for movement of patients between SHAs, so that 'money follows patients'.

Baseline productivity ratios across SHAs vary from 5% above to 6% below the national average. Productivity is highest in South West SHA and lowest in East Midlands, South Central and Yorkshire & the Humber SHAs. In general, relative positions of SHAs hold irrespective of whether the indirect or mixed approach is used to measure inputs, although the actual ratios are sensitive to how the input index is constructed.

Our measure of productivity explicitly incorporates different types (case-mix) as well as volume of patients treated, a quality adjustment based on waiting times and survival rates and adjustment for differences in input prices across regions. This means that the observed geographical variation is not due to such factors. Thus, for instance, an SHA with a ratio below average may not claim that this is a result of doing fewer but more complex activities.

The question then is what may account for the relative differences. These may be due partly to differences in labour productivity. The South West SHA may benefit from a more stable workforce, vacancy rates for non-medical staff being well below the national average (The Information Centre, 2007). Lower productivity in the hospital and community sectors may be because more work is undertaken in primary care. The absence of regional data about the activities undertaken in general practice makes it difficult to establish what General Practitioners (GPs) are doing in different parts of the country.

Let us suppose that all SHA regions could become as productive as the South West or, more accurately, as productive as the average Primary Care Trust in the South West. If this benchmark were met across the country, the NHS could treat the same number of patients with £3.2bn fewer resources each year. This is still a long way short of the £5 billion in annual savings that the Coalition government is seeking to secure from the NHS (Ham, 2010), implying that savings may have to come at the expense of a reduction in NHS output or by reductions in input prices.

This initial look into regional variation in productivity is indicative of a research area that is likely to produce a new, fruitful and policy-relevant perspective on the structure and performance of the NHS. The observed variation in productivity ratios across SHAs clearly shows that gains could be made if underperforming SHAs would operate at the same level as the South West SHA. The next steps would be to identify the reasons why organisations in the South West are more productive than elsewhere and to share best practice.

Appendix

Adjustment for differences in expenditure due to factors outside organisational control

Expenditure on staff and capital is adjusted for geographical differences in factor prices by applying the labour Market Forces Factor (MFF) and an amalgam of the land and buildings MFF for each organisation. Concerns about the 2007/08 MFF led to a revised formulation being used to calculate the 2008/9 MFF and this is what we have used. Data comes from the PCT recurrent revenue allocations exposition book (Department of Health, 2010).

Denote the staffing MFF in organisation p, to be applied to labour input as θ_p^L. We apply a weighted average of the buildings MFF θ_p^{Bld} and land MFF θ_p^{Lnd} indices to capital inputs, such that $\theta_p^K = w_{1p}\theta_p^{Bld} + w_{2p}\theta_p^{Lnd}$ and $w_{1p} + w_{2p} = 1$. The MFF adjusted measure of SHA expenditure, then, is calculated as:

$$Z_s^{IND\_MFF} = \sum_{p=1}^{P} \{\theta_p^L E_{ps}^L + \theta_p^L E_{ps}^A + \theta_p^K E_{ps}^K + E_{ps}^M\}$$

Where E^L is expenditure on NHS labour, E^A is expenditure on agency staff, E^K is expenditure on capital, and E^M is expenditure on intermediate inputs.

Adjustment for migration of patients to hospitals outside their SHA of residence

Outputs are measured for the resident population of each SHA. But input data relate to the organisations located with the SHA boundaries, which treat both residents and non-residents of the SHA in which they are located. To match resources to outputs, we calculate a migration factor σ that captures the movement of patients across SHAs: if there are more patients coming to the SHA than leaving then $\sigma < 1$. This applies only to hospital expenditure, given that (i) the migration factor is based only on those moving for hospital care and (ii) patients are less likely to move for other health services. Applying this adjustment, we have:

$$Z_s^{IND\_MFF\_MIG} = \sigma_s \sum_{p=1}^{f} E_{ps}^{MFF} + \sum_{p=f+1}^{P} E_{ps}^{MFF}$$

Where hospitals are referenced *p*=1…*f* and all other organisations are referenced *p*=*f*+1…*P.*

