ISSN 2042-2695

CENTRE *for* ECONOMIC PERFORMANCE

CEP Discussion Paper No 1484

May 2017

The Hospital as a Multi-Product Firm: The Effect of Hospital Competition on Value-Added Indicators of Clinical Quality

Matthew Skellern

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Abstract

There is increasing international interest in using Patient Reported Outcome Measures (PROMs) to assess health care provider performance. PROMs are a fundamental advance on existing indicators of health care quality in two respects: they equate outcomes with value added (i.e. health gain) from treatment rather than post-treatment health status, and they allow clinical quality to be measured at the level of the individual medical intervention to a far greater extent than existing failure-based indicators of quality such as mortality or readmissions. Most existing econometric studies of hospital competition and quality equate outcomes with post-treatment health status, and use mortality rates of various kinds as indicators of overall hospital performance, in spite of the fact that mortality is a relatively uncommon outcome in the spheres of hospital activity - such as elective surgery - in which competition for patients does occur. This paper contributes to the development of a value-added, multi-product conception of hospital quality by studying the impact of a major competition-promoting reform to the English NHS in 2006, in which patients were allowed to choose which hospital they attended for elective surgery, on PROMs of health gain from hip and knee replacement, groin hernia repair, and varicose vein surgery. In contrast to the existing literature, I find that the competition brought about by the introduction of patient choice of hospital may have had a negative effect on clinical quality. I put forward a theoretical framework that explains these findings, and conclude by arguing that future research should model the hospital as a multi-product firm, and capture clinical quality using value-added outcome measures.

Keywords: health care, hospital competition, price regulation, prospective reimbursement, patient choice, health care quality, vertical product quality, performance measurement, multi-tasking, value added
JEL codes: C21; D21; H42; I11; I18; L1; L15; L32; R12

This paper was produced as part of the Centre's Growth Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

Acknowledgements

I acknowledge invaluable support, guidance and oversight from Zack Cooper, as well as the invaluable input of Gwyn Bevan during the planning stages of this project. In addition, I am very grateful to the following people for their wisdom, suggestions and/or assistance: Tim Besley, Michael Best, Alessandro Gavazza, Maitreesh Ghatak, Stephen Gibbons, Andrew Hutchings, Simon Jones, Henrik Kleven, Sara Machado, Alistair McGuire, Tom O'Keeffe, Ted Pinchbeck, Steve Pischke, Carol Propper, Rosa Sanchis, Sarah Sandford, Jon Skinner, Johannes Spinnewijn, Mohammad Vesal, and Nic Warner. Site code cleaning was performed using a modified version of code kindly provided by Stuart Craig and Zack Cooper. Coding of predicted patient flow competition measures was informed by code published by Martin Gaynor, Carol Propper and Rodrigo Moreno-Serra (Gaynor et al. 2013). All of the deficiencies in the resulting work are the author's responsibility alone.

Matthew Skellern is an Associate of the Centre for Economic Performance's Growth Programme. He is also a Fellow in Health Economics at the Department of Social Policy, London School of Economics.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

There is increasing international interest in using Patient Reported Outcome Measures (PROMs) from health interventions to assess the performance of health care providers. The English NHS distributes PROMs to all patients undergoing selected elective surgical procedures, while the US Centers for Medicare & Medicaid Services has declared its intention to base reimbursement decisions on PROMs as part of its commitment to link 50 per cent of Medicare payments for primary care to quality or value by 2018 (CMS 2016).

PROMs are value-added measures of provider quality because they survey patients on health status both before and after an intervention, and equate outcomes with the gain in health status from the intervention. In this respect, PROMs are a fundamental advance on existing indicators of health care quality, which equate health outcomes with post-treatment health status. The fact that health outcomes are influenced by patient characteristics, which are often poorly observed, means that final outcomes (post-treatment health status) are a problematic indicator of provider quality. A similar situation exists in education, where outcomes (e.g. exam results) are a joint product of student characteristics and school quality. While the economics of education literature has, in recognition of this situation, increasingly focused over the last decade on value-added measures of school quality, the health economics literature has largely continued to equate hospital outcomes with final patient health status, because value-added measures of health outcomes such as PROMs are much less widely available than in education.

Past experience in education suggests that shifting the focus of performance measurement from final outcomes to value added can upend provider rankings and overturn conventional wisdom about the drivers of provider quality (see e.g. Adams & Bengtsson 2017). With this in mind, the present paper returns to a widely studied question within health economics – the impact of fixed-price hospital competition on clinical quality – to examine whether results from the existing literature can be reproduced when value-added quality measures are used. Since April 2009, PROMs surveys have been distributed to all English NHS patients undergoing one of four elective surgical procedures[1] – hip and knee replacement, groin hernia repair and varicose vein stripping. I merge these PROMs with the Hospital Episode Statistics (HES), which includes an observation for every NHS-funded inpatient hospital stay, to study the impact of a major competition-promoting reform in 2006 that enabled patients to choose which hospital they attend for elective surgery. Hospitals received a fixed 'tariff' (or price) for each patient treated – hence patient choice meant that hospitals were forced to compete for patient referrals, instead of being guaranteed a given patient load via bulk contracts with care purchasers as had previously been the case.

This paper studies the impact of the competition engendered by these patient choice reforms on clinical quality as captured by PROMs health gains from elective surgery, using instrumentation strategies to address endogeneity arising from patient choices, and exploiting the value-added nature of PROMs to control for aspects of patient health status that are correlated with competition intensity. While previous studies have generally found that hospital competition after 2006 led to higher hospital quality, this paper finds that, when value-added measures of quality are used, hospital competition may have led to *lower* care quality. While I argue that these divergent results are likely driven by differential responses to competition across hospital outputs, rather than by the use of a value-added indicator of quality *per se*, it is only because of the existence of new, value-added indicators of hospital performance that it has been possible to detect these differential effects.

As well as examining the impact of hospital competition on value-added measures of clinical quality for the first time, this paper contributes to a second new line of investigation concerning the impact of market structure on hospital performance, by modelling the hospital as a multi-product firm. Standard economic models of fixed-price hospital competition (Gaynor 2006; Gaynor & Town 2012) assume that hospitals produce a single type of output, and choose a single, hospital-wide quality level. They predict that increased competition will lead to higher quality so long as the regulated price exceeds the marginal cost with respect to quantity – the intuition being that, if prices are fixed, hospitals only have one choice variable, quality, and will therefore compete for market share on this basis.

The econometric literature on market structure and hospital outcomes largely works within this single-output-type theoretical framework, by focusing on indicators of hospital performance – such as

---

[1] Elective surgery encompasses any surgical procedure that is not urgent or an emergency, and which can therefore be scheduled in advance.

mortality rates – that not only equate outcomes with final patient health status, but also implicitly or explicitly assume that clinical quality is a hospital-wide variable, or has a significant hospital-wide component.[2] Three previous studies (Cooper et al. 2011; Gaynor et al. 2013; Bloom et al. 2015) assessed the impact of the 2006 English NHS patient choice reforms on care quality by using mortality-based performance indicators as proxies for overall hospital quality, and obtained results consistent with the basic theoretical prediction just outlined – competition resulting from the 2006 introduction of patient choice led to higher clinical quality as captured by mortality rates. By contrast, this paper puts forward a model in which the impact of hospital competition on product quality is differentiated by output type, and influenced by the complementarities and substitutabilities between output types in production. Informed by this framework, it uses PROMs to examine the impact of hospital competition on clinical quality at the level of the individual surgical procedure, finding that the competition brought about by patient choice may have led to *lower* clinical quality. Although the estimates reported here are provisional, they point to the need for follow-up work that studies the impact of market structure on hospital performance in a multi-product, value-added setting.

The remainder of the paper is structured as follows. Section Two describes the English NHS patient choice reforms, summarises the literature on hospital competition and clinical quality, and introduces Patient Reported Outcome Measures (PROMs). Section Three presents evidence that there is little correlation between a hospital's performance in relation to mortality, and its elective surgery quality as captured by PROMs health gains. This finding suggests that analysing the impact of hospital competition on quality by focusing exclusively on mortality rates, as the existing literature has done, potentially fails to take account of important dimensions of hospital performance, quality and productivity – and it provides a compelling rationale for looking at the impact of the introduction of patient choice of hospital for elective surgery within the English NHS using elective-surgery-specific outcome measures. Section Four extends existing theoretical models of hospital competition to a multi-product setting, and shows that competition might have a more ambiguous effect on product quality than is suggested by standard single-output-type models, depending on the observability of different outputs and the interaction between outputs in the hospital cost function. Section Five outlines the paper's identification strategy and measures of competition intensity, and presents the data. Section Six presents the results, while Section Seven discusses and concludes.

2 Policy background and literature review

2.1 Market-based reforms to the English NHS

The English NHS is funded by general taxation and offers health care that is largely free at the point of use. Before 1991, the Department of Health paid geographically-defined Health Authorities to directly manage hospitals. In 1991, the Conservative government made hospitals and other care providers into independent 'trusts', thus creating an NHS 'Internal Market' in which Health Authorities and GP 'fund holders' purchased care by entering into bulk contracts with providers.[3] Hospitals were encouraged to compete for contracts on price as well as on quality, yet there was virtually no publicly available information about the quality of care. This situation gave hospitals an incentive to compete on price *at the expense of* quality. Propper et al. (2004; 2008) find that, under the Internal Market, competition led to higher mortality rates for acute myocardial infarction (AMI, or heart attack), as well as shorter elective surgery waiting times. These findings, combined with earlier research showing that competition in the Internal Market led to lower costs and prices (Propper 1996; Propper et al. 1998; Söderlund et al. 1997), suggest that competition during this period led hospitals to focus on observable dimensions of performance (prices and waiting times) at the expense of unobservable dimensions (care quality, as measured by mortality rates).[4]

---

[2] One prominent recent exception is a Working Paper by Colla et al. (2016), whose concerns are complementary to those of this paper.

[3] While the Internal Market was justified using the rhetoric of choice and competition, patients had little say over where they were sent for care.

[4] This interpretation is consistent with predictions from the theoretical literature (Dranove & Satterthwaite 1992; 2000) that, when prices are flexible and signals of quality are poor, the effect of competition on hospital quality is likely to be negative. It can also be seen as an application of Holmstrom & Milgrom's (1991) multi-tasking result, in which incentivising observable dimensions of performance can lead to better or worse performance in unobservable dimensions, depending on whether there are cost complementarities or substitutabilities between observable and unobservable dimensions of performance.

On its election in 1997, the new Labour government declared the end of the Internal Market and announced that health policy would henceforth promote cooperation rather than competition. However, the institutional distinction between providers and purchasers was not abolished, and so the possibility of hospital competition remained, even though it was discouraged at the rhetorical level.

In 2002, the Labour government changed its position on markets within the NHS, and progressively reintroduced competition. This new era of hospital competition had four design pillars. First, price negotiation between providers and purchasers was replaced with a prospective reimbursement regime, Payment by Results (PBR), that paid hospitals a fixed price per procedure, with some adjustment for patient severity, local wage rates, and hospital characteristics. Secondly, a range of new providers (such as NHS Foundation Trusts, and Independent Sector Treatment Centres) were introduced alongside standard NHS trusts, with clearer incentives to increase their market shares. Thirdly, and at the centre of the reform programme, from January 2006 patients requiring elective surgery were entitled to a choice of four or five hospitals, including one private hospital, when booking their first outpatient appointment. From April 2008, patients could choose to be treated at any hospital in England, NHS or private, that was qualified to provide the procedure and willing to accept the standard NHS price. Fourthly, to facilitate informed choice, improved signals of quality were introduced via the establishment in 2007 of the NHS Choices website (http://www.nhs.uk), which provided users with a range of quantitative and qualitative information about the performance of alternative providers.

Compared with the 1990s NHS Internal Market, the hospital market established under Labour was a major improvement, with many design features reflecting an awareness of the factors that led the Internal Market to fail – poor producer and purchaser incentives, quality-reducing price competition, and poor information about quality. Existing econometric evidence, mostly using mortality-based outcome measures, suggests that the competitive reforms of the 2000s did lead to higher hospital quality. Cooper et al. (2011) and Gaynor et al. (2013), both using a DiD strategy, find that hospital competition led to larger reductions in 30-day in-hospital AMI mortality, and in-hospital mortality from all causes (whether in-hospital or after discharge), in high-competition areas than elsewhere. Bloom et al. (2015) use the percentage of parliamentary constituencies in a hospital's market as an instrument for competition intensity – the idea being that hospitals serving marginal constituencies are less likely to be closed, and therefore that markets with many marginal constituencies will be more competitive than other markets – finding that higher competition led to higher overall management quality, as well as lower AMI mortality.

While the econometric rigour of the three aforementioned studies (Cooper et al. 2011; Gaynor et al. 2013; Bloom et al. 2015) is widely acknowledged, there has been controversy over how to interpret their findings, above all because the clinical outcome variables they focus on – AMI or total mortality rates – are not directly relevant to the sphere of hospital activity, elective surgery, that was subject to the introduction of patient choice.[5] Peri-operative mortality is close to zero for most elective surgical procedures; most hospital deaths are instead related to chronic illnesses or terminal conditions, or else are admitted on an emergency basis (as is the case for AMI patients). Indeed, Gaynor et al. (2011) suggest that one way of interpreting these three studies is that increased competition led to higher emergency care quality, as this is where a significant percentage of hospital deaths occur. Thus, a reform affecting one area of hospital activity, elective surgery, appears to have led to performance improvements in other areas of hospital activity. The question naturally arises – how should we understand the link between the two? The argument must be that (1) hospital competition in elective surgery led to quality improvements in *elective surgery*, and that (2) this led to quality improvements in other areas of hospital activity either (2a) because clinical quality is a hospital-wide variable or has a hospital-wide component and/or (2b) because of positive spillovers between elective surgery quality and quality in other areas.

Claim (1), which is the major focus of this paper, has been a source of considerable controversy, yet remains largely unexamined by the econometric literature.[6] Many health policy analysts remain skeptical

---

[5] One argument in defence of these studies is that their focus on outcomes in areas of hospital activity not subject to competition was made deliberately, in order to attenuate bias from endogeneity between hospital quality and measures of market structure arising from patients' choices (Bloom et al. 2011). This is a valid rationale for adopting such an approach, but it does not obviate the difficulties of interpretation that flow from it.

[6] One study (Feng et al. 2015) examines the association between hospital competition in the English NHS and hip replacement surgery quality as captured by PROMs health gains in 2011/12, finding no statistically significant relationship between the two. However, they do not seek to draw causal inferences. By contrast, the present paper examines all four PROMs procedures using four years of data, and seeks to draw causal conclusions by using instrumental variables estimation to address potential sources of endogeneity.

that competition will improve clinical quality, because patients cannot or do not choose a hospital on this basis (Jones & Mays 2009; Fotaki et al. 2008; Colla et al. 2016). Moreover, it is unclear whether the preconditions of an effective hospital market existed in England after the introduction of patient choice. Early assessments suggested that implementation and awareness of the existence of patient choice were poor.[7] In addition, during the four-year period studied in this paper, the NHS Choices website did not report average hospital PROMs health gains to patients undergoing one of the four PROMs surgical procedures (and of course, for all other elective surgical procedures no such indicator of elective surgical quality was collected). Instead, visitors to the NHS Choices website were presented with an overwhelming and confusing list of dozens of different dimensions of comparison, including some clinically relevant variables, but also including numerous clinically irrelevant variables such as availability of car parking and the quality of hospital food catering.[8] Notwithstanding these deficiencies in official efforts to disseminate information about hospital quality, the patient choice reforms may have enabled patients and their General Practitioners to act on any private information they possessed about hospital quality when making a joint decision about referral. Moreover, the mere *threat* of losing market share as a consequence of the patient choice reforms may have spurred hospitals to improve clinical quality, even if few patients did make active choices. Whether the patient choice reforms did indeed lead to higher *elective surgery* quality is ultimately an empirical question – answering this question is the objective of this paper.

If claim (2a) were true, then the existing econometric assessments of the patient choice reforms, by showing that competition led to lower mortality, would have effectively proved claim (1). However, Section 3 shows that there is no evidence that hospital mortality rates are correlated with elective surgery quality as captured by PROMs health gains. Mortality rates cannot, therefore, be taken as a proxy for elective surgery quality.[9]

Claim (2b) is plausible, but it is equally possible that improving quality in one area of hospital activity could lead to a deterioration of quality in other areas (for example, via diversion of managerial attention). This possibility points to the need to develop a better understanding of the way in which changes to incentive structures in one area of a hospital might not only affect performance in that area, but might also have knock-on effects in other areas of hospital activity. As Propper (2012) writes, in the literature on hospital competition and quality there is a “black box” in our understanding of exactly what purchasers, managers, and clinical practitioners do in response to competition that affects outcomes. The finding by Bloom et al. (2015) that increased hospital competition leads to higher across-the-board management quality is presumably one important component of this black box – but the trade-offs faced by managers mean that improved overall management quality need not necessarily imply improved care quality in all areas of hospital activity. This paper aims to open up Propper's ‘black box’ by examining the effect of introducing patient choice of hospital for elective surgery on elective-surgery-specific outcome measures; in so doing, it also aims to shed light on the way in which changes in elective surgery quality are (or are not) transmitted to other parts of the hospital.

---

[7] Surveys indicated that only 30 percent of elective surgery patients recalled being offered choice in 2006, rising to only 47 percent in 2009. Similarly, in 2006, 29 percent of patients were aware they were entitled to a choice of hospital before visiting their GP, rising to only 50 percent in 2009 (Dixon et al. 2010).

[8] For example, a visitor in October 2012 would have found information about a limited set of relevant and procedure-specific variables (number of operations performed, rate of unplanned readmissions, average waiting times, and average time spent in hospital), interspersed with numerous hospital-wide clinical outcome measures (such as MRSA cases – an indicator of hospital cleanliness – and 30-day mortality rates), as well as a thicket of clinically irrelevant variables. Arguably, the presence of these clinically irrelevant details was harmful for two reasons. First, it distracted prospective patients' attention away from clinically relevant information. Secondly, it potentially encouraged hospitals to seek to attract patients by improving the quality of their food, or their number of parking spaces, rather than by improving (and possibly at the expense of) quality of care.

[9] An alternative and complementary approach to that adopted in this paper would be to examine the impact of the patient choice reforms by continuing to use mortality as an outcome measure, but focusing on the relatively limited number of elective surgical procedures for which there is a non-trivial risk of death. For example, Aylin et al. (2013) study five elective surgical procedures that each have mortality rates of between 2 per cent and 3.6 per cent. Mortality rates from these procedures could perhaps be used to study the impact of the patient choice reforms on elective surgery quality. However, one impediment to such a study is that many elective surgical procedures with high mortality rates are performed at only a small number of specialised hospitals, making it difficult to obtain statistically significant estimates. Moreover, even if such a research project were feasible, there would still be a strong case for examining the impact of hospital competition on elective surgery quality using alternative outcome measures that capture clinical quality in relation to the large majority of elective surgical procedures for which death is a rare occurrence.

2.2 Patient Reported Outcome Measures (PROMs)

PROMs are measures of health status or health-related quality of life, as reported by patients. They capture health status at a single point in time, the idea being to capture the *outcome* of a health intervention by surveying patients twice: before the intervention, and after the intervention. The change in health status is then taken as a measure of health gain from the intervention. While PROMs have long been used by clinicians to improve their treatment of individual ailments, only recently have policymakers recognised their potential for use in policy evaluation and performance measurement.[10]

In April 2009, the English NHS, after conducting a pilot programme (Smith et al. 2005; Browne et al. 2007), started collecting PROMs for four surgical procedures – hip replacement, knee replacement, groin hernia repair, and varicose vein treatment. Patients are given the generic EQ-5D survey of health-related quality of life (EuroQol Group 1990) at their pre-surgical assessment or on admission for surgery, and again either three or six months post-operatively (depending on the procedure undertaken). At the same time, patients for all but one of the procedures are given a procedure-specific survey – either the Oxford Hip Score (OHS), the Oxford Knee Score (OKS), or the Aberdeen Varicose Veins Questionnaire (AVVQ). Although completing the surveys is voluntary, it is believed that this is the world's first example of nationwide administrative distribution and collection of a PROM for any surgical procedure.