Mixed approach to measuring inputs

The mixed 'direct and indirect' index, is specified as follows:

$$Z_s^{MX} = \sigma_s \sum_{p=1}^{f} \left[\pi_1 Z_{ps}^L + \left\{\theta_p^L E_{ps}^A + \theta_p^K E_{ps}^K + E_{ps}^M\right\}\right] + \sum_{p=f+1}^{P} \left[\pi_2 Z_{ps}^L + \left\{\theta_p^L E_{ps}^A + \theta_p^K E_{ps}^K + E_{ps}^M\right\}\right]$$

Where π_1 is the a monetary value used to convert FTEs of hospital staff onto an equivalent scale to the expenditure data. The monetary value is calculated as:

$$\pi_1 = \sum_{p=1}^{f} \theta_p^L E_p^L \Big/ \sum_{p=1}^{f} Z_p^L$$

for *p*=1…*f* where $\sum_{p=1}^{f} \theta_p^L E_p^L$ is national MFF-adjusted expenditure on NHS staff working in hospitals and $\sum_{p=1}^{f} Z_p^L$ is national NHS staffing input in hospitals as calculated using the direct method.

And π_2 is the monetary value applying to PCT staff, calculated as:

$$\pi_2 = \sum_{p=f+1}^{P} \theta_p^L E_p^L \Big/ \sum_{p=f+1}^{P} Z_p^L$$

where the numerator is national MFF-adjusted expenditure on NHS staff working in PCTs and the denominator is national PCT staffing input.

Acknowledgements

Thanks are due to Keith Derbyshire, Martin Land, Tom Rogers, Paul Sherrell and Rob Unsworth. The project was funded by the Department of Health in England as part of a programme of policy research at the Centre for Health Economics, University of York. The views expressed are those of the authors and may not reflect those of the funder.

References

ATKINSON, T. 2005. *Atkinson Review: Final Report. Measurement of Government Output and Productivity for the National Accounts,* Basingstoke, Palgrave Macmillan.

ATKINSON, T. 2006. Measurement of Government output and productivity. *Journal of the Royal Stastistical Society, series A,* 169**,** 659-662.

CARTER, N., KLEIN, R. & DAY, P. 1992. *How Organisations Measure Success,* London, Routledge.

CASTELLI, A., DAWSON, D., GRAVELLE, H., JACOBS, R., KIND, P., LOVERIDGE, P., MARTIN, S., O'MAHONY, M., STEVENS, P., STOKES, L., STREET, A. & WEALE, M. 2007a. A new approach to measuring health system output and productivity. *National Institute Economic Review,* 200**,** 105-117.

CASTELLI, A., DAWSON, D., GRAVELLE, H. & STREET, A. 2007b. Improving the measurement of health system output growth. *Health Economics,* 16**,** 1091-1107.

CASTELLI, A., LAUDICELLA, M. & STREET, A. 2008. *Measuring NHS output growth,* York, Centre for Health Economics research paper 43.

CLARKE, A. & MCKEE, M. 1992. The Consultant Episode: an unhelpful measure. *British Medical Journal,* 305**,** 1.

DAWSON, D., GRAVELLE, H., KIND, P., O'MAHONY, M., STREET, A. & WEALE, M. 2004. Developing new approaches to measuring NHS outputs and productivity. *In:* CASTELLI, A., JACOBS, R., MACRAN, S., QUINN, C., STEVENS, P. & STOKES, L. (eds.). York: CHE Technical Paper Series No. 31.

DAWSON, D., GRAVELLE, H., O'MAHONY, M., STREET, A., WEALE, M., CASTELLI, A., JACOBS, R., KIND, P., LOVERIDGE, P., MARTIN, S., STEVENS, P. & STOKES, L. 2005. *Developing new approaches to measuring NHS outputs and productivity, Final Report,* York, Centre for Health Economics research paper 6.

DEPARTMENT OF HEALTH 2007. Combining Age Related & Additional Needs (CARAN) Report. London: Department of Health.

DEPARTMENT OF HEALTH 2010. PCT recurrent revenue allocations exposition book. *In:* DIRECTORATE, R. A. T.-F. A. O. (ed.).

DEPARTMENT OF HEALTH RESOURCE ALLOCATION WORKING PARTY 1976. *Sharing Resources for Health in England,* London, HMSO.

EUROSTAT 2001. *Handbook on price and volume measures in national accounts,* Luxembourg, Office of Official Publications of the European Communities.
FISHER, E., BELL, J. E., TOMEK, I., ESTY, A., GOODMAN, D. & BRONNER, K. 2010. Trends and Regional Variation in Hip Knee and Shoulder Replacement. A Dartmouth Atlas Surgery Report.
FISHER, E., BYNUM, J. & SKINNER, J. 2009. The policy implications of variations in Medicare spending growth. The Dartmouth Institute for Health Policy and Clinical Practice.
HAM, C. 2010. The coalition government's plans for the NHS in England. *British Medical Journal,* 341:3790**,** 111-112.
HEALTHCARE COMMISSION 2007. *Independent sector treatment centres: a review of the quality of care,* London, Commission for Healthcare Audit and Inspection.
KRISTENSEN, T., LAUDICELLA, M., EJERSTED, C. & STREET, A. 2010. Cost variation in diabetes care delivered in English hospitals *Diabetic Medicine,* 27**,** 949-57.
LAKHANI, A., COLES, J., EAYRES, D., SPENCE, C. & RACHET, B. 2005. Creative use of existing clinical and health outcomes data to assess NHS performance in England: Part 1-performance indicators closely linked to clinical care. *British Medical Journal,* 330**,** 1426-1431.
MADIANOS, M. G., ZACHARAKIS, C., TSITSA, C. & STEFANIS, C. 1999. The Mental Health Care Delivery system in Greece: Regional variation and socioeconomic correlates. *The Journal of Mental Health Policy and Economics,* 2**,** 169-76.
NHS EXECUTIVE 1992. *Health Service Indicators handbook 1990/91,* Leeds, NHS Executive.
PATEL, N., ADATIA, R., MELLEMGAARD, A., JACK, R. & MOLLER, H. 2007. Variation in the use of chemotherapy in lung cancer. *British Journal of Cancer,* 96**,** 886-90.
SCHLEINIGER, R. 2008. Regional Quantity, Productivity & Efficiency Measures of the Swiss Health Care System. *Swiss Journal of Economics & Statistics,* 144**,** 18.
SKINNER, J. & FISHER, E. 2010. Reflections on geographic variations in U.S. Health Care. The Dartmouth Institute for Health Policy & Clinical Practice.
STREET, A., SIVEY, P., MASON, A., MIRALDO, M. & SICILIANI, L. 2010. Are English treament centres treating less complex patients? *Health Policy,* 94**,** 150-157.
STREET, A. & WARD, P. 2009. *NHS Input and Productivity Growth 2003/4 - 2007/8,* York, Centre for Health Economics research paper 47.
THE INFORMATION CENTRE, D. O. H. 2007. NHS Workforce vacancies survey 2007. *In:* HEALTH, D. O. (ed.). Leeds.
THOMSON, R. G., MCELROY, H. & KAZANDJIAN, V. A. 1997. Maryland Hospital Quality Indicator Project in the United Kingdom: an approach for promoting continuous quality improvement. *Quality in Health Care,* 6**,** 49-55.
YU, K. & ARISTE, R. 2010. Comparisons of Hospital Output in Canada: National and International Perspectives.