The EQ-5D survey is very widely used in the UK and Europe. The National Institute for Health and Care Excellence (NICE), which approves new medicines and devices for use within the English and Welsh NHS, states that the EQ-5D is "the preferred measure of health-related quality of life in adults" when conducting economic evaluation of health technologies (NICE 2013). The EQ-5D has two components. The first, the EQ-5D Visual Analogue Score (henceforth EQ-VAS), asks patients "how good or bad [their] health is today", on a scale of 0 (worst) to 100 (best). The second and more important component, the EQ-5D profile index score (henceforth EQ-5D), asks patients to indicate their current health status in five dimensions – mobility, ability to undertake self-care, ability to undertake usual activities, pain/discomfort, and anxiety/depression. In each dimension, patients choose from three options – 1 (no problems), 2 (some problems), or 3 (extreme problems) – giving $3^5 = 243$ possible permutations of response.[11] These 243 possible response profiles are then aggregated with weights obtained from population-level surveys, generating a single-dimensional measure of health-related quality of life, with 1 representing perfect health and 0 representing death.[12] The resulting measure can be interpreted cardinally – for example, a health state value of 0.07 means that 1 year lived in that state is equivalent to 0.07 of a year in perfect health.

The OHS, OKS and AVVQ, which ask procedure-specific questions, generally have a greater capacity than the EQ-5D to detect changes in health status resulting from surgery.[13] The OHS and OKS comprise 12 multiple-choice questions which confer between 0 and 4 points each, yielding an overall score between 0 (worst) and 48 (best).[14] The AVVQ comprises 13 multiple-choice questions which confer a certain number of points each, yielding an overall score between 0 (best) and 100 (worst).[15] In this paper, the

---

[10] PROMs are sometimes criticised on the grounds that, unlike 'objective' measures of health status, they are based on 'subjective' assessments by patients of how they are feeling. These assessments, it is sometimes argued, are not reliable, as they are subject to a range of psychological and cognitive biases. However, the incorporation of subjective health states into PROMs is not an undesirable epiphenomenon but is rather intrinsic to their very purpose, as PROMs are premised on the recognition that many individual symptoms of illness (e.g. amount of pain) are best assessed by the patient.

[11] A newer version of the EQ-5D, the EQ-5D-5L, allows patients to choose between five levels of health status in each dimension (Herdman et al. 2011). However, the NHS PROMs programme continues to use the three-level version in the questionnaires distributed to patients.

[12] Values below zero are possible, implying a health profile 'worse than death'. The NHS uses EQ-5D utility weights based on the Measuring and Valuing Health (MVH) study (NHS Digital 2016, p.37), a population-level survey of individuals' preferences concerning different dimensions of health (Dolan 1997). The MVH study surveyed 3,395 representative citizens of England, Wales and Scotland to obtain valuations of 42 representative health profiles using the time trade-off method – that is, respondents were asked how many years of life in the state of perfect health (11111) they considered equivalent to the profile in question. Valuations for the other 201 health profiles were then interpolated from the valuations elicited concerning these 42 health profiles.

[13] For the procedure-specific PROMs studied in this paper, the weights for each question are determined by clinicians rather than by surveying patients or citizens concerning their valuation of different health states. Consequently, the OHS, OKS and AVVQ are best thought of not as measures of health-related quality of life as evaluated by patients or citizens, but rather as clinically relevant measures of health gain from surgery.

[14] The original OHS and OKS score each question between 1 and 5, but for the NHS PROMs programme this is modified to scores between 0 and 4 (NHS Digital 2016).

[15] Due to rounding of the weights used for each question, the maximum AVVQ score is actually 99.658 (NHS Digital 2016).

scale of the AVVQ is reversed so that, like all other measures, higher scores denote better health. Table 2 reports summary statistics for all the outcome variables examined in this paper.

In 2016, the US Centers for Medicare & Medicaid Services (CMS) declared its intention to base reimbursement decisions on PROMs-based measures of clinical quality, as part of its commitment to link 50 per cent of Medicare payments for primary care to quality or value by 2018 (CMS 2016). Efforts to incorporate PROMs into CMS reimbursement decisions have been informed by English experience through a joint initiative between the NHS and the US Department of Health and Human Services (HealthIT.gov 2016). The University of Oxford has allowed royalty-free use of the OHS and OKS to participants in CMS quality reporting programs (HHS, CMS and Isis Innovation 2016), making it likely that these PROMs, which are a central focus of this paper, will be at the heart of CMS efforts to measure orthopaedic surgery quality for reimbursement.

The increasing international attention to PROMs can be understood as a response to the realisation that existing measures of health care provider performance both fail to capture health outcomes for many interventions, and are not robust to the influence of consumer characteristics and actions on outcomes. Grossman's seminal work on demand for health (Grossman 1972) posited that health care only confers utility insofar as it contributes to *health*, a non-market good produced by households using endowments and market goods, including health care. Two implications flow from this insight for performance assessment of health care providers. The first is that performance assessment should be based on the amount of health produced, not on the amount of health care produced – or in other words, on *outcomes* rather than *outputs*. Initial attempts in the 1980s and 1990s to measure health *outcomes* tended to equate health with the absence of sickness, focusing on outcomes such as mortality rates and readmission rates, because these measures could often be derived from administrative data sets. However, failure-based outcome measures of this kind convey only limited information, because, in most spheres of health care, events such as death and readmissions are relatively rare, and therefore "shed little light on the great majority of health service interventions for most patients" (Appleby and Devlin 2010, p.2; see also Shojania & Forster 2008). Table 1 presents average mortality rates for the elective surgical procedures studied in this paper, with mortality rates for AMI as comparator. All four elective procedures have mortality rates of close to zero (or zero) – yet, amongst the vast majority of patients who do not die when undergoing these procedures, health outcomes vary greatly. Mortality-based indicators of hospital performance do not directly capture this variation.[16]

By contrast, the EQ-5D, OHS, OKS and AVVQ have been extensively validated using standard psychometric tests as tools for capturing health gains from the four NHS PROMs procedures (Smith et al. 2005). Table 3 reports the effect sizes – a measure of responsiveness to an intervention, equal to average health gain divided by standard deviation of pre-operative score – of the PROMs studied in this paper. An effect size of 0.2 is considered low, while 0.5 is considered moderate, and 0.8 is considered large (Smith et al. 2005). The procedure-specific PROMs – the OHS, OKS, and AVVQ – have effect sizes ranging from 0.726 to 2.377. The EQ-5D index score performs very well for hip and knee replacement surgery (0.954-1.271), but only moderately well for the other procedures (0.401-0.411). By contrast, the EQ-VAS performs poorly in relation to all procedures other than hip replacement. Pooling all PROMs procedures together yields an average effect size of 0.890 for the EQ-5D index score and 0.212 for the EQ-VAS. This paper designates all PROMs with large effect sizes – the procedure-specific PROMs and the EQ-5D for hip and knee replacement, as well as for all PROMs procedures pooled together – as its main outcome variables of interest, and reports estimates using the other outcome variables in robustness tests. Overall, these effect sizes show that the PROMs collected by the English NHS contain meaningful variation capable of detecting changes in health gain resulting from differential exposure to competition.

A good outcome measure should be correlated with other validated outcome measures: a correlation of 0.2 or above is taken as evidence of convergent (or concurrent) validity (Smith et al. 2005). One can get a sense of the convergent validity of each PROM by checking how closely correlated it is with other PROMs for that procedure. Table 4 shows that, with one exception, all such correlations are above 0.2. There is a particularly strong correlation between the EQ-5D and the Oxford Hip and Knee Scores.

---

[16] While it is possible that mortality rates are correlated with other negative outcomes for the larger subset of elective surgery patients who do not die, the very low mortality rates reported in Table 1 suggest that, for these procedures, mortality is unlikely to effectively signal quality at conventional levels of statistical significance.

These correlations provide evidence that the PROMs studied in this paper are capturing a coherent, underlying concept of health gain from surgery.[17]

The second implication of the Grossman model for health care provider performance measurement is that it must engage with the challenge of casemix adjustment, in order to strip out the confounding effect of patient characteristics and actions on outcomes. The fact that PROMs base health outcomes on the change in health status resulting from treatment makes them 'value-added' measures of health outcomes akin to those increasingly used to evaluate school outcomes in the USA and UK (Kane & Staiger 2008; Rothstein 2010; Chetty et al. 2014; Gibbons et al. 2013a; 2013b). The idea underlying the value-added approach is that the baseline (pre-treatment health status or past exam results) can be used as a sufficient statistic for aspects of prior health or learning that affect current outcomes (Koedel et al. 2015). The pre-treatment PROMs questionnaire thus greatly assists the challenge of disentangling the hospital's contribution to post-treatment health status from the influence of patient characteristics and behaviour.

3 Evidence on within-hospital correlation in quality levels

If hospital mortality rates were correlated with average health gain from elective surgery, then there would be little reason to re-examine the 2006 patient choice reforms using elective-surgery-specific outcome measures, as the existing literature (Cooper et al. 2011; Gaynor et al. 2013; Bloom et al. 2015), by measuring the impact of hospital competition using mortality-based quality indicators, would have effectively also been capturing the impact on elective surgery quality. If, on the other hand, these two dimensions of hospital performance are uncorrelated or weakly correlated, there is a case for looking again at the introduction of patient choice using outcome measures specific to the area of the hospital that was directly affected by this reform – namely elective surgery (Bevan and Skellern 2011). To this end, this section presents evidence on the relationship between elective surgery quality, as measured by EQ-5D health gains, and various mortality-based indicators of hospital performance.

Figure 1 shows a scatter plot of the relationship between hospital trusts'[18] casemix-adjusted mortality ratios and average casemix-adjusted PROMs health gains from elective surgery.[19] Table 5 reports the corresponding correlations for both raw (unadjusted) and risk-adjusted PROMs health gains. There appears to be a small *positive* correlation between hospital trusts' standardised mortality rates (SHM) and unadjusted PROMs health gains for orthopaedic surgery. When casemix-adjusted PROMs health gains are used, this correlation is largely eliminated, but there is now a *positive* correlation between trusts' SHM and average PROMs health gains for varicose vein stripping, which can also be seen in Figure 1. Bivariate regressions of the log of trusts' standardised mortality ratio on the log of trusts' average EQ-5D health gains, reported in Table A4, confirms the overall picture – there is no evidence that hospital performance in relation to mortality is positively correlated with performance in relation to health gains from elective surgery, and indeed there is some evidence that performance in these two dimensions is negatively correlated.

I next present data on the relationship between PROMs health gains and AMI mortality at the hospital site level.[20] Whereas there are questions about the capacity of standardised mortality indicators

---

[17] There is no relationship between hospital mortality rates for the PROMs procedures and average adjusted PROMs health gains. This is because death is an extremely rare outcome for all four PROMs procedures.

[18] In the English NHS, hospital trusts are administrative and financial entities that may include a number of different hospital sites. Most of the analysis in this paper is conducted at the level of individual hospital sites. However, standardised mortality rates are analysed at the trust level, because this is the level at which these data are published.

[19] From 2010/11 onwards, my measure of risk-adjusted hospital mortality is the official NHS Summary Hospital Mortality Indicator or SHMI (HSCIC 2013). For 2009/10, before the SHMI was created, I use the Hospital Standardised Mortality Ratios (HSMR) published by Dr Foster (Dr Foster 2011), divided by 100 to make its scale comparable with the SHMI. Although the HSMR and SHMI are calculated in different ways, they produce similar outputs, namely a number, generally ranging between 0.7 and 1.2 (or 70 and 120), which reports the ratio of actual to expected deaths, with a value lower than one indicating fewer deaths than expected, and a value greater than one indicating more deaths than expected. When reporting correlations with *trust*-level Standardised Mortality Rates, I use official NHS average (adjusted and unadjusted) trust-level PROMs health gains. By contrast, average adjusted and unadjusted PROMs health gains at the hospital *site* level are calculated by the author from patient-level data. The patient-level casemix adjustment strategy is outlined in Appendix 1.

[20] Following Cooper et al. (2011), to maximise comparability between hospitals the sample of AMI patients is restricted to include only patients aged between 39 and 100 that were admitted to hospital on an emergency basis from their permanent or temporary place of residence. To avoid possible bias due to upcoding of diagnoses, patients discharged alive with a total length of stay of less than three days are discarded.

to meaningfully capture differences in care quality (Black 2010; Lilford & Pronovost 2010), there is a clear and well-documented link between AMI survival rates and the quality and timeliness of care (Bradley et al. 2006; Jha et al. 2007). Thus, if the qualities of different treatments within a given hospital are correlated with each other, there should be a negative correlation between AMI mortality and adjusted PROMs health gains. Figure 2 presents scatter plots of hospital sites' AMI mortality rates and adjusted PROMs health gains. Table 6 reports the corresponding correlations for both raw and risk-adjusted PROMs health gains. There appears to be no relationship between the quality of a hospital's elective surgery and its AMI mortality rate. Simple bivariate regressions of the log of hospitals' AMI mortality on the log of average EQ-5D health gain, reported in Table A5, show no significant relationships.

Even if there is little or no cross-sectional relationship between a hospital's mortality rates and elective surgical quality as captured by PROMs health gains, such a relationship may exist in first differences – for example, quality improvements in one section of the hospital may be transmitted to other sections of the hospital, even if some sections offer high quality care while others offer low quality care. To investigate this hypothesis, Table 7 presents correlations between hospital trusts' year-on-year *change in* average risk-adjusted PROMs health gains and *change in* standardised mortality rates, while Table 8 presents correlations between hospital sites' year-on-year *change in* adjusted PROMs health gains and *change in* AMI mortality.[21] Neither table indicates the hypothesised negative correlation between first differenced mortality rates and first differenced adjusted PROMs health gains, and simple bivariate regressions in logs, reported in Tables A6 and A7, confirm no statistically significant relationships at the 5 per cent level.[22]

The apparent lack of any negative relationship between hospital mortality rates and average PROMs health gains – either cross-sectionally or in first differences – highlights the importance of moving beyond an exclusive focus on mortality-based outcome measures when assessing the impact of changes to hospital incentive structures on hospital productivity and clinical quality. It is, moreover, consistent with the finding of inspections of clinical governance in the English NHS in the early 2000s by the Commission for Health Improvement, a predecessor of the Care Quality Commission, that no large acute hospital "performed well across the board" – they typically had "a mix of good and poor services, often with a dysfunctional clinical team" (Bevan & Cornwell 2006, p.359).[23] While some might be tempted to explain this lack of relationship by arguing that PROMs health gains are just random noise, such a dismissal would be unconvincing given the evidence presented in Section 2.2 that the PROMs data *does* capture meaningful variation in health gain from elective surgery. Of course, preventing patient death will always be a key indicator of hospital performance – but Section 3's findings suggest that mortality rates capture just one dimension of quality, and that focusing on them to the exclusion of other dimensions can provide an incomplete picture, especially when the changes being studied target a section of the hospital where mortality is a rare event.

4 Multi-good models of hospital competition with fixed prices

This section presents a theoretical framework to motivate this paper's examination of the 2006 English patient choice reforms using elective-surgery-specific outcome measures. Standard economic models of hospital competition with fixed prices (Gaynor 2006; Gaynor & Town 2012) assume that hospitals produce a single type of output and choose a single, vertical quality level. These models predict that, so long as the regulated price exceeds the marginal cost with respect to quantity, increased competition intensity will lead to higher hospital quality. This section extends this standard model to a setting where a given hospital j produces two types of output – elective surgery ($x = 1$) and emergency care ($x = 2$)

---

[21] This paper conducts its analysis at the financial year level. The UK financial year runs from 5 April until 4 April. For the purpose of this paper, I define a financial year as running from 1 April until 31 March. All references to years in this paper refer to financial years.

[22] For brevity, I do not present graphical evidence of the relationship between first-differenced mortality rates and first-differenced PROMs health gains, nor do I present correlations between first-differenced mortality and first-differenced *unadjusted* PROMs health gains. The graphs indicate that there is no relationship between these variables in first differences, while the unadjusted correlations show a qualitatively similar picture to the adjusted correlations.

[23] This evidence is also consistent with the findings of the New York Cardiac Surgery Reporting System, which showed that there were substantial differences in mortality rates *between individual surgeons at the same hospital* (Chassin 2002). If outcomes can vary so substantially *within* a given hospital department, then it seems clear that outcomes *between* hospital departments cannot necessarily be assumed to be correlated.

– with associated quality levels z_{j1} and z_{j2}.[24] Elective surgery is assumed to be subject to competition, while emergency care is not. The model initially assumes that quality of both elective surgery and emergency care is observable (as captured, for example, by PROMs health gains and mortality rates respectively). In light of the fact that PROMs health gains from elective surgery were not reported on the NHS Choices website doing the period of study, it then considers what happens if quality of elective surgery is unobservable.

Prices for each output type, $\bar{p}_1$ and $\bar{p}_2$, are fixed and paid by the government. The demand experienced by hospital j for output x, q_{jx}, is equal to market share s_{jx} multiplied by overall market demand D_x: $q_{jx} = s_{jx}D_x$. As NHS patients do not face any of the costs associated with hospitalisation, market shares, as well as overall market demand for each good, are independent of prices; overall market demand is a function only of exogenous demand shifters θ_x (e.g. illness): $D_x = D_x(\theta_x)$.

For elective surgery, market share $s_{j1}(z_{j1}, \boldsymbol{z}_{-j1}, N)$ is a function of the number of hospitals in the market N, own electives quality z_{j1}, and the vector of electives quality of all other hospitals $\boldsymbol{z}_{-j1}$, with $\frac{\partial s_{j1}}{\partial z_{j1}} > 0$, $\frac{\partial s_{j1}}{\partial z_{k1}} \leq \mathbf{0}\ \forall k \neq j$, and $\frac{\partial s_{j1}}{\partial N} < 0$. That is, electives market share is increasing in own electives quality, weakly decreasing in electives quality of all other hospitals, and decreasing in the number of competitors. Increased hospital competition is represented in the model as an increase in N.[25] Crucially, it is also assumed that $\frac{\partial^2 s_{j1}}{\partial z_{j1}\partial N} > 0$ – the sensitivity of market share to own quality is increasing in the number of competitors. Emergency patients are assumed to simply attend the nearest appropriate hospital, so demand is not a function of emergency care quality: $q_{j2} = s_{j2}D_2(\theta_2)$.

The two output types interact via the cost structure: the cost of producing each output type is dependent not only on the quality of that output type, but also on the quality of the other output type. In this way, the model aims to capture possible complementarities and substitutabilities between output types in production. Total cost of producing output type x is $c_{jx} = c_x(q_{jx}, z_{jx}, z_{j,-x}, W_x) + F_x$, where W_x denotes exogenous cost shifters and F_x denotes fixed costs. If $\frac{\partial c_x}{\partial z_{-x}} < 0$ and $\frac{\partial^2 c_x}{\partial z_x \partial z_{-x}} < 0$, the output types are cost complements – or in other words, there are economies of scope between the two output types. If $\frac{\partial c_x}{\partial z_{-x}} > 0$ and $\frac{\partial^2 c_x}{\partial z_x \partial z_{-x}} > 0$, the output types are cost substitutes, or in other words, there are diseconomies of scope.[26]

NHS hospitals are not profit-maximisers, but do have an incentive to generate operating surpluses (that is, profits), or at least not to run deficits.[27] In addition, hospital managers are assumed to value the provision of quality in its own right, whether for altruistic or other (e.g. reputational) reasons, and are therefore assumed to maximise some combination of profits and quality – $U_j = u(\pi_j, z_{j1}, z_{j2})$. For simplicity, managerial utility is assumed to be additively separable in all arguments, so $U_j = \pi_j + v_1(z_{j1}) + v_2(z_{j2})$. The hospital's problem is therefore:

$$\max_{z_{j1}, z_{j2}} U_j = \bar{p}_1[s_{j1}(z_{j1}, \boldsymbol{z}_{-j1}, N)D_1(\theta_1)] + \bar{p}_2[s_{j2}D_2(\theta_2)] + \sum_{x=1}^{2}[v_x(z_{jx}) - c(q_{jx}, z_{jx}, z_{j,-x}, W) - F_x]$$

---

[24] The model presented here is an extension of that presented in Gaynor et al. (2011), which includes two output types but assumes that the hospital chooses a single level of quality that is common to both output types.

[25] The main reason for representing the increase in competition resulting from the patient choice reforms as an increase in N is that moving from the previous regime of selective contracting, in which a patient's choices are restricted to the hospitals with whom their care purchaser maintains a bulk contract, to free patient choice of hospital, involves an expansion in patients' choice sets, even if no new providers actually enter the market. Additionally, new provider entry did occur alongside the patient choice reforms, as a consequence of the establishment of privately owned and managed specialty surgical centres (Independent Sector Treatment Centres) for the provision of routine diagnostic procedures and treatments (Cooper et al. 2016).