Table I: NHS activity sources and summary statistics by SHA

Data Source	NHS activity	Strategic Health Authority									
		East Midlands	East of England	London	North East	North West	South Central	South East Coast	South West	West Midlands	Yorkshire & The Humber
HES	**Hospital output**										
	Elective and day cases										
	Volume of activity	591,163	702,976	904,398	392,370	1,031,592	472,474	501,778	774,067	688,024	716,513
	Mean 30-day post discharge survival rate	0.996	0.996	0.997	0.996	0.997	0.996	0.996	0.996	0.996	0.996
	Mean life expectancy	23.1	23.0	26.1	23.7	23.7	23.8	22.5	22.8	23.7	24.3
	80th percentile waiting times	80	91	83	68	76	85	105	79	82	78
	Non-electives										
	Volume of activity	523,159	571,338	871,491	353,883	975,740	442,051	452,657	578,285	649,173	646,434
	Mean 30-day post discharge survival rate	0.954	0.949	0.966	0.953	0.957	0.959	0.949	0.951	0.955	0.955
	Mean life expectancy	32.6	30.7	33.8	31.1	32.8	34.3	30.4	30.4	33.0	32.0
	Mental Health inpatient										
	Elective and day cases										
	Volume of activity	3,310	3,672	5,602	1,113	3,982	2,415	2,786	5,296	4,275	3,738
	Non-electives										
	Volume of activity	8,795	8,002	18,215	4,137	22,430	6,281	7,462	7,329	11,367	10,215
Referenc Costs	**Mental Health non-admitted care**										
	Volume of activity	1,133,994	1,669,854	3,924,379	1,108,196	4,065,130	1,291,482	1,532,909	2,269,121	2,228,098	1,838,739
	Outpatient										
	Volume of activity	4,998,738	6,332,670	11,444,988	3,975,337	11,235,076	4,213,883	5,077,311	7,067,161	7,141,006	8,208,435
	Mean waiting time (weeks)*	3.43	3.35	3.67	3.25	3.19	3.10	3.81	3.30	3.47	3.18
	Community care										
	Volume of activity	6,676,370	8,202,332	12,194,902	6,551,584	14,895,042	5,090,955	5,967,712	8,113,322	8,631,712	9,171,091
	Other NHS activity**										
	Volume of activity	19,348,979	29,597,743	42,883,706	18,150,095	46,974,180	28,094,575	25,376,235	35,968,183	36,400,119	35,652,957

Table II: Real expenditure on NHS inputs (£million) by SHA

Data Source	NHS Inputs	Strategic Health Authority									
		East Midlands	East of England	London	North East	North West	South Central	South East	South West	West Midlands	Yorkshire & The Humber
Financial Returns	**Labour**										
(NHS Hospital,	NHS Staff	3,111,378	3,383,384	5,557,184	2,214,979	5,726,862	2,378,939	2,627,801	3,569,918	3,965,573	4,011,992
Foundation,	Agency Staff	69,998	9,388	284,398	66,042	156,534	110,207	93,775	81,573	107,184	79,038
Mental Health and	Deflator*	NHS Pay Index									
Community Trusts,											
Ambulance trusts,	**Intermediate Inputs**										
PCTs, SHAs, DH)	Intermediate Inputs	1,404,014	1,792,781	3,389,227	1,028,235	2,624,653	1,358,278	1,458,517	1,672,366	1,959,480	1,775,377
	Deflator*	Pay and Prices deflator, FHS deflator									
	Capital Inputs										
	Capital Inputs	713,766	661,866	825,833	410,084	1,490,504	553,381	551,240	758,075	1,039,539	891,356
	Deflator**	Specific capital assets deflators (MM17 Price Index)									
	Total Trusts and PCTs	**5,299,156**	**5,847,419**	**10,056,642**	**3,719,340**	**9,998,553**	**4,400,805**	**4,731,333**	**6,081,932**	**7,071,776**	**6,757,763**

Table III: Observed and Quality-Adjusted Hospital and Community Care Outputs by SHA