[26] For example, consider $c_1 = q_1(z_1^2 + \phi z_1 z_2)$, so $\frac{\partial c_1}{\partial z_2} = \phi q_1 z_1$ and $\frac{\partial^2 c_1}{\partial z_1 \partial z_2} = \phi q_1$. If ϕ is positive, the output types are cost substitutes; if ϕ is negative, they are cost complements. There does not seem to be any reason, *ex ante*, to assume that hospital outputs are more likely to be cost substitutes or cost complements – one can easily think of reasons why both types of relationship might arise. For example, hospital outputs might be cost complements because innovations in one part of the hospital can be translated to other parts of the hospital. Alternatively, hospital outputs might be cost substitutes because of limited managerial attention, so that quality increases in one part of the hospital can only come at the expense of quality in other parts of the hospital.

[27] Hospitals with Foundation Trust (FT) status could retain any surplus generated within a financial year for investment as they saw fit; operating surpluses therefore enabled them to finance whatever other objectives they may have had. Hospitals without FT status could not retain surpluses, but were assessed for FT status in part on their financial performance, so they too had an incentive to run surpluses, or at least to avoid deficits.

Dropping the j subscripts, the hospital's two first order conditions (FOC) are:

$$\begin{array}{lccc} z_1 \text{ (electives care quality):} & \left[\bar{p}_1 - \frac{\partial c_1}{\partial q_1}\right] \frac{\partial s_1(z_1^*)}{\partial z_1} D_1(\theta_1) + \frac{\partial v_1(z_1^*)}{\partial z_1} & = & \frac{\partial c_1(z_1^*, z_2^*)}{\partial z_1} + \frac{\partial c_2(z_1^*, z_2^*)}{\partial z_1} \\ z_2 \text{ (emergency care quality):} & \frac{\partial v_2(z_2^*)}{\partial z_2} & = & \frac{\partial c_1(z_1^*, z_2^*)}{\partial z_2} + \frac{\partial c_2(z_1^*, z_2^*)}{\partial z_2} \end{array} \quad (1)$$

In the first FOC, the left hand side denotes the marginal benefit of providing elective surgery quality – the first term is the marginal monetary benefit, which is proportional to the gap between the regulated price and marginal cost, while the second term is the marginal altruistic benefit. The right hand side denotes the marginal cost of providing elective surgery quality. The first FOC implies that, subject to $\bar{p}_1$ greater than marginal cost, an increase in competition (increase in N) leads to unambiguously higher elective surgery quality by increasing the sensitivity of market share to electives quality ($\frac{\partial^2 s_1(z_1^*)}{\partial z_1 \partial N} > 0$).

The effect of increased competition on emergency care quality, however, is not so clear-cut. The second FOC, for emergency care quality, shows that, if the two types of output are cost complements ($\frac{\partial^2 c_1}{\partial z_1 \partial z_2} < 0$: economies of scope), the increase in elective surgery quality reduces marginal costs, implying an increase in z_2. If, on the other hand, the two output types are cost substitutes ($\frac{\partial^2 c_1}{\partial z_1 \partial z_2} > 0$: diseconomies of scope), the increase in z_1 leads to higher marginal costs, implying a *decrease* in z_2. Thus, increased competition leads to higher emergency care quality if the two outputs are cost complements, but lower emergency care quality if the two outputs are cost substitutes.[28]

One way of interpreting the studies of the English patient choice reforms focused on mortality-based quality indicators (Cooper et al. 2011; Gaynor et al. 2013; Bloom et al. 2015) is that increased competition led to higher emergency care quality, as this is where a large percentage of hospital deaths occur. On this interpretation, the first order conditions just presented imply that elective surgery and emergency care must be cost complements, as quality in elective surgery should have unambiguously increased. However, the model on which these first order conditions are based assumes that the quality of both output types is observable. If the quality of elective surgery is unobservable – as was arguably the case during the period under consideration, given that PROMs health gains were not published on the NHS Choices website – then it cannot influence electives demand. In other words, $\frac{\partial s_{j1}}{\partial z_{k1}} = \mathbf{0} \ \forall k$. In this case, competition will have no impact on the quality of either output, as it leads to first order conditions in which the quality of both outputs is set by simply equating the marginal altruistic benefit of quality with the marginal cost:

$$\frac{\partial v_x(z_x^*)}{\partial z_x} = \frac{\partial c_x(z_x^*, z_{-x}^*)}{\partial z_x} + \frac{\partial c_{-x}(z_{-x}^*, z_x^*)}{\partial z_x} \ \forall x = 1, 2 \quad (2)$$

If, in addition to elective surgery quality being unobservable, patients (rightly or wrongly) take emergency care quality as a proxy for elective care quality – for example, because when they go to the NHS Choices website to learn about options for their elective surgery procedure, hospital standardised mortality rates are listed as one of the bases for comparison – then the model should be modified to assume that $\frac{\partial s_{j1}}{\partial z_{k1}} = \mathbf{0} \ \forall k$, $\frac{\partial s_{j1}}{\partial z_{j2}} > 0$, $\frac{\partial s_{j1}}{\partial z_{k2}} \leq \mathbf{0} \ \forall k \neq j$, and $\frac{\partial^2 s_{j1}}{\partial z_{j2} \partial N} > 0$. That is, electives market share is unresponsive to own elective surgery quality, increasing in own emergency care quality, weakly decreasing in emergency care quality of all other hospitals, and more responsive to own emergency care quality when competition intensity (N) is higher. The first order conditions for the hospital's optimisation problem are now:

$$\begin{array}{lccc} z_1 \text{ (electives care quality):} & \frac{\partial v_1(z_1^*)}{\partial z_1} & = & \frac{\partial c_1(z_1^*, z_2^*)}{\partial z_1} + \frac{\partial c_2(z_1^*, z_2^*)}{\partial z_1} \\ z_2 \text{ (emergency care quality):} & \left[\bar{p}_2 - \frac{\partial c_2}{\partial q_2}\right] \frac{\partial s_1(z_1^*)}{\partial z_2} D_1(\theta_1) + \frac{\partial v_2(z_2^*)}{\partial z_2} & = & \frac{\partial c_1(z_1^*, z_2^*)}{\partial z_2} + \frac{\partial c_2(z_1^*, z_2^*)}{\partial z_2} \end{array} \quad (3)$$

These first order conditions imply that increasing competition in elective surgery leads to unambiguously higher emergency care quality but, perversely, to *lower* elective surgery quality if there are diseconomies

---

[28] This result can be understood as an application of the Holmstrom-Milgrom (1991) multi-tasking model. Instead of z_2 being unobservable, as in the Holmstrom-Milgrom model, the problem is that it is not possible to incentivise improvements in z_2 because demand for emergency care is inelastic (patients are simply sent to the closest appropriate hospital). As in the Holmstrom-Milgrom setting, the inability to incentivise z_2 means that only incentivising z_1 will have a negative effect on z_2 if the two activities are cost substitutes. The essential message of this model, for the purpose of empirical studies of hospital competition of quality, is that assuming *ex ante* that the quality of emergency care – which can reasonably be captured by a hospital's total or AMI mortality rate – is either identical to, or a proxy for, the quality of elective surgery elides potentially important issues concerning the interaction between production of different hospital outputs.

of scope (i.e. if the two types of hospital output are cost substitutes).[29] Recent empirical evidence suggests that there are in fact diseconomies of scope between emergency and elective care in English NHS hospitals (Freeman et al. 2016).[30] These recent findings provide a further argument in support of looking at the effect of the 2006 patient choice reforms on elective-surgery-specific quality measures, rather than assuming *ex ante* that any changes to mortality-based performance indicators resulting from competition will also have occurred in relation to elective surgery.

5 Identification strategy and data

5.1 Regression specification

The English NHS has collected PROMs since April 2009, while patient choice of hospital for elective surgery dates from 2006. It is not therefore possible to estimate the effect of patient choice on PROMs using the a DiD-style methods of Cooper et al. (2011) and Gaynor et al. (2013). Cross-sectional variation in treatment intensity does exist, however, because the strength of competition to which a hospital is exposed varies geographically – the reform had a greater impact on hospitals that had many competitors in their nearby vicinity than on hospitals with few competitors in their nearby vicinity, because it is easier for patients of the former to switch to an alternative hospital. This paper identifies the effect of hospital competition on care quality using this cross-sectional variation, by running the following regression for patient i undertaking procedure p[31] at hospital site j[32] and year t:

$$\begin{aligned}\text{gain}_{ijpt} = \beta_0 + \beta_1 \text{ comp}_{jt} + \beta_2 \text{ cases}_{jpt} + \beta_3 \text{ cases}^2_{jpt} + \beta_4 \text{ admissions}_{jt} + \beta_5 \text{ admissions}^2_{jt} \\ + \boldsymbol{\beta}_6' \boldsymbol{X}_{ijpt} + \boldsymbol{\beta}_7' \boldsymbol{Y}_{jt} + \boldsymbol{\beta}_8' \boldsymbol{Z}_j + u_{ijpt}\end{aligned} \tag{4}$$

The left hand side variable is (casemix-adjusted) health gain from surgery as captured by PROMs, while comp_{jt} is the competition intensity experienced by hospital j at time t, cases_{jpt} is the number of cases for that hospital site-procedure-year, admissions_{jt} is total annual admissions per trust, and u_{ijpt} is an error term.[33] $\boldsymbol{X}_{ijpt}$ denotes a vector of patient-level controls, $\boldsymbol{Y}_{jt}$ denotes time-varying hospital-level controls, and $\boldsymbol{Z}_j$ denotes time-invariant hospital level controls; the contents of these vectors are defined in Section 5.5. All regressions cluster standard errors at the hospital level.

The coefficient of interest is β_1, the effect of competition intensity on casemix-adjusted health gain from surgery. Actual competition intensity between hospitals cannot be measured; consequently, all measures of competition intensity are measures of market structure, or of the *potential* for competition. This paper measures competition using the (negative log) Herfindahl-Hirschman Index (HHI), which is equal to the sum of squared market shares of each competitor in the market.[34] A separate HHI is calculated for each financial year for each of six high-volume elective surgical procedures: the four PROMs procedures, plus knee arthroscopy and cataract repair, using a definition of market size that is discussed

---

[29] It is important to be clear that this hypothesis of the model just presented is not consistent with rational expectations, which imply that patients should not choose which hospital to attend for elective surgery on the basis of emergency care quality if the latter has a negative relationship with elective surgery quality. Nonetheless, the hypothesis is plausible given the range of information available to elective surgery patients on the NHS Choices website during the study period.

[30] Freeman et al. (2016) study the relationship between emergency and elective care costs and volumes, finding that increased elective surgery volumes are associated with an increase in emergency care costs. While Freeman et al. only look at cost-quantity relationships, their findings suggest that there may also be diseconomies of scope between these two output types in relation to vertical product quality.

[31] Hip and knee replacement observations (including both primary procedures and revisions) are analysed together because they are all performed by a hospital's orthopaedic department. In all orthopaedic surgery regressions, a dummy variable for knee replacement is included, as are separate dummy variables for revisions to hip and knee replacements, to capture level differences in health gains between these four surgical procedures. Estimates from running regressions separately for hip and knee replacement are reported in Table 15.

[32] NHS hospital trusts often consist of multiple hospital sites which can be located up to 100km from each other. Although finances are managed at the trust level, individual hospital sites within a hospital trust are for many purposes run independently, and can act as effective competitors with each other. For this reason, analysis is conducted at the hospital site level rather than trust level. Unlike the HES trust code field, which is always complete, the site code field is missing in approximately 10 percent of cases, and contains invalid data in approximately 10 percent more. In the vast majority of such cases, however, it is possible to impute the correct site codes with certainty, for example when only one site within a trust performs a given procedure. In the small number of remaining cases – around 4.4 percent – site codes are randomly imputed from a list of all sites in a trust that perform the procedure in question.

[33] Total and procedure-specific hospital admissions may be influenced by hospital quality after the introduction of patient choice; this issue is discussed in Section 5.3.

[34] Logs are taken to capture the idea that treatment effects will be constant with respect to percentage changes in competition intensity; the scale is reversed so that a higher value of comp_{jt} denotes higher competition intensity.

below. An overall HHI for each hospital and financial year is then created by taking a weighted average of the procedure-specific HHIs. Although treatment intensity is equated with a financial-year-specific HHI, for reasons discussed below, in our main specification we instrument this current-period HHI by the average value of the HHI in the three financial years leading up to the introduction of patient choice of hospital (2002/3 to 2004/5).

The most worrisome potential sources of endogeneity when estimating the effect of hospital competition on care quality fall under two broad headings. The first is endogeneity caused by patient choice behaviour, and encompasses selection bias (in which choice of hospital is driven by unobservable patient characteristics that influence outcomes) and reverse causality (in which patient choices influence one's measure of competition intensity). The second is omitted variable bias due to unobserved geographical correlates of competition intensity. The next section outlines how I use instrumental variables to address bias arising from patient choices. After that, I show how value-added measures of health outcomes control for the most worrisome geographical correlate of competition – namely patient health status.

5.2 Instruments for competition to address endogeneity due to patient choice behaviour

When patients choose a hospital, their choices may be systematically influenced by patient-level characteristics that influence outcomes. If sicker patients select disproportionately into attending higher-quality hospitals because they have more at stake from surgery, and if quality is equated with post-treatment health status, then the observed distribution of hospital quality will be compressed or reversed relative to the true distribution. For example, Great Ormond Street Hospital has one of the highest child mortality rates of any hospital in England, but this is because the sickest children are sent there, not because it is a poor quality hospital. While the sign of this selection bias is not so clear cut when value-added outcome measures are used, it is nonetheless a significant potential source of endogeneity.[35]

As well as leading to selection bias, patients' choices can also lead to reverse causality as a result of the way in which competition intensity is calculated. The primary methodological challenge in measuring competition intensity is how to define the size of the market within which a given hospital operates. A common approach is to centre markets on hospitals and to assume that hospital j's market includes all hospitals within the radius required to encompass the home address of a certain percentage (e.g. 95%) of j's patients. The problem with these 'variable radius' methods is that, when patients can choose which hospital they attend, percentiles of patient distance travelled will in general be endogenous to hospital quality. For example, a high quality hospital may attract patients from farther afield, thus giving it a larger market radius, and making it appear more competitive.[36] This is an example of reverse causality, as the objective is to estimate the causal effect of competition intensity on hospital quality, but hospital quality is now influencing (one's definition of) competition intensity; estimates of the effect of competition on quality using standard regression methods will therefore be biased.

This paper uses three different strategies to address these dual sources of endogeneity arising from the effect of patient choices. Firstly, to ameliorate the problem of reverse causality arising from hospital-centred definitions of market size, this paper's main measures of competition intensity instead centre hospital markets on patients' neighbourhoods of residence. A patient's neighbourhood is defined as their Middle Super Output Area (MSOA), a geographical statistical unit that usually contains between 6,000 and 9,000 residents.[37] Competition measures centred on an exogenously defined 'neighbourhood' will still be influenced by patient choices, but are less subject to bias due to reverse causality because it is hard to think of reasons why the direction of any such influence would be systematic, as it likely would be with a variable radius definition of market size.

---

[35] This problem of selection bias can be understood as a form of omitted variable bias because, if all relevant patient characteristics could were observable, the problem could be eliminated by controlling for these patient characteristics in one's regressions.

[36] The bias could, in theory, run in either direction. A counter-example would be a hospital offering such high quality of care that all patients living nearby attend the hospital, resulting in a very small market size when measured using percentiles of patient distance travelled. While the example is perhaps far-fetched in relation to hospital markets, this is exactly what happens in many British state schools, where places are allocated on the basis of distance from home to school.

[37] England had 6,791 MSOAs in 2011. MSOA boundaries are kept as stable as possible, but are redefined as required to keep MSOA populations between 5,000 and 15,000.

Secondly, to further address reverse causality, this paper instruments its current-period measure of market structure with the average value of market structure for the three years preceding the introduction of patient choice, 2002/3 to 2004/5.[38] As patients could not choose which hospital they attended during these years, instrumenting the HHI with its average pre-reform level should address any concerns that competition intensity is partly a function of hospital quality.

Thirdly, to address selection bias, I use a conditional logit model (Kessler & McClellan 2000; Gaynor et al. 2013) to predict patient choice of hospital on the basis of exogenous variables, and calculate the HHI using these predicted patient choices, rather than actual patient choices. Let patient i's utility from prospective hospital j be an additive function of a systematic component $\boldsymbol{\theta}_j'\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij}$ and a random component ε_{ij}. The vector $\boldsymbol{x}_i$ – with hospital-specific coefficients $\boldsymbol{\theta}_j$ – contains individual-specific explanatory variables, while the vector $\boldsymbol{z}_{ij}$ – with coefficients $\boldsymbol{\gamma}$ – denotes hospital-level or patient-level variables that are a function of choice of hospital.

$$U_{ij} = \boldsymbol{\theta}_j'\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij} + \varepsilon_{ij}$$

The probability that i will choose j, π_{ij}, is the probability that j is the utility maximising choice of hospital:

$$\pi_{ij} = \Pr(H_i = j) = \Pr\left(\max\{U_{i1}, U_{i2}, \ldots, U_{iJ}\} = U_{ij}\right)$$

If it is assumed that ε_{ij} is distributed standard Type 1 extreme value with cumulative distribution function $F(\varepsilon) = \exp\left(-\exp\left(-\varepsilon\right)\right)$, then it can be shown (Maddala 1983; for further details, see Appendix 2) that:

$$\pi_{ij} = \frac{\exp(\boldsymbol{\theta}_j'\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij})}{\sum_{\lambda=1}^{J} \exp\left(\boldsymbol{\theta}_j'\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij}\right)}$$

The parameters of the model are estimated by maximum likelihood separately for each surgical procedure and financial year in the dataset. For each patient, the model gives a probability of attendance at each hospital. These probabilities, which sum to 1 for each individual, are used to calculate predicted market shares in each MSOA for each hospital. To calculate this competition index, for each of the six elective surgical procedure listed above, I first calculate a procedure-MSOA-year-level HHI equal to the sum of each provider's predicted squared market shares in the MSOA. I then create a procedure-hospital-year-level HHI equal to the weighted average of the HHIs of all the MSOAs that it serves, and finally a single HHI for each hospital and year as a weighted average of the procedure-specific HHIs.

HHIs based on predicted patient choices are essentially a complex form of instrumentation, in which instead of instrumenting competition intensity itself, each patient's choice of hospital is instrumented, and the resulting predicted patient choices are then used as inputs into the construction of competition indices.[39] This approach eliminates any influence of hospital quality on patient decisions, thus comprehensively addressing the problem of selection bias, as well as addressing any residual concerns about reverse causality.

Patient distance to hospital is a critical variable in predicted patient choice models for two reasons. The first is that, as patients generally bear some or all of the travel costs incurred when obtaining treatment, a patient's distance to a given hospital is the biggest predictor of whether they will attend that hospital. This study specifies a choice model in which a patient's utility from attending a given hospital is not only a function of distance to that hospital, but is also a function of the difference between

---

[38] As patients could not choose which hospital they attended, a HHI calculated during these years is best understood as a measure of market structure – or alternatively, of the potential for competition – rather than of the intensity of actually operative competition.

[39] Given the complex functional form of this instrument, it must be estimated in two separate stages. The first stage estimates patient choices on the basis of exogenous variables, and constructs competition indices on the basis of these choices, while the second regresses hospital quality on our measure of competition intensity constructed using predicted patient choices. Performing instrumental variables estimation in two separate stages will lead to incorrect standard errors in the second stage, as the standard errors from the first stage regression are not taken into account. Gaynor et al. (2013) investigate the severity of this problem in relation to HHI indices based on predicted patient choices by generating ten bootstrap samples of hospital admissions from their dataset and constructing HHIs for each sample. They find that the correlation between hospitals' predicted HHIs across samples was above 0.99, suggesting that there is little need to account for sampling variation in the first stage. They argue that this result arises from the large number of observations used to construct predicted HHIs.

distance to that hospital and distance to the next closest alternative hospital, based on a number of hospital characteristics. The second reason for this variable's importance is that it is used to satisfy the exclusion restriction. That is, it is assumed that distance to hospital does not affect outcomes except via its effect on choice of hospital and therefore competition. Given that all the other predictors of patient choice included in the conditional logit model, such as patient characteristics, are included in the second stage regression of competition on outcomes, distance to hospital, and derivatives of this variable, are therefore the primary means by which regressions identify the causal effect of competition on clinical quality using HHIs constructed from predicted patient choices.