	Hospital Patients			Non-Admitted Patients			All Patients	
SHA	Unadjusted Activity	Cost weighted activity	Quality-adjusted Activity	Unadjusted Activity	Cost weighted activity	Quality-adjusted Activity	Cost weighted activity	Quality-adjusted Activity
East Midlands	1,114,322	1,113,016	1,109,450	32,170,186	1,295,199	1,294,924	2,408,215	2,404,374
East of England	1,274,314	1,319,863	1,281,547	45,814,273	1,640,059	1,639,640	2,959,922	2,921,187
London	1,775,889	1,708,157	1,793,668	70,471,792	3,198,900	3,200,102	4,907,057	4,993,770
North East	746,253	744,365	756,550	29,869,462	1,017,389	1,017,301	1,761,754	1,773,851
North West	2,007,332	1,947,529	1,981,711	77,195,840	2,891,156	2,891,726	4,838,685	4,873,437
South Central	914,525	933,137	935,473	38,699,591	1,146,660	1,146,411	2,079,797	2,081,883
South East Coast	954,435	1,015,335	975,841	37,964,415	1,296,722	1,296,464	2,312,057	2,272,304
South West	1,352,352	1,405,641	1,357,267	53,430,412	1,868,107	1,867,984	3,273,748	3,225,251
West Midlands	1,337,197	1,343,961	1,345,970	54,416,577	1,949,548	1,949,656	3,293,509	3,295,626
Yorkshire & The Humber	1,362,947	1,308,560	1,314,987	54,885,175	1,834,078	1,834,236	3,142,638	3,149,223
Total	*12,839,566*	*12,839,563*	*12,852,464*	*494,917,723*	*18,137,820*	*18,138,443*	*30,977,383*	*30,990,907*

Table IV: Total NHS Expenditure by Input Type (£000)

SHA	NHS Staff		Agency Staff	Intermediate Inputs	Capital	Total inputs	
	E_S^L	Z_S^L	E_S^A	E_S^M	E_S^K	Eq.6	Eq.8
East Midlands	£3,111,378	£3,130,719	£69,998	£1,404,014	£713,766	£5,299,155	£5,318,497
East of England	£3,383,384	£3,367,033	£93,880	£1,792,781	£661,866	£5,931,911	£5,915,560
London	£5,557,184	£5,711,543	£284,398	£3,389,227	£825,833	£10,056,642	£10,211,002
North East	£2,214,979	£2,106,062	£66,042	£1,028,235	£410,084	£3,719,339	£3,610,423
North West	£5,726,862	£5,660,641	£156,534	£2,624,653	£1,490,504	£9,998,553	£9,932,331
South Central	£2,378,939	£2,429,525	£110,207	£1,358,278	£553,381	£4,400,805	£4,451,390
South East	£2,627,801	£2,622,018	£93,775	£1,458,517	£551,240	£4,731,333	£4,725,550
South West	£3,569,918	£3,816,704	£81,573	£1,672,366	£758,075	£6,081,932	£6,328,719
West Midlands	£3,965,573	£3,836,873	£107,184	£1,959,480	£1,039,539	£7,071,777	£6,943,076
Yorkshire & The Humber	£4,011,992	£3,866,890	£79,038	£1,775,377	£891,356	£6,757,763	£6,612,662
Total	£36,548,010	£36,548,010	£1,142,629	£18,462,927	£7,895,644	*£64,049,210*	£64,049,210