The exclusion restriction could be violated for two reasons. First, very sick patients could move house in order to live close to a hospital, in which case distance to hospital would predict health outcomes. While this phenomenon likely occurs to some limited extent, it is unlikely to bias the measures of competition intensity used in this paper, because patients who move house in order to live near a hospital (for example, patients with terminal cancer, or patients with kidney disease who require regular dialysis) are likely to be so sick that they are ineligible to undergo the elective surgical procedures covered by the PROMs programme.

Secondly, and more problematically, patient distance to hospital will be correlated with urbanness and therefore (potentially) with competition intensity, so whether the exclusion restriction is satisfied will depend on whether the correlates of urbanness which affect outcomes (such as poverty and health status) are satisfactorily controlled for by other means. In other words, HHIs based on predicted patient choices do not solve the problem of omitted variable bias due to the unobserved correlates of geography. This problem is the focus of the next section.

5.3 Value-added measures to address the correlates of geography

This paper identifies the causal effect of hospital competition on care quality using cross-sectional estimation, in which variation in competition intensity comes from the geographically-based nature of hospital markets – yet there are many correlates of geography that may also influence outcomes.[40] These correlates of geography will lead to omitted variable bias if not adequately controlled for. In England, inner-city residents tend to be poorer, and therefore also sicker, than their suburban and rural counterparts. If competition intensity is also higher in inner-city areas, the resulting correlation between competition intensity and health status will, if quality is equated with post-treatment health status, lead to downward-biased estimates of the effect of competition on quality. While this correlation between patient health status and competition intensity is the most important potential geographically-driven source of omitted variable bias, there are other potential sources of sources of geographically-driven heterogeneity at the hospital level that may bias one's estimates.

The 'best practice' solution to the problem of the unobserved correlates of geography has, since Kessler & McClellan (2000), been to include hospital fixed effects in one's model alongside a competition index based on predicted patient choices. The within-hospital year-on-year variation in predicted patient choice HHIs *can*, unlike equivalent variation from HHIs based on actual patient choices, theoretically be used to identify the causal effect of competition on quality in a model with hospital fixed effects.[41] However, my prior is that the four years covered by the PROMs data is unlikely to have provided sufficient time for any exogenous drivers of competition to have generated sufficient within-hospital variation in competition intensity to enable statistically significant treatment effects to be estimated. I report and discuss the estimates from such a regression in the robustness tests.

While including hospital fixed effects is not possible in this study, it turns out that the use of a value-added indicator of hospital performance (that is, equating hospital quality with *health gain from*

---

[40] The dataset used in this paper incorporates observations from four years, and the competition indices – as well as any other hospital-level averages – used are calculated at year level. This means that there is, strictly speaking, some within-hospital variation in competition intensity over time. However, as time-invariant instruments for competition are also employed, and hospital fixed effects are not included in the regressions, the effect of hospital competition on clinical quality is effectively being identified using only cross-sectional variation (notwithstanding that the fitted values from the first stage might vary slightly within a given hospital from year to year).

[41] Whereas any within-hospital year-on-year variation using HHIs calculated from actual patient choices will be an endogenous outcome of market participants' behaviour, an HHI calculated using predicted patient choices is based only on exogenous determinants of patient choice. Any within-hospital variation in an HHI based on predicted patient choices (which could be driven by exogenous factors such as demographic changes driven by migration, or changes in preferences concerning willingness to travel) should therefore, in theory, be able to identify a causal effect.

treatment rather than with *post-treatment health status*) provides an arguably even more powerful way of controlling for the most important correlate of geography that influences health outcomes, namely pre-treatment patient health status. The primary rationale for including hospital fixed effects in one's regressions is that there may be unobserved components of (pre-treatment) patient health status that affect outcomes. Hospital fixed affects are, in fact, not an ideal solution to this problem, as they do not control for pre-treatment patient health status at the level of the individual patient, but rather only control for any time-invariant features of average patient health status at a given hospital. PROMs, by providing information about pre-treatment health status as well as post-treatment health status, allows one to control directly for pre-treatment patient health status at the individual level, thus increasing precision relative to the inclusion of a hospital fixed effect, as well as avoiding any bias due to changing patient characteristics at the hospital level that can arise when hospital fixed effects are used.

The use of a value-added indicator of hospital performance eliminates the problem of classical casemix bias, in which hospitals whose patients start out unobservably sicker have worse average health outcomes, and therefore appear to offer lower quality care than is the case in reality. This problem is replaced with another, more subtle source of potential bias if unobserved components of pre-treatment health status are correlated with health *gain* from surgery. For example, if unobservably sicker patients have a higher average health gain from surgery, then hospitals with unobservably sicker patients will appear to offer *higher* quality care than is the case in reality.

Observably sicker patients will certainly have higher health gains from surgery as measured by PROMs due to ceiling effects arising from the bounded nature of PROMs scores – a patient who reports close to perfect health before treatment is incapable of experiencing large health gains from treatment as captured by the PROMs surveys. However, by risk-adjusting health gains from surgery, and including pre-treatment health status (and its quadratic) as explanatory variables in the risk adjustment regression, this paper strips out this mechanical influence of pre-treatment health status on outcomes.[42] The effect of this risk-adjustment exercise is demonstrated in Figure 3, which shows, for hip replacement patients, the relationship between pre-treatment health status and adjusted/unadjusted health gains. Panels (A) and (C) show how unadjusted health gains from surgery are a mechanical function of pre-treatment health status, while panels (B) and (D) show how the risk adjustment process eliminates this relationship.[43] The risk-adjustment of outcomes means that this paper is effectively estimating a kind of 'lagged-score' value-added model rather than a 'gain-score' model, because it allows for the possibility that pre-treatment health status influences post-treatment health status with a coefficient different to one.

This paper's estimates will only be subject to casemix bias from the unobserved correlates of geography if *unobserved* components of pre-treatment health status are correlated with health gains from surgery. In other words, post-treatment health status must be influenced by *unobserved* components of pre-treatment health status. But pre-treatment health status is captured using a questionnaire identical to that used to capture post-treatment health status. I therefore argue that the pre-treatment PROMs survey is a sufficient statistic for aspects of pre-treatment health status relevant to post-treatment health status as captured by PROMs. Therefore, estimating the effect of market structure on hospital performance using risk-adjusted PROMs health gains eliminates the problem of casemix bias due to the unobserved correlates of geography.

---

[42] When publishing hospital-level PROMs scores, NHS Digital (and its predecessors, HSCIC and the NHS IC) provides casemix-adjusted health gains from surgery, in addition to unadjusted outcomes. NHS Digital does not, however, perform the same adjustment on patient level data. I therefore replicate the NHS's hospital-level casemix adjustment strategy, to generate patient-level risk-adjusted post-treatment health status, and risk-adjusted health gain from surgery, for all survey respondents who could be linked to HES. Appendix 1 outlines this casemix adjustment methodology, discusses associated methodological questions, provides a list of the variables used, and reports coefficient estimates from the risk adjustment regressions.

[43] It is a deliberate design feature of the official NHS PROMs risk adjustment methodology that adjusted health gains from surgery are not completely orthogonal to pre-treatment health status. If adjusted health gains were simply calculated as a residual after stripping out the effect of observables (including pre-treatment health status) then adjusted health gains would be orthogonal to pre-treatment health status by construction. However, the NHS PROMs risk adjustment methodology includes in the adjustment regression a hospital fixed effect (which can be thought of as capturing hospital quality), and strips out only the effect of observable patient characteristics on health gains *over and above* the component of health gains that can be attributed to the hospital via the fixed effect. Therefore, adjusted health gains will continue to be correlated with pre-treatment health status to the extent that pre-treatment health status is correlated with the hospital fixed effects. This is a strength not a weakness of the risk adjustment methodology, as it allows for the possibility that hospital quality, as captured by the hospital fixed effects, is empirically correlated with pre-treatment health status.

While unobserved aspects of patient health status are the most significant potential source of omitted variable bias due to the geographical correlates of competition, there are other unobserved hospital-level characteristics that are correlated with competition intensity and influence outcomes. Inner-city hospitals tend to be older and more prestigious than other hospitals – for example because they are teaching hospitals, or connected to a university (and therefore involved in medical research). If these hospitals attract better doctors, and if competition intensity is higher in inner-city areas, the resulting correlation between competition intensity and doctor quality may, if not adequately controlled for, lead to biased estimates. Relatedly, it has been argued that doctors who elect to live in London may be systematically different from those who do not. In addition, care quality tends to be higher, and costs lower, in larger health care markets, and in larger hospitals. These economies of scale, or volume effects, can lead to biased estimates if areas with larger markets and hospitals also have higher competition intensity.

While all the sources of hospital-level heterogeneity affecting outcomes outlined in the previous paragraph are a threat to identification of a causal effect, I argue that they can be adequately controlled for using observables. I address concerns that teaching and university hospitals may attract better doctors by including dummy variables for these hospital types in all regressions.[44] I address concerns about systematic differences between London and elsewhere by running a robustness test that excludes all London hospitals from the regressions. And economies of scale at the hospital and market level are addressed by including controls for the site's total annual case load for that procedure and its quadratic, the trust's total annual number of admissions for all causes and its quadratic, and catchment area population density (people per hectare). As total and procedure-specific admissions may be influenced by hospital quality in the period after the introduction of patient choice, in the main specification lagged values of these variables and their quadratics (their average and total values respectively over the three years from 2002/3 to 2004/5) are used in place of current-period values.[45]

While other sources of omitted variable bias due to unobserved correlates of geography cannot be completely ruled out, these controls do a good job of addressing the most worrisome non-casemix sources of hospital-level heterogeneity that influence outcomes. Therefore, I argue that, in conjunction with a value-added indicator of outcomes to control for casemix bias at the individual level, they enable the identification of the causal effect of hospital competition on elective surgery quality.

5.4 Additional competition indices

This paper centres hospital markets on patients' neighbourhoods of residence, rather than on hospitals themselves, in order to ameliorate concerns about the possible influence of hospital quality on (its measures of) competition intensity. An alternative method of addressing this concern, employed by Cooper et al. (2011), is to centre hospital markets on GP surgeries. In the UK, patients are required, in the vast majority of cases, to register with a GP that is close to their home. The address of the GP surgery thus provides a good proxy for the patient's home address. As a check on the results obtained using neighbourhood-centred markets, yearly GP-centred HHIs are also calculated. For each elective surgery observation, the straight line distance from GP surgery to hospital is calculated, and the GP surgery's market for each year and surgical procedure is defined as the GP-centred circle that encompasses 95 per cent of the treatment locations of the GP surgery's patients. Any hospital that lies within this circle is considered to be in the GP surgery's market, irrespective of whether the GP surgery refers any patients to the hospital. That is, each GP's market includes all patients that attend hospitals within the GP surgery's 95 per cent market radius, irrespective of the GP surgery at which they are registered. An HHI is calculated for each GP-procedure-year combination, and an overall GP-year HHI is then calculated as a weighted average of procedure-level HHIs. Finally, a hospital-level HHI is calculated as the weighted sum of the GP-level HHIs of its patients' GP surgeries.[46]

I also calculate a competition index similar to that used by Propper et al. (2004) and Bloom et al. (2015), in which intensity of competition experienced by hospital j is equal to the number of hospitals

---

[44] In addition, to address the concern that Foundation Trust hospitals (which have greater autonomy than standard NHS hospitals) may attract higher-quality staff, a robustness test includes a dummy variable for Foundation Trust status. This dummy variable is not included in the main regressions because acquisition of Foundation Trust status was endogenous to hospital quality during this period.

[45] Likewise, catchment area population density is calculated as the average population density (from the 2011 census) of the MSOAs of residence of the hospital's patients between 2002/3 and 2004/5, instead of using the MSOAs of residence of the hospital's current patients.

[46] This final stage of aggregation, which was not undertaken by Cooper et al. (2011), ensures that treatment intensity is the same for every patient attending a given hospital in a given year.

within 30km of j.[47] Table 9 reports summary statistics for all the competition indices examined in this paper. Table 10 reports the correlation between these competition indices, a dummy variable indicating that the patient lives in an urban area, and population density in the hospital's catchment area.

5.5 Data sources and control variables

This paper is based on two NHS datasets – the Hospital Episode Statistics (HES), which contains a record for every hospital visit by an NHS patient in England, and the PROMs survey responses by individual patients.[48] The HES dataset used in this study encompasses eleven financial years, from 2002/3 to 2012/13. In addition to containing all elective admissions for the four PROMs procedures, the dataset includes all admissions for two additional elective procedures, knee arthroscopy and cataract repair, which are used to construct the competition measures.[49] Finally, the dataset includes all non-elective AMI admissions. In total there are 8.6 million observations.

The PROMs dataset for 2009/10 to 2012/13 contains 525,538 non-duplicate survey responses where the epikey field (which allows the record to be linked to HES) is present. NHS Digital (2016) indicates that this constitutes approximately 60 per cent of all surgical procedures eligible for inclusion in the PROMs programme. Of these, 505,396 – or 96.2 per cent – were successfully matched to the HES dataset. After excluding observations with 50 or fewer cases for the procedure and financial year in question, as well as applying other minor exclusion criteria, there are 495,347 observations available for the regressions. Regression sample sizes are smaller than this number due to unavailability of the post-operative survey, or (in a smaller number of cases) missing data on hospital competition intensity or volumes.

Patient-level controls include dummies for gender crossed with age (in five-year intervals), and dummies for day and month of operation (to control for day-of-week and seasonality effects). They also include dummies indicating a day case, 'low' and 'high' severity (respectively, one diagnosis, and three or more diagnoses), and residence in an urban area, as well as the patient's Charlson score, which indicates the patient's 10-year survival probability based on their health status in relation to 17 conditions likely to lead to death. Finally, a control is included for the Index of Multiple Deprivations income deprivation score (Noble et al. 2004), which measures the percentage of households in the patient's Lower Super Output Area (LSOA) of residence that are income deprived.[50]

Dummies are also included to indicate whether the hospital (site) is part of a specialist trust, a teaching trust, a university trust, or a private provider (a dummy variable denoting a standard acute trust is omitted).[51] As discussed in Section 5.3, scale effects are addressed via controls for pre-reform average number of procedures undertaken at the hospital site and its quadratic, pre-reform average total admissions for the hospital trust and its quadratic, and catchment area population density. Finally, year-specific dummies are included to indicate the region of England in which the hospital site is located, to account for changing health policies at the Strategic Health Authority level, as well as any other region-level trends. Table 11 provides average values for key control variables used in this paper.

---

[47] Simple 'fixed radius' measures of this kind are subject to bias due to differences in travel time between rural and urban areas – they overestimate competition intensity in urban areas, and underestimate it in rural areas. Nonetheless, they provide an interesting alternative perspective to the other competition indices used in this paper.

[48] NHS patients can be treated either in NHS (public) hospitals, or in private hospitals that are registered to accept NHS patients. A dummy variable is included in all regressions to denote treatment in a private hospital. HES does not include private (e.g. privately insured) patients treated at private hospitals. However, private patients comprise only a small percentage (less than 10 per cent) of the total hospital market in England.

[49] The elective surgical procedures used to construct the competition indices in this paper follow Cooper et al. (2011). Varicose vein stripping is added to the five elective surgical procedures used in that paper. This ensures that all the PROMs procedures are used as inputs to the competition indices used here. For the four PROMs procedures, a broader definition of each procedure is used than that employed by the PROMs programme, in order to define the 'market' for each procedure in a meaningful and intuitive way for the purpose of constructing the competition indices; see Appendix 3 for procedure and diagnosis codes.

[50] As poverty is associated with poor health, the IMD income deprivation score may control for unobserved dimensions of health status that influence outcomes.

[51] To make the hospital dummy variables exclusive, specialist university trusts are categorised as specialist, and teaching university trusts are categorised as teaching. A very small number of community care trusts enter our dataset for groin hernia repair and varicose vein stripping; these are grouped with standard acute trusts. A dummy denoting whether a hospital is a Foundation Trust is not included, for reasons discussed in Footnote 44.

6 Results

6.1 Main estimates

Table 12 reports cross-sectional estimates of the impact of competition on elective surgery quality as captured by casemix-adjusted PROMs health gains. Competition is measured as a neighbourhood-centred, hospital-level HHI based on actual patient choices. Table 12 indicates that competition has a negative effect on orthopaedic surgery quality as captured by the Oxford Hip/Knee Score, and varicose vein surgery quality as captured by the Aberdeen Questionnaire, both significant at the 5 per cent level. There are no statistically significant treatment effects using the EQ-5D.

Table 13 reports the first stage estimates from this paper's headline specification, which includes two forms of instrumentation – an HHI constructed using patient choices predicted on the basis of exogenous variables, and instrumentation of current-period HHIs by their average pre-reform (2002/3 to 2004/5) value. Table 13 shows that the coefficient on the excluded instrument, pre-reform average HHI, is between 0.940 and 0.964, and highly statistically significant, in all specifications, indicating that hospital-level HHIs based on predicted patient choices exhibit very little year-on-year variation during the study period.

Table 14 reports the second stage estimates. It indicates a negative effect of competition on the quality of orthopaedic surgery as captured by the EQ-5D (10 per cent level) and the Oxford Hip/Knee Score (5 per cent level), and varicose vein stripping surgery as captured by the Aberdeen Questionnaire (5 per cent level). Pooling all four PROMs procedures together using the EQ-5D, competition has a negative effect on elective surgery quality significant at the 10 per cent level.

Patients from urban areas have lower health gains from surgery. Although the coefficients on variables capturing patient severity should be interpreted with caution given that outcomes are already casemix-adjusted, they indicate that sicker patients (those with a higher Charlson score, or with three or more diagnoses) have lower health gains from surgery. Interestingly, the procedure-specific controls for hospital volume are not statistically significant, while for all outcomes except the Aberdeen Questionnaire, the Trust-wide controls for hospital volume are statistically significant. In all cases where a volume control is statistically significant, the linear term is negative while the quadratic term is positive, indicating that, for sufficiently large volumes, higher volumes imply larger health gains per patient.

It is possible to make use of the cardinality of the EQ-5D to provide a sense of the magnitudes of these treatment effects. A one standard deviation increase in competition intensity (0.6777) leads to a decrease in health gain from orthopaedic surgery of 0.005 ($= -0.00756 * 0.6777$). An average orthopaedic surgery patient experiences an increase in health status from 0.358 before surgery to 0.735 after surgery. If their hospital experienced a one standard deviation increase in competition intensity, their post-surgical health status would be only 0.730. Thus, they would be indifferent between one year in their post-operative heath state and 0.730 years in perfect health, as opposed to 0.735 years before the increase in competition. These negative impacts of competition on health gains from elective surgery are small, but nonetheless distinguishable from zero.

6.2 Robustness tests

Table 15 presents results when the paper's headline regressions for orthopaedic surgery are disaggregated into separate regressions for hip replacement and knee replacement. They show that the results differ little between the two surgical procedures, but are marginally more statistically significant for knee replacement surgery when the EQ-5D is used. Table 16 reports estimates using the PROMs with smaller effect sizes – the EQ-5D for varicose vein stripping and groin hernia repair, and the EQ-VAS for all PROMs procedures. As expected, these estimates are less statistically significant than those reported in Tables 12 and 14. Nonetheless, it is noteworthy that all but one of the reported coefficients indicate a negative effect of competition on quality.

Table 17 reports robustness tests using alternative measures of competition intensity. Row (1) uses an HHI based on actual patient choices rather than predicted patient choices. The results remain significant for orthopaedic surgery, but not for the other outcomes. Row (2) reports estimates using a GP-centred, hospital-level HHI defined using the 95th percentile of distance from GP to hospital. Using this index of competition, the estimates for orthopaedic surgery using the Oxford Hip/Knee Score remain statistically significant. Finally, Row (3) reports estimates using a simple fixed distance measure, in which competition is defined as the number of competitors within 30 kilometres. In this specification, the estimates for varicose vein stripping using the Aberdeen Questionnaire remain statistically significant. Overall, these robustness tests using alternative competition indices, while not always as statistically significant as our headline estimates, strengthen the overall impression that hospital competition during this era had a negative effect on elective surgery quality, and on orthopaedic surgery quality in particular.

Table 18 reports estimates using alternative specifications but the same outcome variables and competition indices as in Table 14, with an HHI constructed using predicted patient choices and current-period competition intensity instrumented by its average pre-reform value. Studies of hospital competition in England are often criticised on the grounds that they simply pick up differences between London and the rest of the country. Row (1) reports estimates when all London hospitals are excluded. The results are largely unchanged. Row (2) reports estimates when a dummy variable indicating a Foundation Trust is included, while Row (3) reports estimates when the sample is restricted to public (NHS) hospitals. The results in both cases are virtually identical to those reported in Table 14. It might be worried that interacting year dummies and region of England dummies soaks up too much variation, and thus reduces the statistical significance of the estimates reported in Table 14. Row (4) reports estimates when year dummies and region of England dummies are included as separate controls, instead of interacted with each other. Contrary to the worry that motivated this robustness test, the results are reduced in magnitude and statistical significance.