Figure 1: Immigration and Emigration as a Percentage of Total Activity

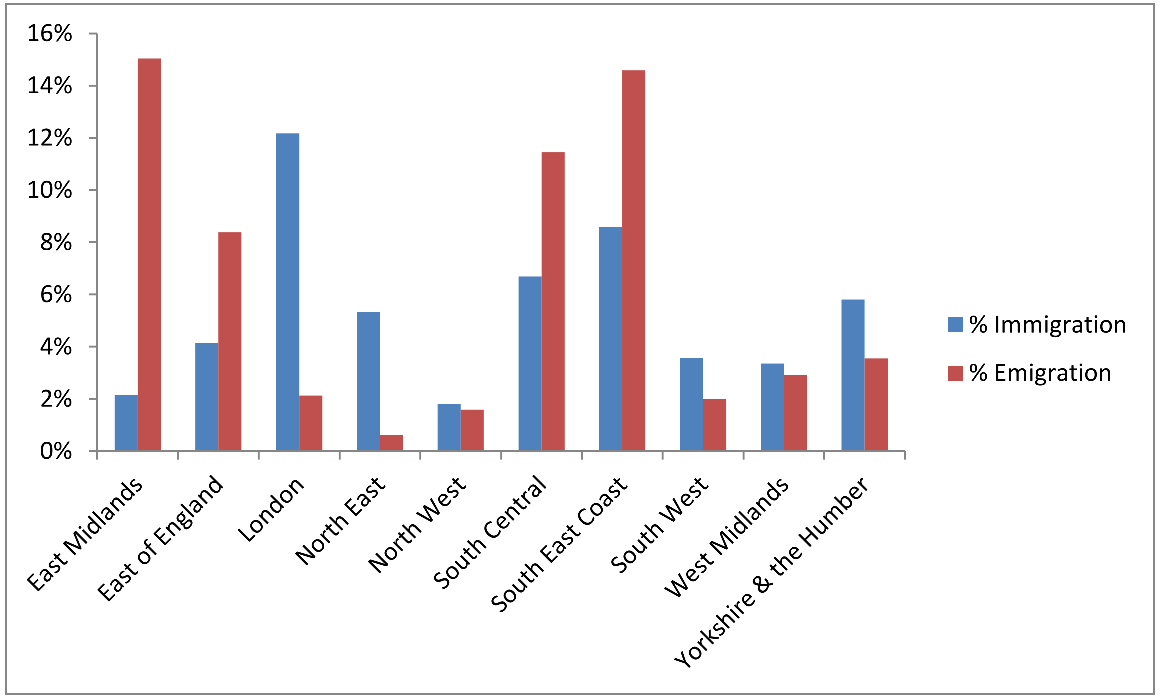

Figure 2: Quality Adjustors for Hospital Activity and Non-admitted Activity by SHA