Row (5) reports estimates when the control for catchment area population density and the dummy variable for living in an urban area are not included, while Rows (6) and (7) report estimates when only one of these variables is omitted. The results are much more statistically significant in all cases. This is a particularly important finding, as controls for catchment area population density are rarely included when estimating the causal effect of competition on quality. Row (8) reports estimates when catchment area population density is calculated based on the home MSOAs of current patients, as opposed to patients in the three years before the introduction of patient choice (2002/3-2004/5). The results are very similar to those reported in Table 14.

To ameliorate concerns about endogeneity of the hospital scale controls (total and procedure-specific admissions, and their respective quadratics), the main specification uses the pre-reform average values of these variables in place of their current-period values. Row (9) reports estimates using an alternative approach, namely instrumenting current-period values of these variables with their pre-reform averages. The results are very similar to those reported in Table 14, albeit with a small loss of statistical significance in some cases. Row (10) reports estimates when all scale effects controls (total and procedure-level admissions, and their respective quadratics) are omitted. Only the estimates using the Aberdeen Varicose Vein Questionnaire remain statistically significant. The findings in Row (10) do not raise questions about the robustness of our main results, but merely indicate that hospital scale effects are an important driver of outcomes, and need to be controlled for.[52]

Row (11) reports estimates when all non-binary variables are logged. The results are qualitatively identical to those reported in Table 14. Row (12) reports estimates when hospital fixed effects are included in the regression. Consistent with my initial hypothesis that there is insufficient exogenous within-hospital variation in competition intensity to enable the identification of treatment effects when hospital fixed effects are included, the estimates reported in Row (12) are not statistically significant.

Finally, Table 19 reports estimates when a linear time trend interacted with competition is included in addition to competition on its own. With the possible exception of the Aberdeen Varicose Vein Questionnaire, the results do not support the hypothesis that the effect of competition on elective surgery quality exhibited a trend over the study period. Rather, the negative effect of competition on elective surgery quality does not seem to vary over time.

---

[52] Omitting all hospital scale effects controls increases sample size by 18,000, as some volume controls are missing for some observations. I have confirmed that the loss of statistical significance reported in Row (10) of Table 18 is driven by the omission of hospital scale effects, not by the inclusion of these additional observations.

7 Discussion and conclusions

7.1 Discussion

Overall, the evidence seems to indicate that the hospital competition brought about by the 2006 English NHS patient choice reforms had a negative impact on quality of orthopaedic surgery and varicose vein stripping surgery significant at the 5 per cent level. Pooling all PROMs procedures together using the EQ-5D, competition had a negative effect on elective surgery quality significant at the 10 per cent level. The estimates for orthopaedic surgery using the Oxford Hip/Knee Score are particularly robust to alternative specifications. However, even when the coefficients on competition using other outcome measures and surgical procedures are not statistically significant in the robustness tests, they are almost always negative in sign.

Two questions naturally arise from these findings. Why does competition appear to have had a negative effect on orthopaedic and varicose vein surgery quality? And, how should these findings be understood in relation to the existing literature (Cooper et al. 2011; Gaynor et al. 2013; Bloom et al. 2015), which indicates that competition in the English NHS during this era led to lower mortality rates?

Section Four put forward a model that could simultaneously explain both these findings from the literature, and this paper's findings. While information about hospitals' performance in relation to PROMs health gains has been available to researchers since 2010, little attempt was made, during the period of study, to communicate these data to patients undergoing a PROMs procedure when choosing a hospital. On the other hand, the NHS Choices website, which helps such patients to choose a hospital by providing information about the performance of nearby hospitals, *did* report overall hospital mortality rates to patients undergoing a PROMs procedure, even though this statistic is of little direct relevance to the elective surgical procedures covered by PROMs. It may be that hospitals, knowing this situation, focused on improving their performance in relation to publicly reported dimensions of performance, such as mortality rates, at the expense of other areas of activity in which quality is not reported to patients.

A more general hypothesis is that the hospital competition engendered by the patient choice reforms had a positive effect on hospital performance via a more diffuse mechanism than that considered by formal economic models. That is, perhaps competition leads to improved behaviour not via actual exertion of patient choice, leading to changes in market shares and hence behaviour, but instead by making hospital managers in more competitive markets feel, in a more general sense, that their performance is under scrutiny, and that they therefore need to lift their game in relation to observable and high-profile performance indicators such as mortality. Such a mechanism could explain why competition for elective surgery patients appears to have led to lower mortality rates in high competition areas, but not to improved elective surgery quality as captured by less prominent performance indicators.

Alternatively, a similar more diffuse effect could operate via patient choice – perhaps a hospital's mortality rates affect its overall reputation for quality, and perhaps elective surgery patients choose which hospital to attend on the basis of this general reputation, rather than on the basis of knowledge about hospital quality in the specific surgical specialty that encompasses their procedure. If patients choose in such a manner, it would make perfect sense for a hospital to focus on reducing mortality instead of on – and perhaps at the expense of – improving elective surgery quality.

7.2 Conclusions

Previous studies of hospital competition and quality have tended to focus on mortality-based indicators of hospital performance, yet in the spheres of hospital activity where competition for patients does occur, such as elective surgery, mortality is a relatively uncommon outcome. The approach of these previous studies has been informed by a theoretical framework that explicitly or implicitly assumes that quality is a hospital-wide variable, or at least has a significant hospital-wide component.

This paper, by contrast, has conceptualised the hospital as a multi-product firm, in which managers make separate quality choices in different areas of activity. I have shown that there is little evidence that a hospital's emergency care quality (as captured by mortality rates) is correlated with its elective surgery quality (as captured by PROMs health gains), and put forward a theoretical model in which the effect of hospital competition on clinical quality is differentiated by output type, and influenced by complementarities and substitutabilities between output types in production. I test this theoretical model using new performance indicators – Patient Reported Outcome Measures of health gain from four high-volume elective surgical procedures – that, by equating outcomes with value added (health gain) from treatment rather than post-treatment health status, enables hospital performance to be measured at the level of the individual surgical procedure to a far greater extent than existing failure-based indicators of hospital performance such as mortality or readmissions. As well as enabling a more granular approach to health care provider performance measurement, the use of value-added outcome measures also represents a fundamental advance in the challenge of disentangling the contribution of provider performance and patient characteristics to health outcomes.

In contrast to the existing literature, this paper finds that, when value-added, elective-surgery-specific outcome measures are used, it appears that the introduction of patient choice of hospital for elective surgery to the English NHS during the 2000s may have had a negative effect on clinical quality in at least some areas. This finding, though provisional, calls into question a common interpretation of previous econometric studies of the English NHS patient choice reforms (Cooper et al. 2011; Gaynor et al. 2013; Bloom et al. 2015) – namely that they show that hospital competition led to across-the-board improvements in care quality. This paper's finding that hospital mortality rates and PROMs health gains are essentially uncorrelated suggests that, in fact, the very notion of a 'hospital-wide' indicator of clinical quality is questionable.

On the other hand, this paper's findings do *not* support one commonly heard criticism of the econometric literature on the English NHS – namely that hospital behaviour is simply unresponsive to the economic environment. In fact, this paper's findings suggest that hospitals are, if anything, *more* responsive to the incentives environment within which they operate than has been suggested by the existing literature, in the sense that hospital responses to the competitive environment may be differentiated by area of activity, depending on the observability of performance in each area, the extent to which performance is incentivised in each area, and the interactions between incentivised and unincentivised areas in the production process. These findings are consistent with the literature on public organisation (see e.g. Dixit 2002; Propper & Wilson 2003; Holmstrom & Milgrom 1991), which predicts that standard economic incentive structures drawn from the private sector will have more complex effects – including, potentially, perverse effects – when applied to the public sector, given the existence, within the latter, of multiple dimensions of performance, many of which are not fully measured or observable.

This paper's findings, although tentative, run counter to the conventional wisdom that fixed-price hospital competition should lead to higher clinical quality. This conventional wisdom forms the basis for much health policy – for instance, it has influenced market-based reforms to government-run health care systems, and informs the thinking of anti-trust regulators responsible for approving mergers of health care providers. The fact that this paper's findings deviate substantially from those of the existing literature suggests the need for follow-up work seeking to replicate these findings in different settings, and with different identification settings, but using similar value-added, elective-surgery-specific outcome measures.

It would be particularly valuable to seek to replicate this paper's findings in a setting where it is possible to identify the effect of competition on elective surgery quality in a model that includes hospital fixed effects, the traditional method of controlling for the geographical correlates of competition. I argue that the use of a (casemix-adjusted) value-added outcome measure is a convincing alternative method of controlling for the most worrying geographical correlate of competition, namely patient health status, and that other potential geographically-driven sources of bias can be controlled for using observables. Nonetheless, uncertainties remain in relation to both arguments, and employing hospital fixed effects alongside a value-added outcome measure would be one way of resolving them. Whatever these follow-up studies find, the present paper's findings highlight the limitations of single-output-type models of the hospital, the value of conceptualising the hospital as a multi-product firm, and the importance of assessing hospital performance using indicators that are differentiated by output type, and based on the value added from hospital treatment.

In a paper showing that hospital competition in the English NHS in the 1990s led to higher mortality rates, Propper et al. (2004, p.1267) conclude by discussing the fact that hospital mortality rates were not publicly available during the 1990s, noting drily that "it may have been a mistake to delay the publication of quality signals until some 10 years after the introduction of a market meant to rely on them". The present study's findings suggest that a similar point might apply to the post-2006 era of hospital competition driven by patient choice of hospital for elective surgery. To be sure, as compared with the 1990s Internal Market, vastly more data about hospital performance was available from 2006 onwards to both patients and researchers. Nonetheless, there were few or no elective-surgery-specific indicators of clinical quality provided to patients. The NHS PROMs programme, which was developed in response to a recognition of this deficiency, commenced in 2009/10, several years after the inauguration of a new era of provider competition premised on the existence of meaningful signals of elective surgery quality.

Even after the PROMs programme commenced, the resulting data were published as "experimental statistics" on the back alley of an NHS website of data sources. Given this situation, it is perhaps not surprising that NHS hospitals may have responded to this new era of competition by prioritising performance improvements in salient and high-profile areas such as mortality, in spite of the fact that few hospital deaths occur in parts of the hospital subject to competitive forces, and furthermore that these performance improvements may even have come at the expense of performance in lower-profile areas, including those directly influencing health outcomes in elective surgery.

References

Adams, Richard and Helena Bengtsson (2017), ‘Grammar schools lose top spots after league table shakeup’, *The Guardian*, 19 January, available from https://www.theguardian.com/education/2017/jan/19/grammar-schools-lose-top-spots-after-league-table-shakeup, accessed 5 February 2017.

Aylin, P, R Alexandrescu, M H Jen, E K Mayer, and A Bottle (2013), ‘Day of week of procedure and 30 day mortality for elective surgery: retrospective analysis of hospital episode statistics’, *British Medical Journal* 346:f2424.

Basu, Anirban and Andrea Manca (2012), ‘Regression Estimators for Generic Health-Related Quality of Life and Quality-Adjusted Life Years’, *Medical Decision Making* 32(1), Jan/Feb, pp.56–69.

Bevan, Gwyn and Jocelyn Cornwell (2006), ‘Structure and logic of regulation and governance of quality of health care: was OFSTED a model for the Commission for Health Improvement?’, *Health economics, policy and law* 1(4), October, pp.343–70.

Bevan, Gwyn and Matthew Skellern (2011), ‘Does competition between hospitals improve clinical quality? A review of evidence from two eras of competition in the English NHS’, *British Medical Journal* 343:d6470.

Black, Nick (2010), ‘Assessing the quality of hospitals’, *British Medical Journal* 340:c2066.

Bloom, Nicholas, Zack Cooper, Martin Gaynor, Stephen Gibbons, Simon Jones, Alistair McGuire, Rodrigo Moreno-Serra, Carol Propper, John Van Reenen and Stephan Seiler (2011), ‘A response to Pollock et al. (2011)’, available from http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/cpresponse.pdf accessed 8 September 2016.

Bloom, Nicholas, Carol Propper, Stephan Seiler and John Van Reenen (2015), ‘The Impact of Competition on Management Quality: Evidence from Public Hospitals’, *Review of Economic Studies* 82(2), pp.457–89.

Bradley, Elizabeth H, Jeph Herrin, Brian Elbel, Robert L McNamara, David J Magid, Brahmajee Nallamothu, Yongfei Wang, Sharon-Lise T Normand, John A Spertus and Harlan M Krumholz (2006), ‘Hospital Quality for Acute Myocardial Infarction: Correlation Among Process Measures and Relationship With Short Term Mortality’, *Journal of the American Medical Association* 296(1), 5 July, pp.72–8.

Browne, John, Liz Jamieson, James Lewsey, Jan van der Meulen, Nick Black, John Cairns, Donna Lamping, Sarah Smith, Lynn Copley, and Jackie Horrocks (2007), *Patient Reported Outcome Measures (PROMs) in Elective Surgery*, Report to the Department of Health, 12 December, available from http://www.lshtm.ac.uk/php/hsrp/research/proms_report_12_dec_07.pdf, accessed 16 August 2013.

Chassin, Mark R (2002), ‘Achieving And Sustaining Improved Quality: Lessons From New York State And Cardiac Surgery’, *Health Affairs* 21(4), pp.40–51.

Chetty, Raj, John N Friedman and Jonah E Rockoff (2015), ‘Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates’, American Economic Review 104(9), September, pp.2593–632.

CMS [Centers for Medicare & Medicaid Services (2016), *CMS Quality Measure Development Plan: Supporting the transition to the Merit-based Incentive Payment System (MIPS) and Alternative Payment Models (APMs)*, 2 May, Centers for Medicare & Medicaid Services, Baltimore MD, available from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/Final-MDP.pdf, accessed 5 February 2017.

Coles, James (2010), ‘PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures’, Report by Northgate Public Services, September, available from http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf, accessed 16 August 2013.

Coles, James (2012), ‘PROMs Risk Adjustment: a case-mix approach’, presentation to *Patient reported outcome measures (PROMs) research conference*, 22 November, King’s Fund, London, available from http://www.kingsfund.org.uk/sites/files/kf/proms-risk-adjustment-a-case-mix-approach-nov12.pdf, accessed 16 August 2013.

Colla, Carrie, Julie Bynum, Andrea Austin and Jonathan Skinner (2016), ‘Hospital Competition, Quality, and Expenditures in the U.S. Medicare Population’, NBER Working Paper 22826, National Bureau of Economic Research, Cambridge MA, available from http://www.nber.org/papers/w22826, accessed 26 November 2016.

Cooper, Zack, Stephen Gibbons, Simon Jones and Alistair McGuire (2011), ‘Does Hospital Competition Save Lives? Evidence From the English NHS Patient Choice Reforms’, *Economic Journal* 121, pp.F228–60.

Cooper, Zack, Stephen Gibbons and Matthew Skellern (2016), ‘Does Competition from Private Surgical Centres Improve Public Hospitals’ Performance? Evidence from the English National Health

Service', CEP Discussion Paper 1434, Centre for Economic Performance, London School of Economics.

Devlin, Nancy J and John Appleby (2010), *Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision-making*, The King's Fund, London, available from https://www.kingsfund.org.uk/sites/files/kf/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf, accessed 16 August 2013.

DH [Department of Health] (2012a), 'Patient Reported Outcome Measures (PROMs) in England: The case-mix adjustment methodology', 11 April, Department of Health, London, available from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216507/dh_133449.pdf, accessed 16 August 2013.

DH [Department of Health] (2012b), 'Patient Reported Outcome Measures (PROMs) in England: A Methodology for applying casemix adjustment Annex A: Coefficients for Groin Hernia Models', 11 April, Department of Health, London, available from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216508/dh_133450.pdf, accessed 16 August 2013.

DH [Department of Health] (2012c), 'Patient Reported Outcome Measures (PROMs) in England: A Methodology for applying casemix adjustment Annex B: Coefficients for Varicose Vein Models', 11 April, Department of Health, London, available from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216509/dh_133581.pdf, accessed 16 August 2013.

DH [Department of Health] (2012d), 'Patient Reported Outcome Measures (PROMs) in England: A Methodology for applying casemix adjustment Annex C: Coefficients for Hip Replacement Models', 11 April, Department of Health, London, available from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216510/dh_133452.pdf, accessed 16 August 2013.

DH [Department of Health] (2012e), 'Patient Reported Outcome Measures (PROMs) in England: A Methodology for applying casemix adjustment Annex D: Coefficients for Knee Replacement Models', 11 April, Department of Health, London, available from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216511/dh_133453.pdf, accessed 16 August 2013.

Dixit, Avinash (2002), 'Incentives and Organizations in the Public Sector: An Interpretative Review', *The Journal of Human Resources* 37(4), Autumn, pp.696–727.

Dixon, Anna, Ruth Robertson, John Appleby, Peter Burge, Nancy Devlin and Helen Magee (2010), *Patient Choice: How patients choose and how providers respond*, The King's Fund, London, available from http://www.kingsfund.org.uk/sites/files/kf/Patient-choice-final-report-Kings-Fund-Anna%20Dixon-Ruth-Robertson-John-Appleby-Peter-Purge-Nancy-Devlin-Helen-Magee-June-2010.pdf, accessed 14 May 2014.

Dolan, Paul (1997), 'Modelling Valuations for EuroQol Health States', *Medical Care* 35(11), pp.1095–108.

Dr Foster [Dr Foster Research Limited] (2011), 'Dr Foster Report Card', available from http://www.drfosterhealth.co.uk/quality-reports, accessed 28 September 2011.

EuroQol Group, The (1990), 'EuroQol - a new facility for the measurement of health-related quality of life', *Health Policy* 16(3), December, pp.199–208.

Feng, Yan, Michele Pistollato, Anita Charlesworth, Nancy Devlin, Carol Propper and Jon Sussex (2015), 'Association between market concentration of hospitals and patient health gain following hip replacement surgery', *Journal of Health Services Research & Policy* 20(1), January, pp.11–7.

Freeman, Michael, Nicos Savva and Stefan Scholtes (2016), 'Economies of Scale and Scope in Hospitals: Evidence of Productivity Spillovers Across Hospital Services', mimeo, London Business School, September, available from http://faculty.london.edu/nsavva/ScaleAndScope.pdf, accessed 19 May 2017.

Fotaki, Marianna, Martin Roland, Alan Boyd, Ruth McDonald, Rod Scheaff and Liz Smith (2008), 'What benefits will choice bring to patients? Literature review and assessment of implications', *Journal of Health Services Research & Policy* 13(3), pp.178–84.

Gaynor, Martin, Rodrigo Moreno-Serra and Carol Propper (2011), 'Death by Market Power: Reform, Competition and Patient Outcomes in the National Health Service', CMPO Working Paper No. 10/242, July 2010 (updated August 2011), Centre for Market and Public Organisation, University of Bristol, available from http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/wp242.pdf, accessed 25 August 2016.

Gaynor, Martin, Rodrigo Moreno-Serra, and Carol Propper (2013), 'Death by Market Power: Reform, Competition and Patient Outcomes in the National Health Service', *American Economic Journal: Economic Policy* 5(4), pp.134–66.

Gaynor, Martin and Robert J Town (2012), 'Chapter 9: Competition in Health Care Markets', *Handbook of Health Economics* Vol. 2, pp.499–637.

Gaynor, Martin (2006), 'What Do We Know About Competition and Quality in Health Care Markets?', 6 December, Carnegie Mellon University, mimeo.

Gibbons, Stephen, Olmo Silva and Felix Weinhardt (2013b), 'Everybody needs good neighbours? Evidence from student outcomes in England', *Economic Journal* 123(571), September, pp.831–74.

Gibbons, Stephen, Stephen Machin and Olmo Silva (2013a), 'Valuing school quality using boundary discontinuity regressions', *Journal of Urban Economics* 75, May, pp.15–28.

Grossman, Michael (1972), 'On the Concept of Health Capital and the Demand for Health', *Journal of Political Economy* 80(2), March-April, pp.223–55.

Gutacker, Nils, Chris Bojke, Silvio Daidone, Nancy Devlin and Andrew Street (2013), 'Hospital Variation in Patient-Reported Outcomes at the Level of EQ-5D Dimensions: Evidence from England', *Medical Decision Making* 33(6), August, pp.804–18.

HealthIT.gov (2016), 'Health information technology use in the United States and the United Kingdom', available from https://www.healthit.gov/policy-researchers-implementers/health-information-technology-use-united-states-and-united-kingdom, accessed 5 February 2017.

Herdman, M, C Gudex, A Lloyd, MF Janssen, P Kind, D Parkin, G Bonsel and X Badia (2011), 'Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L)', *Quality of Life Research* 20(10), pp.1727–36.

Hernández Alava, Mónica, Allan J Wailoo, Roberta Ara (2012), 'Tails from the Peak District: Adjusted Censored Mixture Models of EQ-5D Health State Utility Values', *Value in Health* 15(3), May, pp.550–61.

HHS, CMS and Isis Innovation (2016), 'Memorandum of Understanding Between Office of the National Coordinator for Health Information Technology and Centres for Medicare & Medicaid Services and ISIS Innovation Ltd.', available from https://www.healthit.gov/sites/default/files/mou_hhs_i-innovations_final_may_2016.pdf, accessed 5 February 2017.

HSCIC [Health and Social Care Information Centre] (2013), 'Indicator Specification: Summary Hospital-level Mortality Indicator', Indicator Reference Number I00699, Version 1.14, 11 April, Health and Social Care Information Centre, available from http://www.hscic.gov.uk/CHttpHandler.ashx?id=11151&p=0, accessed 1 October 2013.

Jha, Ashish K, E John Orav, Zhonghe Li and Arnold M Epstein (2007), 'The Inverse Relationship Between Mortality Rates and Performance In The Hospital Quality Alliance Measures', *Health Affairs* 26(4), July, pp.1104–10.

Jones, Lorelei and Nicholas Mays (2009), 'Systematic review of the impact of patient choice of provider in the English NHS', Systematic Review for the Health Reform Evaluation Programme, London School of Hygiene and Tropical Medicine, available from http://hrep.lshtm.ac.uk/publications/Choice%20review%20March%202009.pdf, accessed 25 April 2016.

Kane, Thomas J and Douglas O Staiger, 'Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation', NBER Working Paper 14607, National Bureau of Economic Research, Cambridge MA, available from http://www.nber.org/papers/w14607, accessed 8 September 2016.

Kessler, Daniel P and Mark B McClellan (2000), 'Is Hospital Competition Socially Wasteful?', *The Quarterly Journal of Economics* 115(2), May, pp.577–615.

Koedel, Cory, Kata Mihaly and Jonah E Rockoff (2015), 'Value-added modeling: a review', *Economics of Education Review* 47, August, pp.180–95.

Lilford, Richard and Peter Pronovost (2010), 'Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away', *British Medical Journal* 340:c2016.

Maddala, G S (1983), *Limited-Dependent and Qualitative Variables in Economics*, Cambridge University Press, Cambridge.

NHS Digital (2016), 'Patient Reported Outcome Measures (PROMs) in England: A guide to PROMs methodology', Version 11, 11 August, NHS Digital, available from https://www.england.nhs.uk/statistics/statistical-work-areas/proms/, accessed 4 April 2017.

NHS England (2013a), *Patient Reported Outcome Measures (PROMs): An alternative aggregation methodology for case-mix adjustment,* 1 June, available from https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-agg-meth-adju.pdf, accessed 4 April 2017.

NHS England (2013b), *Patient Reported Outcome Measures (PROMs) in England: Update to reporting and case-mix adjusting hip and knee procedure data,* October, available from https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/10/proms-meth-prim-revis.pdf, accessed 4 April 2017.

NICE [National Institute for Health and Care Excellence] (2013), *Guide to the methods of technology appraisal 2013*, 4 April, available from http://nice.org.uk/process/pmg9, accessed 8 September 2016.

Noble M, G Wright, C Dibben, GAN Smith, D McLennan, C Anttila, H Barnes, C Mokhtar, S Noble, D Avenell, J Gardner, I Covizzi and M Lloyd (2004), *Indices of Deprivation 2004: Report to the Office of the Deputy Prime Minister*, Neighbourhood Renewal Unit, Office of the Deputy Prime Minister, London.

Propper, Carol (2012), 'Competition, Incentives and the English NHS', *Health Economics* 21(1), January, pp.33–40.

Propper, Carol and Deborah Wilson (2003), 'The use and usefulness of performance measures in the public sector', *Oxford Review of Economic Policy* 19(2), pp.250–67.

Propper, Carol, Simon Burgess and Denise Gossage (2008), 'Competition and Quality: Evidence from the NHS Internal Market 1991-99', *Economic Journal* 118(525), January, pp.138–70.

Propper, Carol, Simon Burgess and Katherine Green (2004), 'Does competition between hospitals improve the quality of care? Hospital death rates and the NHS internal market', *Journal of Public Economics* 88(7/8), July, pp.1247–72.

Propper, Carol (1996), 'Market structure and prices: the responses of hospitals in the UK national health service to competition', *Journal of Public Economics* 61(3), pp.307–35.

Propper, Carol, Deborah Wilson, and Neil Söderlund (1998), 'The effects of regulation and competition in the NHS internal market: the case of GP fundholder prices', *Journal of Health Economics* 17(6), pp.645–74.

Rodríguez, German (2007), *Lecture Notes on Generalized Linear Models*, available from http://data.princeton.edu/wws509/notes/, accessed 16 August 2013.

Rothstein, Jesse (2010), 'Teacher quality in educational production: tracking, decay and student achievement', *Quarterly Journal of Economics* 125(1), pp.175–214.

Shojania, Kaveh G and Alan J Forster (2008), 'Hospital mortality: when failure is not a good measure of success', *CMAJ-JAMC (Canadian Medical Association Journal – Journal de l'Association Médicale Canadienne)* 179(2), 15 July, pp.153–7.

Smith, Sarah C, Stefan Cano, Donna L Lamping, Sophie Staniszweska, John Browne, James Lewsety, Jan van der Meulen, John Cairns and Nick Black (2005), 'Patient-Reported Outcome Measures (PROMs) for routine use in Treatment Centres: recommendations based on a review of the scientific evidence', Final report to the Department of Health, Health Services Research Unit, London School of Hygiene & Tropical Medicine, London, December, available from https://www.lshtm.ac.uk/php/hsrp/research/promsnickblack2005.pdf, accessed 16 August 2013.

Söderlund, Neil, Ivan Csaba, Alastair Gray, Ruairidh Milne, and James Raftery (1997), 'Impact of the NHS reforms on English hospital productivity: an analysis of the first three years', *British Medical Journal* 315, 1 November, pp.1126–9.

Appendix 1: Replicating the NHS PROMs casemix adjustment methodology for patient-level data

This paper has adapted the current NHS PROMs casemix adjustment methodology for provider-level data to derive risk-adjusted health gain from surgery at the patient level.[53] This methodology is somewhat simpler than the NHS methodology, as aggregation to the provider level is not required. The current NHS casemix adjustment methodology involves the estimation of a linear model with provider (hospital) fixed effects:[54]

$$Q2_i = \alpha + \beta_1 Q1_i + \boldsymbol{\beta}'_2 \boldsymbol{x}_i + \boldsymbol{\beta}'_3 \boldsymbol{z}_{ij} + u_j + \varepsilon_{ij} \tag{5}$$

In Equation (5), $Q2_i$ and $Q1_i$ denote, respectively, the post-operative and pre-operative survey score for patient i attending hospital j; $\boldsymbol{x}_i$ is a vector of patient characteristics; $\boldsymbol{z}_{ij}$ is a vector of variables containing information about the patient's hospital stay;[55] u_j is a hospital fixed effect; and ε_{ij} is an error term. Standard errors are clustered at the hospital site level. The risk adjustment is performed separately for each procedure and outcome variable included in the PROMs programme; revisions to hip and knee replacement surgery are treated separately to primary hip and knee replacements. The control variables included in $\boldsymbol{x}_i$ and $\boldsymbol{z}_{ij}$ are chosen by regressing Equation (5) using a large, standard set of controls (listed in Tables A2 and A3; NHS England 2013b), and then re-running the regression using only those variables that are significant at the 5 per cent level. From this regression the fitted $Q2_i$, $\widehat{Q2_i}$, is calculated:

$$\widehat{Q2_i} = \widehat{\alpha} + \widehat{\beta_1} Q1_i + \widehat{\boldsymbol{\beta}'_2} \boldsymbol{x}_i + \widehat{\boldsymbol{\beta}'_3} \boldsymbol{z}_{ij} + \widehat{u}_j$$

Secondly, the 'predicted' $Q2_i$, $\widetilde{Q2_i}$, is defined as the post-operative score that would have been expected in the absence of the hospital's contribution to the patient's health gain. This is calculated by subtracting the hospital fixed effect, $\hat{u}_j$, from the fitted $Q2_i$, and, as a normalisation, replacing it with $\overline{\widehat{u}}$, the (weighted) mean value of the hospital fixed effect.[56] That is:

$$\widetilde{Q2_i} = \widehat{Q2_i} - \widehat{u}_j + \overline{\widehat{u}} = \widehat{\alpha} + \widehat{\beta_1} Q1_i + \widehat{\boldsymbol{\beta}'_2} \boldsymbol{x}_i + \widehat{\boldsymbol{\beta}'_3} \boldsymbol{z}_{ij} + \overline{\widehat{u}}$$

Thirdly, the provider's Relative Performance Factor (RPF) – the ratio of actual post-operative health status to predicted post-operative health status – is calculated for patient i:[57]

$$RPF_i = \frac{Q2_i}{\widetilde{Q2_i}}$$

Finally, the adjusted $Q2$ score ($Q2a_i$) and adjusted health gain (ΔQa_i) are calculated with reference to $\overline{Q2}$ and $\overline{Q1}$, the national average (by procedure and year) $Q2$ and $Q1$ scores:[58]

$$\begin{aligned} Q2a_i &= RPF_i \cdot \overline{Q2} \\ \Delta Qa_i &= Q2a_i - \overline{Q1} \end{aligned}$$

---

[53] A rigorous and fully specified risk-adjustment methodology was published in 2012 (DH 2012a; b; c; d; e). This methodology was modified in 2013, but many of the changes made at this time are poorly documented and/or justified (NHS England 2013a; b). The approach adopted here therefore adopts some of the 2013 changes, but retains features of the 2012 methodology when those changes are either questionable or incompletely documented.

[54] The present study defines a provider as a hospital site.

[55] Both $\boldsymbol{x}_i$ and $\boldsymbol{z}_{ij}$ are vectors of patient-level control variables – $\boldsymbol{z}_{ij}$ cannot contain any variables that are invariant at the hospital level, as these are incorporated into the hospital fixed effect.

[56] If N is the total number of patients for a given procedure and financial year, and n_j is the number of hospital j's patients, so $\sum_{\forall j} n_j = N$, then $\overline{\widehat{u}} = \sum_{\forall j} \frac{n_j}{N} \hat{u}_j$

[57] Calculation and interpretation of the RPF is complicated when the outcome variable can take either positive or negative values, as is the case for the EQ-5D Index Score. In response to related problems, in 2013 the NHS moved to calculating an additive rather than a multiplicative RPF (NHS England 2013a). This paper retains the use of a multiplicative RPF as set out in DH 2012a, and instead adds 0.594 (the maximum negative value for the EQ-5D Index Score) to all $Q1$ and $Q2$ scores before performing risk adjustment, and then subtracting 0.594 back out from the $Q1$ and $Q2$ scores, and any derivatives of these (including the adjusted $Q2$ score) after performing the risk adjustment.

[58] Adjusted gain is calculated by taking the difference between the adjusted $Q2$ score and the *national average* $Q1$ score, rather than the individual patient's $Q1$ score, because the individual $Q1$ score has already been controlled for when risk adjusting $Q2$. An alternative method would be to omit the $Q1$ score from $Q2$ risk adjustment, adjust both $Q2$ *and* $Q1$ scores for casemix separately, and then calculate adjusted health gain as the difference between these two adjusted scores. The method employed here has the advantage of not requiring any adjustment of the $Q1$ scores.

As is well known from the example of estimating probabilities, estimating limited dependent variables using linear regression methods can lead to predicted values that are outside the support of the dependent variable. This is a known potential problem with existing PROMs casemix adjustment methods (Coles 2010, p.10). While there is a literature seeking to address this concern (Hernández et al. 2012; Basu and Manca 2012; Gutacker et al. 2013), it is not clear that the use of linear regression models for risk adjustment poses serious problems in contexts where, unlike the case of estimating probabilities, the absolute level of the dependent variable is not all that important, such as when comparing the performance of health care providers. Furthermore, there may be advantages to *not* truncating the support of the dependent variable when adjusting for casemix. Consider the situation of a patient who records a post-operative EQ-5D profile of 11111 – or perfect health, implying an EQ-5D index score of 1. If this patient's characteristics (including pre-operative score) made such an outcome very unlikely, the patient *should*, arguably, receive a post-operative score greater than 1, even though this implies a conceptually problematic 'more than perfect health'.

Table A1 provides the minima and maxima of the scale, observed, and adjusted minimum and maximum post-operative (Q2) scores for the different PROMs used in this paper. Table A1 demonstrates that, while the adjusted scores project outside the original support of the outcome variable in all cases, they do not do so in a way that drastically distorts the interpretation of the outcome measure. Table A2 lists the variables used for casemix adjustment, and the coefficients on each variable for each of the paper's main outcome variables.

Appendix 2: Details of predicted patient choice model

This Appendix presents the details of the predicted patient choice model used in the present study, which is based on Kessler & McClellan (2000) and Gaynor et al. (2013).[59]

A2.1 The Conditional Logit Model

This study uses a conditional logit model to predict each patient's choice of hospital based on plausibly exogenous parameters. The conditional logit model is an extension of the multinomial logit model that allows determinants of outcomes (here, hospital choices) to be a function of characteristics of those outcomes (hospitals) and not just, as in the multinomial logit model, a function of characteristics of the individuals themselves. Let H_i denote patient i's choice of hospital: $H_i = j$ denotes that hospital j is chosen. Let $\boldsymbol{x}_i$ (with hospital-specific coefficients $\boldsymbol{\theta}_j$) be a vector of individual-specific explanatory variables that are independent of choice of hospital, and let $\boldsymbol{z}_{ij}$ (with coefficients $\boldsymbol{\gamma}$) be a vector of explanatory variables that may either be hospital-level variables or patient-level variables that are a function of choice of hospital. Let patient i's utility from alternative j be an additive function of a systematic component $\boldsymbol{\theta}'_j\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij}$ and a random component ε_{ij}:

$$U_{ij} = \boldsymbol{\theta}'_j\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij} + \varepsilon_{ij}$$

Since utility has a random component (ε_{ij}), the probability that i will choose j, π_{ij}, is the probability that j is the utility maximising choice of hospital:

$$\pi_{ij} = \Pr(H_i = j) = \{\Pr(\max\{U_{i1}, U_{i2}, \ldots, U_{iJ}\}) = U_{ij}\}$$

Let η_{ij} denote the log of the odds that i will choose hospital j against reference hospital J. Then, if it is assumed that ε_{ij} is distributed standard Type 1 extreme value with cumulative distribution function $F(\varepsilon) = \exp(-\exp(-\varepsilon))$,[60] it can be shown (Maddala 1983) that η_{ij} is a linear function of the explanatory variables, $\boldsymbol{x}_i$ and $\boldsymbol{z}_{ij}$:

$$\begin{aligned} \eta_{ij} &= \log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \boldsymbol{\theta}'_j\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij} \\ \pi_{ij} &= \pi_{iJ}\exp(\eta_{ij}) = \pi_{iJ}\exp(\boldsymbol{\theta}'_j\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij}) \qquad (6) \end{aligned}$$

Summing Equation (6) over all J hospitals and noting that $\sum_{j=1}^{J}\pi_{ij} = 1$ yields that:

$$\pi_{iJ} = \frac{1}{\sum_{j=1}^{J}\exp\left(\eta_{ij}\right)} \qquad (7)$$

Plugging the expression for π_{iJ} from Equation (7) into Equation (6), and replacing the js in the former with λs, yields:

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{\sum_{\lambda=1}^{J}\exp\left(\eta_{i\lambda}\right)} = \frac{\exp(\boldsymbol{\theta}'_j\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij})}{\sum_{\lambda=1}^{J}\exp\left(\boldsymbol{\theta}'_j\boldsymbol{x}_i + \boldsymbol{\gamma}'\boldsymbol{z}_{ij}\right)}$$

The parameters of the model ($\boldsymbol{\theta}_j$s and $\boldsymbol{\gamma}$) are estimated by maximum likelihood.

A2.2 Model setup

The hospitals in a patient's choice set are defined to include their chosen hospital plus any hospital within 100km of their MSOA of residence. The choice set must also include the two closest: teaching

---

[59] This Appendix draws from Maddala (1983) and Rodríguez (2007), and above all from Gaynor et al. (2013).

[60] The Type 1 (Gumbel) extreme value distribution, also known as the double exponential distribution, has parameters μ and σ and CDF $F(x) = \exp(-\exp(-(x-\mu)/\sigma))$. The mean is $\mu + \sigma\gamma$, where γ is Euler's constant (≈ 0.577), and the variance is $\frac{1}{6}\pi^2\sigma^2$ where π is the constant *pi*. The "standard" Type 1 extreme value distribution is the case where $\mu = 0$ and $\sigma = 1$, so $F(x) = \exp(-\exp(-x))$.

hospitals; non-teaching hospitals; big hospitals (defined as larger than the median, with reference to trust admissions); small hospitals (defined as smaller than the median); NHS hospitals; and private hospitals. The choice set must include these hospitals because the model postulates that patients may have preferences over the type of hospital they attend (whether in relation to teaching status, size, or NHS vs private), and that utility from attending a given hospital is a function not only of distance to that hospital, but also of the difference between distance to that hospital, and distance to an alternative hospital with similar (or different) characteristics.

Let $h \in \{1, 2, 3\}$ denote the three dimensions of hospital type over which preferences are defined – $h = 1$ refers to the distinction between teaching and non-teaching hospitals, $h = 2$ to the distinction between big and small hospitals, and $h = 3$ to the distinction between NHS and private hospitals. Let z_j^h be a binary indicator of whether hospital j possesses characteristic h: $z_j^1 = 1$ denotes a teaching hospital, $z_j^2 = 1$ a big hospital, and $z_j^3 = 1$ an NHS hospital.

Let d_{ij} denote distance from the centroid of patient i's MSOA to hospital j, and let d_{ij+}^h denote the distance to the closest hospital that is a good substitute for hospital j in terms of characteristic h. That is, if $h = 1$ and j is a teaching hospital, then d_{ij+}^h denotes the distance to the closest teaching hospital (other than hospital j, if it is the closest). Likewise, let d_{ij-}^h denote the distance to the closest hospital that is a poor substitute for hospital j in terms of characteristic h. Thus, with $h = 1$ and j a teaching hospital, d_{ij-}^h would denote the distance to the closest non-teaching hospital. Utility from attending hospital j is defined as a function of the difference between distance to j and the distance to the nearest good/poor substitute for j in terms of each of the three dimensions of hospital types included in the model.[61] Specifically, patient i's utility from attending hospital j is defined as:

$$U_{ij} = \sum_{h=1}^{3} \left\{ \begin{array}{c} \theta_1^h \left(d_{ij} - d_{ij+}^h\right) z_j^h + \theta_2^h \left(d_{ij} - d_{ij+}^h\right) \left(1 - z_j^h\right) \\ +\theta_3^h \left(d_{ij} - d_{ij-}^h\right) z_j^h + \theta_4^h \left(d_{ij} - d_{ij-}^h\right) \left(1 - z_j^h\right) \\ +\theta_5^h \left(female_i \cdot mid_i \cdot z_j^h\right) + \theta_6^h \left(female_i \cdot old_i \cdot z_j^h\right) \\ +\theta_7^h \left(male_i \cdot young_i \cdot z_j^h\right) + \theta_8^j \left(male_i \cdot mid_i \cdot z_j^h\right) + \theta_9^h \left(male \cdot old_i \cdot z_j^h\right) \\ +\theta_{10}^h \left(lowseverity_i \cdot z_j^h\right) + \theta_{11}^h \left(highseverity_i \cdot z_j^h\right) + \theta_{12}^h \left(charlson_i \cdot z_j^h\right) \\ +\theta_{13}^h \left(urban_i \cdot z_j^h\right) + \theta_{14}^h \left(poor_i \cdot z_j^h\right) + \boldsymbol{\gamma}^h \left(\boldsymbol{region}_i \cdot z_j^h\right) + \varepsilon_{ij} \end{array} \right\} \tag{8}$$

Only two of the first four terms will be turned on for any given dimension h. For example, if j is a private hospital, then $z_j^3 = 0$ and so the θ_1^3 and θ_3^3 terms will be turned off; the θ_2^3 term will then capture utility from differential distance between j and the nearest private hospital, while the θ_4^3 term will capture utility from differential distance between j and the nearest NHS hospital.