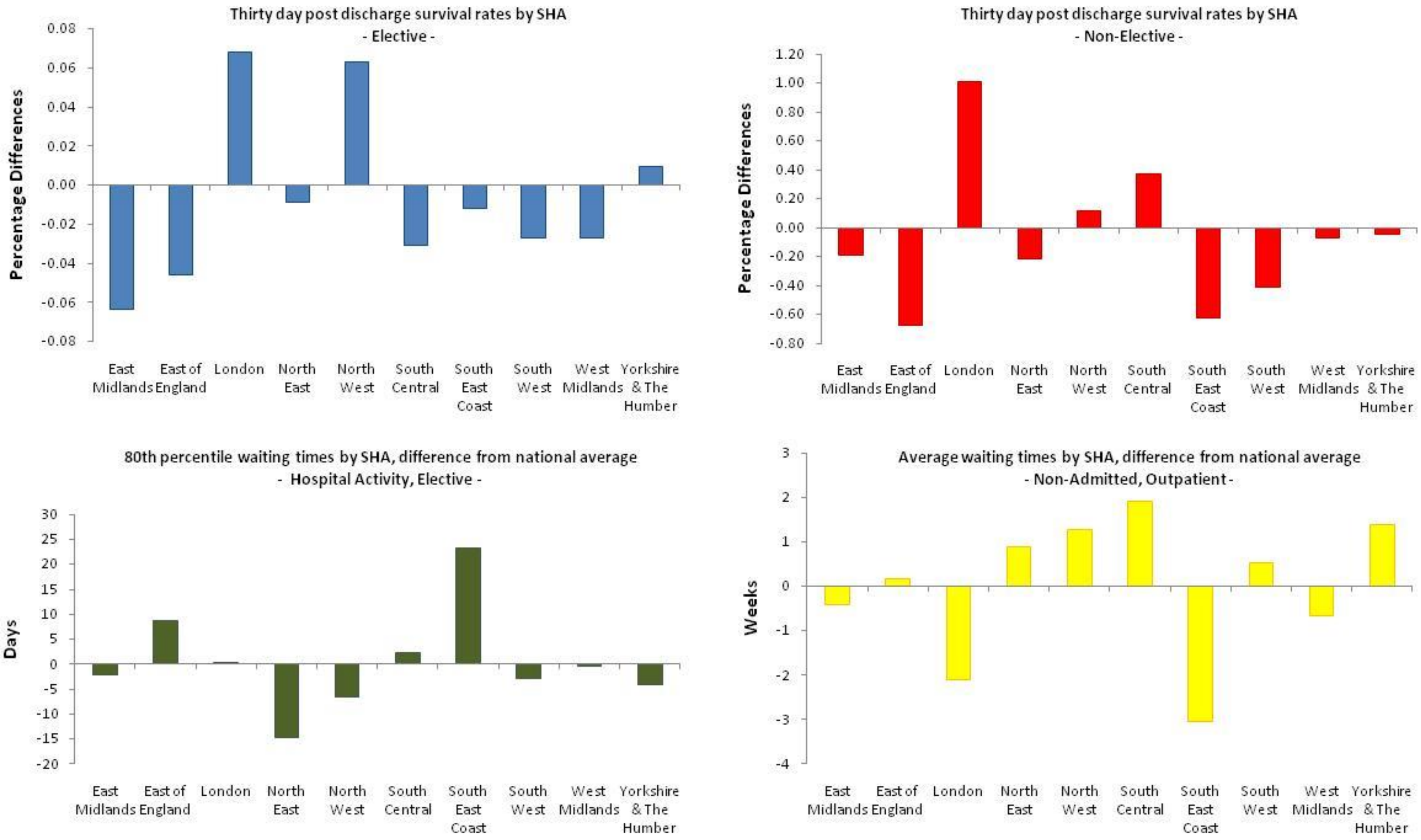

Figure 3: Standardised Productivity by SHA

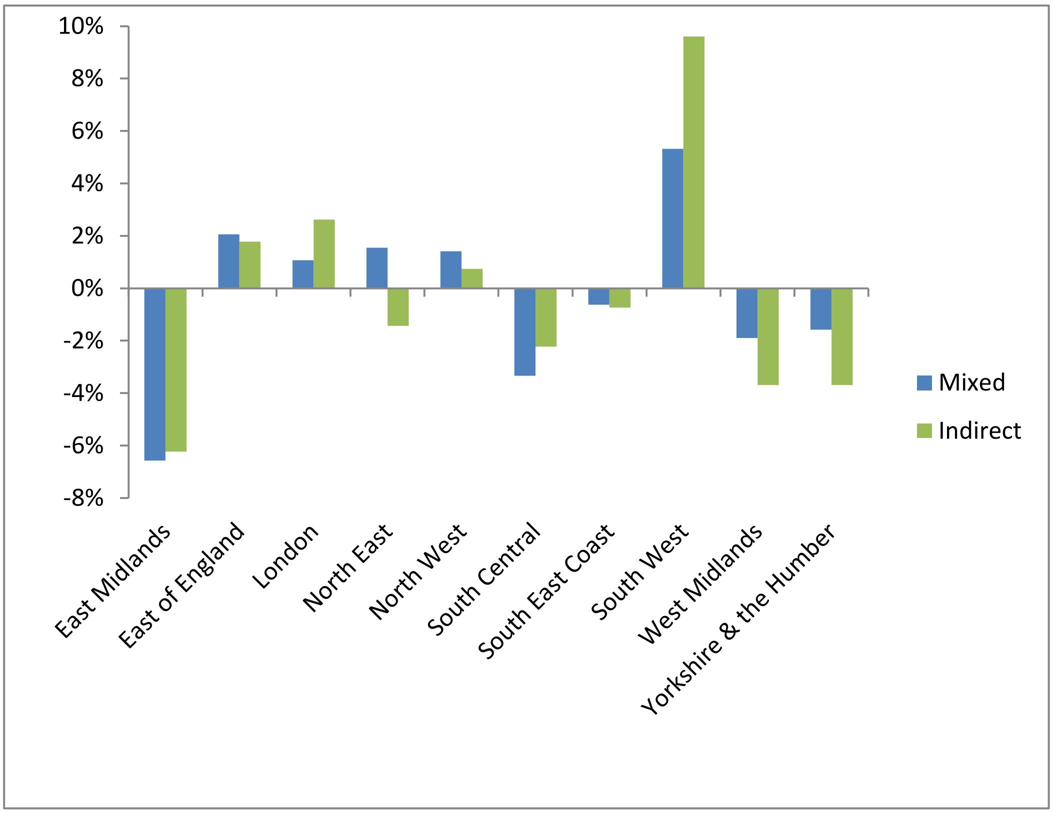

Figure 4: Outputs and Inputs by Comparable Population SHAs