In addition to these differential distance terms, which are used to satisfy the exclusion restriction, the model seeks to capture possible differences in preferences for different hospital types based on patient characteristics. Terms are therefore included to capture differences in utility from attending a teaching hospital, a big hospital, or an NHS hospital (relative to attending a non-teaching hospital, a small hospital, or a private hospital respectively) based on a range of exogenous variables describing patient characteristics. The number of variables that can be included in the model is constrained by the fact that computation time is relative to the square of the number of choice determinants. Casemix is therefore accounted for by dividing patients into three age categories – young (below 60), mid (61 to 75), and old (over 75), and crossing these with gender to give six dummies (one of which is omitted). Dummies are also included for low and high severity (respectively, any patient with only one diagnosis code, or with three or more diagnosis codes), as well as for the patient's Charlson score. Finally, dummies are included for urban status (any patient living in an urban area), poverty (any patient living in an area where more than 10 per cent of households are classified as being income-deprived), and the nine regions of England (the bold coefficient and variable denote vectors). All of these variables are included in the conditional logit model three times, as they are interacted with each of the three dimensions of hospital heterogeneity over which patients have preferences, so that each patient characteristic can separately enter preferences concerning each dimension.

---

[61] This definition of utility in terms of differential distances is the reason the choice set needs to include the closest two hospitals in terms of each dimension of hospital heterogeneity included in the model.

The parameters of the model are estimated separately for each surgical procedure and financial year. When estimating the model, the dataset is collapsed to include a single entry for all patients that are identical in terms of the model (that is, who attend the same hospital, live in the same MSOA, and have the same patient characteristics). All such patients within a given 'subtype' have the same choice set and the same differential distances, as distances are measured in terms of distance from MSOA centroids to hospitals. After collapsing all such identical patients, the model is then estimated using frequency weights to reflect the number of patients in each subtype.

A2.3 Model outputs

For each patient subtype, the model gives a probability of attendance for each hospital in the choice set. These probabilities sum to one, and are used in place of the subtype's actual choice of hospital. Thus, if there are 10 patients in a given subtype, and the conditional logit model gives probabilities of $\{0.2, 0, 0.4, 0.4\}$ for hospitals A, B, C and D, then the predicted patient choice model would allocate 2, 0, 4 and 4 patients from this subtype to each of these hospitals respectively. A (predicted) HHI is calculated for each MSOA by aggregating across subtypes within the MSOA to calculate the sum of hospitals' squared market shares for that MSOA; a hospital's predicted HHI is then calculated as the weighted sum of predicted HHIs of all the MSOAs that it serves.

The resulting HHIs have a correlation of about 0.4 with MSOA-centred HHIs based on actual patient choices. The choice estimates are robust to alternative specifications of the distance used to define each patient subtype's choice set.

Appendix 3: Procedure and diagnosis definitions

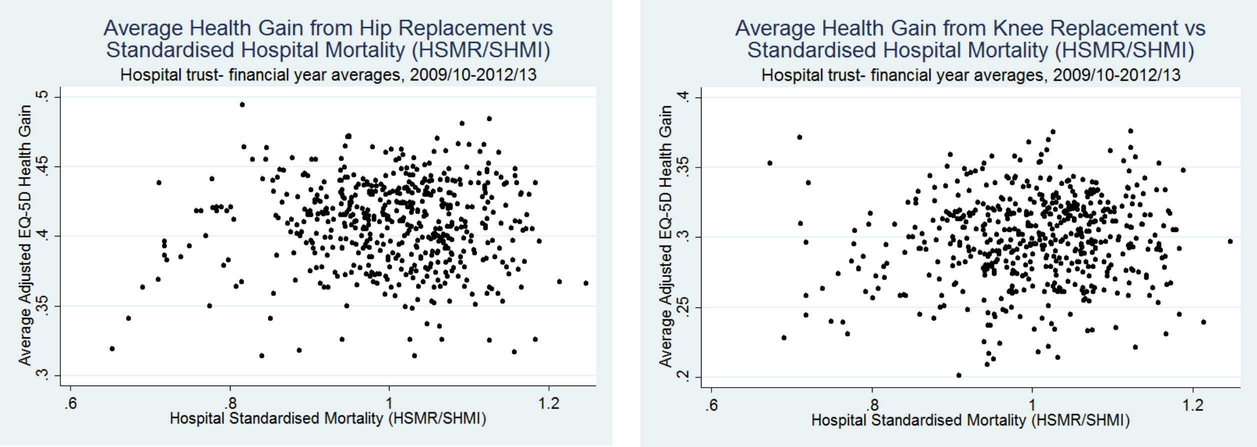
NHS Digital (2016) outlines the OPCS4 procedure codes used to define participation in the PROMs programme, as well as to distinguish between primary hip and knee replacements, and revisions to these procedures. A broader set of definitions was used to construct the dataset used in this paper, in order to define the market for each procedure in an intuitive and meaningful way for the purpose of calculating competition intensity. For example, whereas the PROMs programme only surveys groin hernia patients, I include all hernia patients in the dataset. Also, whereas patients undergoing bilateral hip and knee replacement surgery are excluded from the PROMs programme, I include these patients in the dataset. Of course, the additional records in the dataset resulting from these expanded definitions are not included in the final regressions, as they cannot be linked to a PROMs survey response.

The following OPCS4 procedure codes and ICD10 diagnosis codes were used to construct the dataset, matching on all 24 procedure fields and all 20 diagnosis fields in an episode.

> Hip replacement: (1) W37, W38, W39, W46, W47, W48, W93, W94 or W95, or (2) W52, W53, W54, or W58, in conjunction with Z761, Z756, or Z843.
>
> Knee replacement: (1) O18, W40, W41, or W42, or (2) W52, W53, or W54, in conjunction with Z765, Z771, Z774, Z844, Z845, or Z846.
>
> Varicose veins: L84, L85, L86, L87, L88, or L93.
>
> Hernia: T19 through to T27.
>
> Knee arthroscopy: W82 through to W89.
>
> Cataracts: C71 through to C77, in conjunction with ICD10 diagnosis code H25, H26, H28, or Q120.
>
> Acute myocardial infarction: ICD10 diagnosis codes I21 or I22.

Figures

Figure 1: Adjusted EQ-5D health gain vs Standardised Hospital Mortality (SHM)


(A) Adjusted EQ-5D health gain from hip replacement vs Standardised Hospital Mortality (SHM)

(B) Adjusted EQ-5D health gain from knee replacement vs Standardised Hospital Mortality (SHM)

(C) Adjusted EQ-5D health gain from groin hernia repair vs Standardised Hospital Mortality (SHM)

(D) Adjusted EQ-5D health gain from varicose vein stripping vs Standardised Hospital Mortality (SHM)

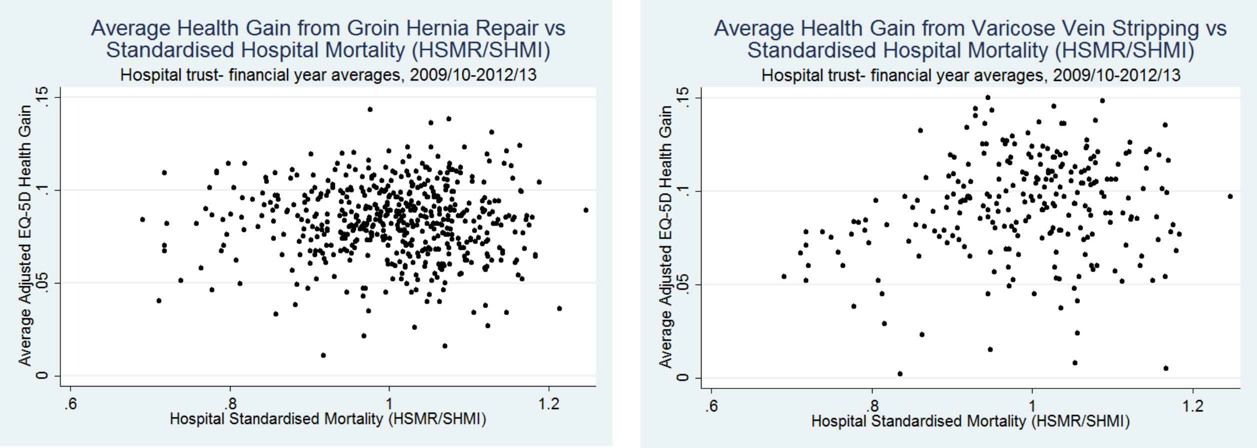
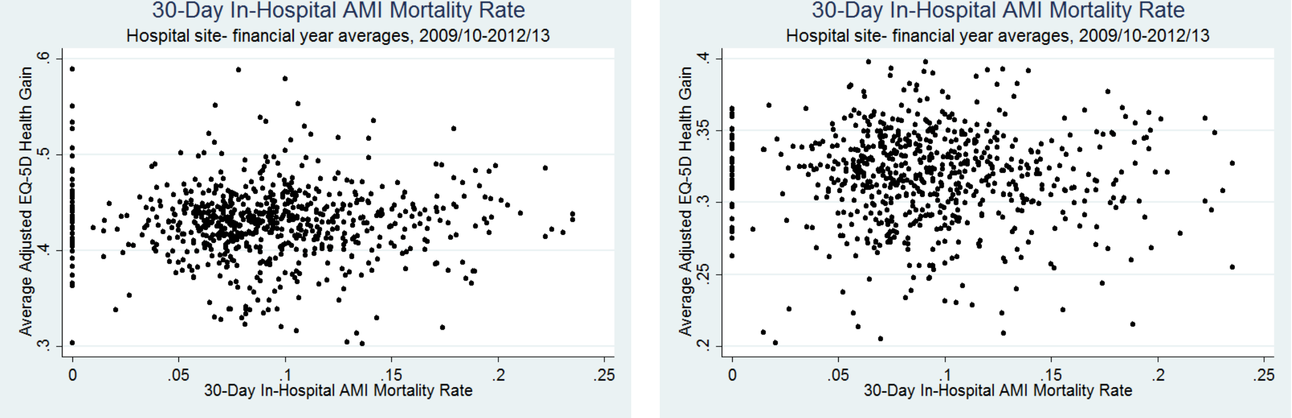


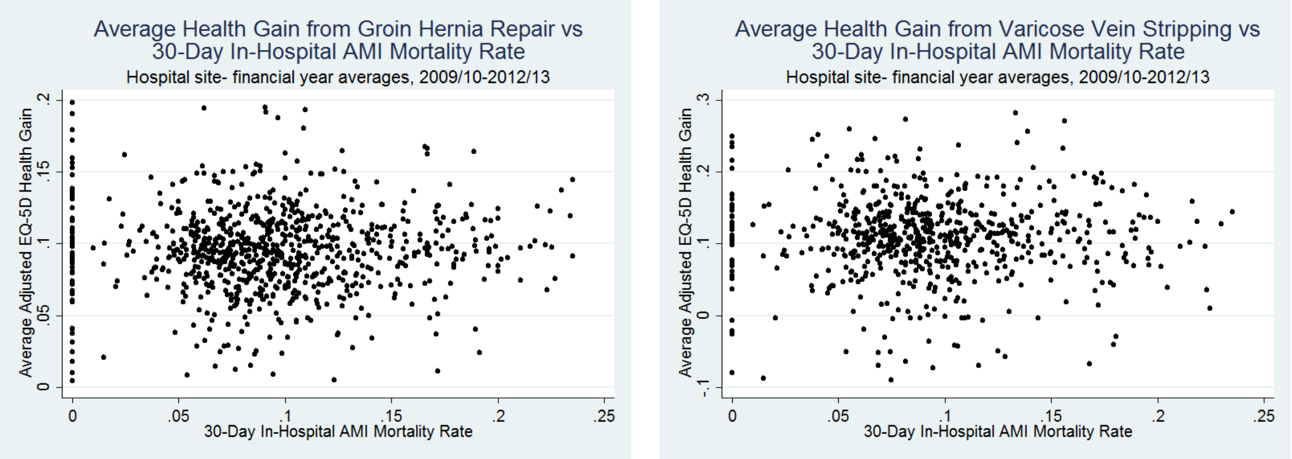
Figure graphs the relationship between hospital trusts' standardised (risk-adjusted) mortality rates (Standardised Hospital Mortality or SHM) and average casemix-adjusted health gain from elective surgery as captured by the EQ-5D PROM. In each panel, there is one observation for each hospital trust and financial year, covering the financial years 2009/10 to 2012/13. For 2009/10, SHM is captured by Dr Foster's Hospital Standardised Mortality Ratios (HSMR) (Dr Foster 2011). For 2010/11 onwards, SHM is captured by the NHS Summary Hospital Mortality Indicator (SHMI) (HSCIC 2013). In 2012/13, the NHS began reporting average PROMs health gains for revisions to hip and knee replacements separately to PROMs health gains for primary hip and knee replacements. Therefore, for this year only, PROMs health gains for hip and knee replacements are based only primary procedures, and do not include health gains from revisions.

Figure 2: Adjusted EQ-5D health gain vs AMI mortality



(A) Adjusted EQ-5D health gain from hip replacement vs AMI mortality

(B) Adjusted EQ-5D health gain from knee replacement vs AMI mortality



(C) Adjusted EQ-5D health gain from groin hernia repair vs AMI mortality

(D) Adjusted EQ-5D health gain from varicose vein stripping vs AMI mortality

Figure graphs the relationship between individual hospital sites' mortality rates from acute myocardial infarction (AMI) and average casemix-adjusted health gain from elective surgery as captured by the EQ-5D PROM. In each panel, there is one observation for each hospital site and financial year, covering the financial years 2009/10 to 2012/13. The AMI mortality rate is calculated as the 30-day in-hospital mortality rate for patients aged 39 to 100, omitting all patients discharged alive with a total length of stay of less than three days, and including only patients admitted on an emergency basis from their place of residence. Average PROMs health gains from hip and knee replacements include both primary procedures and revisions.

Figure 3: Adjusted and Unadjusted PROMs Health Gains vs Pre-Treatment Health Status

(A) EQ-5D Index Score for Hip Replacement: Unadjusted Health Gains vs Pre-Treatment Health Status

(B) EQ-5D Index Score for Hip Replacement: Adjusted Health Gains vs Pre-Treatment Health Status

(C) Oxford Hip Score for Hip Replacement: Unadjusted Health Gains vs Pre-Treatment Health Status

(D) Oxford Hip Score for Hip Replacement: Adjusted Health Gains vs Pre-Treatment Health Status

Figure graphs the relationship between hip replacement patients' pre-treatment health status and their adjusted and unadjusted health gains from treatment. Panels (A) and (B) show this relationship for the EQ-5D Index Score, with unadjusted and adjusted health gains on the y-axis respectively. Panels (C) and (D) show this relationship for the Oxford Hip Score, with unadjusted and adjusted health gains on the y-axis respectively.

Tables

Table 1: Average mortality rates for PROMs elective procedures and acute myocardial infarction

Procedure/diagnosis	Average mortality rate (per cent)
Hip replacement	0.0617
Knee replacement	0.0710
Groin hernia repair	0.0117
Varicose vein stripping	0.0000
Acute myocardial infarction (AMI)	9.1707

Table reports 30-day in-hospital mortality rates in the four years from 2009/10 to 2012/13. Includes all elective admissions for the four listed procedures, and all non-elective admissions for acute myocardial infarction (AMI). See Figure 2 notes for AMI sample restrictions.

Table 2: Outcome variables – summary statistics

Outcome	Surgical Procedure	Observations	Average health gain	Standard deviation	Min	Max
EQ-5D	Hip Replacement	122,701	0.4324	0.2251	-0.9352	1.4821
EQ-5D	Knee Replacement	130,309	0.3226	0.2359	-0.9826	2.0375
EQ-5D	Groin Hernia	78,364	0.0943	0.168	-1.3728	1.0361
EQ-5D	Varicose Veins	24,309	0.1044	0.1865	-1.341	1.3389
OHS	Hip Replacement	135,538	20.518	8.6309	-18.8077	60.1532
OKS	Knee Replacement	142,448	15.6333	9.0858	-18.3818	83.101
AVVQ	Varicose Veins	25,711	8.202	8.705	-65.8169	72.7042
EQ-VAS	Hip Replacement	118,264	10.8475	16.6161	-64.3165	95.0956
EQ-VAS	Knee Replacement	125,250	4.9042	17.0208	-66.922	120.9261
EQ-VAS	Groin Hernia	76,282	-0.0827	13.5718	-79.2098	98.3347
EQ-VAS	Varicose Veins	23,473	0.1899	14.3065	-78.4885	94.7759

Table reports summary statistics by surgical procedure for each of the PROMs (casemix-adjusted health gain) used in this paper. Abbreviations: EQ-5D (EQ-5D Index Score), OHS (Oxford Hip Score), OKS (Oxford Knee Score), AVVQ (Aberdeen Varicose Vein Questionnaire), EQ-VAS (EQ-5D Visual Analogue Score).

Table 3: Effect sizes of PROMs outcome measures

	Hip replacement	Knee replacement	Groin hernia	Varicose veins	All PROMs
EQ-5D	1.271	0.954	0.411	0.401	0.890
EQ-VAS	0.449	0.194	-0.028	-0.010	0.212
OHS/OKS/AVVQ	2.377	1.936	N/A	0.726	N/A

Table reports effect sizes (average health gain divided by standard deviation of Q1 score) of the PROMs outcome measures. See Table 2 notes for abbreviations.

Table 4: Convergent validity of PROMs outcome measures

(A) Correlation between hip replacement PROMs

	OHS	EQ-5D	EQ-VAS
OHS	1		
EQ-5D	0.6338	1	
EQ-VAS	0.3343	0.3112	1

(B) Correlation between knee replacement PROMs

	OKS	EQ-5D	EQ-VAS
OKS	1		
EQ-5D	0.5912	1	
EQ-VAS	0.3014	0.2688	1

(C) Correlation between groin hernia repair PROMs

	EQ-5D	EQ-VAS
EQ-5D	1	
EQ-VAS	0.2785	1

(D) Correlation between varicose vein stripping PROMs

	AVVQ	EQ-5D	EQ-VAS
AVVQ	1		
EQ-5D	0.3087	1	
EQ-VAS	0.1490	0.2247	1

Table reports the within-observation pairwise correlation between unadjusted health gains from treatment as captured by different PROMs for each of the surgical procedures studied in this paper. See Table 2 notes for abbreviations.

Table 5: Correlation between hospital trusts' average EQ-5D health gain from surgery and Standardised Hospital Mortality (SHM)

	SHM	Hip replacement	Knee replacement	Groin hernia	Varicose veins
(A) Correlation between unadjusted EQ-5D health gain and Standardised Hospital Mortality (SHM)					
Standardised mortality (SHM)	1				
Hip replacement: average EQ-5D health gain	0.212***	1			
Knee replacement: average EQ-5D health gain	0.1193**	0.395***	1		
Groin hernia: average EQ-5D health gain	0.0915**	0.0679	0.0273	1	
Varicose veins: average EQ-5D health gain	0.0608	0.0585	-0.0175	0.1091**	1
(B) Correlation between adjusted EQ-5D health gain and Standardised Hospital Mortality (SHM)					
Standardised mortality (SHM)	1				
Hip replacement: average adjusted EQ-5D health gain	-0.0234	1			
Knee replacement: average adjusted EQ-5D health gain	0.0903**	0.3224***	1		
Groin hernia: average adjusted EQ-5D health gain	-0.0172	0.1367***	0.0536	1	
Varicose veins: average adjusted EQ-5D health gain	0.1749***	0.1147*	0.1755***	0.0985	1

Table reports the pairwise correlations between hospital trusts' annual standardised (risk-adjusted) mortality rates (Standardised Hospital Mortality or SHM) and average annual health gains, as captured by the EQ-5D, from the four elective surgical procedures included in the PROMs programme, covering the years 2009/10 to 2012/13. See Figure 1 notes for SHM sources. Panel (A) uses raw (unadjusted) EQ-5D health gains, while Panel (B) uses casemix-adjusted EQ-5D health gains. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1. Statistical significance after applying a Bonferroni correction for multiple comparisons is denoted by the underlined stars. Average EQ-5D health gains from hip and knee replacement in 2012/13 exclude data from patients undergoing revisions to these procedures; see Figure 1 notes for more.

Table 6: Correlation between hospitals sites' average EQ-5D health gain from surgery and mortality rate from acute myocardial infarction (AMI)

	AMI mortality	Hip Replacement	Knee Replacement	Groin Hernia	Varicose Veins
(A) Correlation between unadjusted EQ-5D health gain and AMI mortality rate					
Acute myocardial infarction (AMI) mortality rate	1				
Hip replacement: average EQ-5D health gain	0.0216	1			
Knee replacement: average EQ-5D health gain	-0.0112	0.1678***	1		
Groin hernia: average EQ-5D health gain	0.0159	0.0486	0.0271	1	
Varicose veins: average EQ-5D health gain	-0.0101	-0.0474	-0.0311	0.1252***	1
(B) Correlation between adjusted EQ-5D health gain and AMI mortality rate					
Acute myocardial infarction (AMI) mortality rate	1				
Hip replacement: average adjusted EQ-5D health gain	-0.0096	1			
Knee replacement: average adjusted EQ-5D health gain	0.0149	0.1952***	1		
Groin hernia: average adjusted EQ-5D health gain	0.0219	0.0643*	-0.0167	1	
Varicose veins: average adjusted EQ-5D health gain	-0.0388	-0.0639	0.0347	0.0234	1

Table reports the pairwise correlations between hospital sites' annual mortality rates from acute myocardial infarction (AMI) and average annual health gains, as captured by the EQ-5D, from the four elective surgical procedures included in the PROMs programme, covering the years 2009/10 to 2012/13. See Figure 2 notes for AMI sample restrictions. Panel (A) uses raw (unadjusted) EQ-5D health gains, while Panel (B) uses casemix-adjusted EQ-5D health gains. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1. Statistical significance after applying a Bonferroni correction for multiple comparisons is denoted by the underlined stars. Average PROMs health gains from hip and knee replacements include both primary procedures and revisions.

Table 7: Correlation between change in hospital trusts' average adjusted EQ-5D health gain from surgery and change in Standardised Hospital Mortality (SHM)

	SHM	Hip replacement	Knee replacement	Groin hernia	Varicose veins
Change in Standardised Hospital Mortality (SHM)	1				
Hip replacement: change in average adjusted EQ-5D health gain	0.0016	1			
Knee replacement: change in average adjusted EQ-5D health gain	-0.0213	0.1003**	1		
Groin hernia: change in average adjusted EQ-5D health gain	-0.0123	0.0417	0.0585	1	
Varicose veins: change in average adjusted EQ-5D health gain	-0.0944	0.1279	0.1298	0.045	1

Table reports the pairwise correlations between hospital trusts' first differenced (year-to-year) Standardised Hospital Mortality (SHM) and first differenced (year-to-year) average casemix-adjusted health gain, as captured by the EQ-5D, from the four elective surgical procedures included in the PROMs programme, covering the years 2009/10 to 2012/13. See Figure 1 notes for sources of SHM data. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1. Statistical significance after applying a Bonferroni correction for multiple comparisons is denoted by the underlined stars. Average EQ-5D health gains from hip and knee replacement in 2012/13 exclude data from patients undergoing revisions to these procedures; see Figure 1 notes for more.

Table 8: Correlation between change in hospital sites' average adjusted EQ-5D health gain from surgery and change in AMI mortality

	AMI mortality	Hip replacement	Knee replacement	Groin hernia	Varicose veins
Change in acute myocardial infarction (AMI) mortality rate	1				
Hip replacement: change in average adjusted EQ-5D health gain	-0.0614	1			
Knee replacement: change in average adjusted EQ-5D health gain	-0.0446	0.413***	1		
Groin hernia: change in average adjusted EQ-5D health gain	0.0901**	0.0641	-0.0606	1	
Varicose veins: change in average adjusted EQ-5D health gain	-0.0709	-0.033	0.0369	0.0789*	1

Table reports the pairwise correlations between hospital sites' first differenced (year-to-year) mortality rate from acute myocardial infarction (AMI) and first differenced (year-to-year) average casemix-adjusted health gain, as captured by the EQ-5D, from the four elective surgical procedures included in the PROMs programme, covering the years 2009/10 to 2012/13. See Figure 2 notes for AMI sample restrictions. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1. Statistical significance after applying a Bonferroni correction for multiple comparisons is denoted by the underlined stars. Average PROMs health gains from hip and knee replacements include both primary procedures and revisions.

Table 9: Competition indices – summary statistics

Competition Index	Observations	Average value	Standard deviation	Min	Max
MSOA-centred HHI (predicted choices)	474,572	1.0547	0.6979	0.0667	5.1651
MSOA-centred HHI (actual choices)	494,867	0.5796	0.2344	0	1.5054
GP-centred 95% radius HHI	495,347	1.2061	0.4472	0.1785	6.572
Number of hospital sites within 30km	422,955	11.0321	9.8007	1	43.6667

Table reports summary statistics for the competition indices used in this paper. All HHIs are a weighted average of the negative log of HHI across six high-volume elective surgical procedures – hip and knee replacement, hernia repair, varicose vein surgery, knee arthroscopy, and cataract repair.

Table 10: Correlation between different indicators of competition intensity

	MSOA-centred HHI (predicted)	MSOA-centred HHI (actual)	GP-centred 95% radius HHI	No. of hospitals within 30km	Urban dummy	Pop density (pre-reform patients)	Pop density (current patients)
MSOA-centred HHI (predicted choices)	1						
MSOA-centred HHI (actual choices)	0.3974	1					
GP-centred 95% radius HHI	0.4183	0.7955	1				
Number of hospital sites within 30km	0.7112	0.1963	0.3582	1			
Dummy: lives in urban area	0.2454	0.0322	0.0577	0.2767	1		
Catchment population density (pre-reform patients)	0.7292	0.1458	0.2061	0.7822	0.2806	1	
Catchment population density (current patients)	0.6198	0.0815	0.1627	0.7881	0.2828	0.9836	1

Table reports the pairwise correlations between measures of competition intensity used in the paper, a dummy indicating that the patient resides in an urban area, and hospitals' catchment area population density based on both pre-reform patient flows and current patient flows. All HHIs are a weighted average of the negative log of HHI across six high-volume elective surgical procedures – hip and knee replacement, hernia repair, varicose vein surgery, knee arthroscopy, and cataract repair.

Table 11: Control variables – averages

	Hip Replacement	Knee Replacement	Groin Hernia	Varicose Veins
Patient age	67.75	69.15	58.7	50.56
Patient is female (%)	59.26	56.9	7.22	62.36
Annual case load (site-procedure level) (current)	424	418	451	289
Total case load (site-procedure level) (pre-reform)	793	782	1236	728
Total annual trust admissions (current)	88,727	90,172	91,869	113,641
Total annual trust admissions (pre-reform)	78,001	79,470	81,243	91,055
Patient lives in an urban area (%)	71.72	75.31	75.56	80.05
Hospital catchment area population density (current)	25.36	26.29	27.03	34.13
Hospital catchment area population density (pre-reform)	26.89	27.84	27.68	35.03
Charlson score	1.33	1.51	0.82	0.45
Dummy: low severity (1 diagnosis) (%)	15	13.35	39.9	60.14
Dummy: high severity (3 or more diagnoses) (%)	64.61	67.9	34.78	18.77
IMD Income Deprivation score (%)	12.39	13.43	13.2	15.21
Standard acute hospital (%)	50.64	51.74	56.36	56.66
Treated at a specialist hospital trust (%)	4.57	3.64	0.04	0
Treated at a teaching hospital trust (%)	12.73	12.46	13.61	21.69
Treated at a non-teaching university hospital trust (%)	16.26	16.39	14.13	15.22
Treated at a private hospital (%)	15.81	15.78	15.87	6.44
Patient treated as a day case (%)	N/A	N/A	69.15	87.27
Revision to primary hip or knee replacement (%)	8.29	5.15	N/A	N/A

Table reports average values of control variables used in this paper, separated by PROMs procedure. All variables expressed as percentages in this table are entered as proportions (between 0 and 1) in the regressions. Italicised variables are not included in the regressions, but instead are proxied by their pre-reform values, the averages for which are reported immediately below the italicised variables. 'Annual case load (site-procedure level)' is equal to the annual procedure-specific case load (Finished Consultant Episodes or FCEs) at that hospital site. It is proxied by *total* procedure-specific FCEs in the three financial years before the introduction of patient choice (2002/3-2004/5). 'Total annual trust admissions' is proxied by its *average* value over these three years. Catchment area population density (people per hectare) is defined as the average population density (from the 2011 census) of the Middle Super Output Areas (MSOAs) of residence of the hospital's patients (for the six surgical procedures used to calculate the competition indices) in the current financial year. It is proxied by a measure that, instead of using patients from the current financial year, uses patients in the three years before the introduction of patient choice. In the regressions, the dummy variable indicating a standard acute hospital is omitted.

Table 12: Cross-sectional estimates: Impact of hospital competition (negative log of actual patient choice HHI) on casemix-adjusted health gain from elective surgery

	(1)	(2)	(3)	(4)
Procedure	Orthopaedic Surgery		Varicose Vein Stripping	All PROMs Procedures
Casemix-Adjusted Outcome	EQ-5D health gain	Oxford Score health gain	Aberdeen Questionnaire health gain	EQ-5D health gain
Competition intensity	-0.00587	-0.366**	-0.657**	-0.00406
(negative log HHI)	(0.00383)	(0.174)	(0.271)	(0.00283)
Annual case load	-4.669	-150.0	-899.5	-3.434
(site-procedure level) (/10^6)	(7.523)	(330.2)	(726.5)	(4.215)
Annual case load squared	1.067	38.44	590.0	0.960
(site-procedure level) (/10^9)	(2.835)	(126.2)	(421.9)	(1.408)
Total annual trust	-0.198*	-9.593**	-12.26	-0.181**
admissions (/10^6)	(0.104)	(4.828)	(12.42)	(0.0835)
Total annual trust	1.193**	61.86**	44.22	1.035**
admissions squared (/10^{12})	(0.518)	(23.85)	(56.44)	(0.401)
Dummy: patient lives	-0.00608***	-0.359***	0.0775	-0.00493***
in an urban area	(0.00144)	(0.0485)	(0.148)	(0.00111)
Hospital catchment area	-6.214	-1,654	-2,790	-30.79
population density (/10^6)	(113.0)	(5,045)	(9,183)	(80.02)
Charlson score (/10^3)	-0.549**	3.235	-56.07	-0.608**
	(0.270)	(10.12)	(60.55)	(0.242)
Dummy: low severity	0.000400	0.0759	0.0563	0.00103
(1 diagnosis)	(0.00223)	(0.0707)	(0.160)	(0.00138)
Dummy: high severity	-0.00333**	-0.111**	-0.0122	-0.00356***
(3 or more diagnoses)	(0.00144)	(0.0558)	(0.208)	(0.00121)
IMD Income	-0.0102	-1.089***	-0.577	-0.00828
Deprivation score	(0.00692)	(0.272)	(0.739)	(0.00571)
Dummy: part of a specialist	-0.00251	0.202	-	-0.00322
hospital trust	(0.00605)	(0.299)	-	(0.00500)
Dummy: part of a teaching	-0.00332	-0.177	0.195	-0.00154
hospital trust	(0.00444)	(0.195)	(0.285)	(0.00294)
Dummy: part of a university	-0.000444	0.0462	0.571**	0.000841
hospital trust (non-teaching)	(0.00283)	(0.131)	(0.251)	(0.00231)
Dummy: private hospital	-0.00879	-0.373	-	-0.0110*
	(0.00723)	(0.227)	-	(0.00664)
Dummy: patient	-	-	-0.252	0.00259
treated as a day case	-	-	(0.175)	(0.00165)
Dummy: knee replacement	-0.117***	-5.184***	-	-0.117***
	(0.00139)	(0.0652)	-	(0.00136)
Dummy: groin hernia repair	-	-	-	-0.345***
	-	-	-	(0.00225)
Dummy: varicose vein	-	-	-	-0.332***
stripping	-	-	-	(0.00275)
Dummy: revision to hip	-0.140***	-7.430***	-	-0.140***
replacement	(0.00361)	(0.156)	-	(0.00360)
Dummy: revision to knee	-0.0565***	-4.279***	-	-0.0564***
replacement	(0.00456)	(0.150)	-	(0.00457)
Observations	195,517	215,113	22,327	277,425
R-squared	0.068	0.101	0.017	0.283

Table reports regression coefficients and standard errors clustered at the hospital (site) level. The coefficients on competition intensity give the treatment effects of interest. Competition intensity is measured as the negative log of a neighbourhood-centred, hospital-level HHI based on actual patient choices and averaged over six high-volume elective surgical procedures – see text for further information. A constant term is included, but its coefficient is not reported. Statistical significance is reported as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Columns (1) and (2) pool hip and knee replacement surgery observations in a single regression, while Column (4) pools all PROMs procedures in a single regression; dummy variables are included for all procedures except hip replacement, as well as for revisions to hip and knee replacement, to capture differences in the level of health gains across procedures. Some variables (marked in the table) are divided by a power of 10 to ease reporting of very small coefficients. A missing coefficient indicates that Stata dropped the variable due to multicollinearity. 'Annual case load' is equal to the number of finished consultant episodes (FCEs) for the hospital (site) and surgical procedure in question. 'Total annual trust admissions' is equal to the total number of admissions for the hospital trust in question. Annual case load and its quadratic are proxied by *total* procedure-specific FCEs and its quadratic in the three financial years before the introduction of patient choice (2002/3-2004/5), while total annual trust admissions and its quadratic are proxied by their average values over these three years. 'Hospital catchment area population density' is measured in people per hectare, and is calculated at the hospital site level. It is equal to the average population density (persons per hectare) across each patient's Middle Super Output Area of residence, averaging over the three financial years before the introduction of patient choice (2002/3-2004/5) and the six surgical procedures used to construct the competition indices used in this paper. The Charlson score indicates the patient's 10-year survival probability based on their health status in relation to 17 conditions likely to lead to death. The IMD income deprivation score measures the percentage of households in the patient's Lower Super Output Area of residence that are income deprived. All regressions also include the following control variables, the coefficients for which are not reported: casemix dummies with gender interacted with five-year age bins; dummies for the nine regions of England (including London) crossed with financial year; and dummies for day of week and month of year.

Table 13: First-stage estimates with pre-reform HHI instrument: Impact of hospital competition (negative log of predicted patient choice HHI) on casemix-adjusted health gain from surgery

	(1)	(2)	(3)	(4)
First Stage Outcome	Current-period competition intensity (negative log HHI)			
Procedure	Orthopaedic Surgery		Varicose Vein Stripping	All PROMs Procedures
Second Stage (Casemix-Adjusted) Outcome	EQ-5D health gain	Oxford Score health gain	Aberdeen Questionnaire health gain	EQ-5D health gain
Pre-reform average competition intensity	0.944***	0.945***	0.964***	0.940***
	(0.0234)	(0.0232)	(0.0334)	(0.0235)
Annual case load (site-procedure level) ($/10^6$)	193.5***	193.9***	98.88	76.54**
	(48.43)	(48.43)	(95.18)	(34.17)
Annual case load squared (site-procedure level) ($/10^9$)	-83.32***	-83.65***	-10.20	-23.27
	(21.42)	(21.38)	(65.21)	(16.18)
Total annual trust admissions ($/10^6$)	-1.818*	-1.798*	-1.772	-2.072**
	(0.936)	(0.937)	(1.304)	(0.907)
Total annual trust admissions squared ($/10^{12}$)	6.809	6.728	4.176	7.978*
	(4.447)	(4.448)	(6.561)	(4.414)
Dummy: patient lives in an urban area	0.0125**	0.0123**	0.00907	0.0129**
	(0.00544)	(0.00544)	(0.00835)	(0.00598)
Hospital catchment area population density ($/10^6$)	2,103***	2,099***	3,799***	2,123***
	(793.0)	(790.9)	(1,222)	(743.7)
Charlson score ($/10^3$)	-0.254	-0.262	0.0268	-0.284
	(0.339)	(0.342)	(0.662)	(0.330)
Dummy: low severity (1 diagnosis)	0.00555	0.00568	0.00253	0.00401
	(0.00496)	(0.00500)	(0.00651)	(0.00364)
Dummy: high severity (3 or more diagnoses)	0.00622	0.00623	-0.00206	0.00791**
	(0.00395)	(0.00398)	(0.00308)	(0.00349)
IMD Income Deprivation score	-0.0393*	-0.0389*	-0.00780	-0.0300
	(0.0203)	(0.0202)	(0.0188)	(0.0194)
Dummy: part of a specialist hospital trust	-0.162***	-0.161***	-	-0.195***
	(0.0521)	(0.0519)	-	(0.0600)
Dummy: part of a teaching hospital trust	-0.0760**	-0.0760**	-0.0491	-0.0704**
	(0.0322)	(0.0321)	(0.0327)	(0.0294)
Dummy: part of a university hospital trust (non-teaching)	-0.0883***	-0.0879***	-0.114***	-0.0885***
	(0.0219)	(0.0219)	(0.0422)	(0.0218)
Dummy: private hospital	0.960***	0.956***	-	1.004***
	(0.149)	(0.149)	-	(0.119)
Dummy: patient treated as a day case	-	-	0.00946	0.00258
	-	-	(0.0111)	(0.00555)
Dummy: knee replacement	-0.00270	-0.00328*	-	0.00111
	(0.00202)	(0.00199)	-	(0.00201)
Dummy: groin hernia repair	-	-	-	-0.0106
	-	-	-	(0.0137)
Dummy: varicose vein stripping	-	-	-	-0.0105
	-	-	-	(0.0124)
Dummy: revision to hip replacement	0.00379	0.00301	-	0.00241
	(0.00330)	(0.00327)	-	(0.00372)
Dummy: revision to knee replacement	0.00769	0.00824*	-	0.00522
	(0.00514)	(0.00498)	-	(0.00523)
Observations	195,314	214,904	22,327	276,980

Table reports first-stage estimates from instrumental variables estimation of the effect of hospital competition on casemix-adjusted health gain from elective surgery, where competition intensity is instrumented by its average level in the three financial years before the introduction of patient choice of hospital (2002/3 to 2004/5). The first stage for different outcome variables within a given surgical procedure involves running exactly the same regression, but each yields slightly different results because some observations will be included in one regression but not others due to survey non-completion. A constant term is included, but its coefficient is not reported. Standard errors clustered at the hospital (site) level are reported in parentheses. Statistical significance is reported as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Competition intensity is measured as the negative log of an MSOA-centred, hospital-level HHI based on predicted patient choices and averaged over six high-volume elective surgical procedures – see text for further information. See Table 12 notes for further information, all of which also applies unless it contradicts information provided here.

Table 14: Second-stage estimates with pre-reform HHI instrument: Impact of hospital competition (negative log of predicted patient choice HHI) on casemix-adjusted health gain from surgery

	(1)	(2)	(3)	(4)
First Stage Outcome	Current-period competition intensity (negative log HHI)			
Procedure	Orthopaedic Surgery		Varicose Vein Stripping	All PROMs Procedures
Second Stage (Casemix-Adjusted) Outcome	EQ-5D health gain	Oxford Score health gain	Aberdeen Questionnaire health gain	EQ-5D health gain
Competition intensity	-0.00756*	-0.458**	-0.690**	-0.00556*
(negative log HHI)	(0.00415)	(0.186)	(0.277)	(0.00306)
Annual case load	-5.620	-187.2	-902.5	-3.842
(site-procedure level) (/10^6)	(7.635)	(336.7)	(725.4)	(4.251)
Annual case load squared	1.415	52.71	591.9	1.063
(site-procedure level) (/10^9)	(2.867)	(128.4)	(422.0)	(1.413)
Total annual trust	-0.208**	-10.16**	-12.38	-0.191**
admissions (/10^6)	(0.105)	(4.906)	(12.47)	(0.0839)
Total annual trust	1.264**	65.44***	44.85	1.096***
admissions squared (/10^{12})	(0.522)	(24.22)	(56.75)	(0.404)
Dummy: patient lives	-0.00586***	-0.349***	0.0812	-0.00473***
in an urban area	(0.00143)	(0.0483)	(0.149)	(0.00111)
Hospital catchment area	19.28	-266.0	-2,208	-7.683
population density (/10^6)	(112.4)	(5,056)	(9,402)	(80.66)
Charlson score (/10^3)	-0.565**	2.908	-56.04	-0.625***
	(0.269)	(10.11)	(60.27)	(0.241)
Dummy: low severity	0.000417	0.0765	0.0563	0.00105
(1 diagnosis)	(0.00222)	(0.0703)	(0.159)	(0.00138)
Dummy: high severity	-0.00329**	-0.110**	-0.0124	-0.00353***
(3 or more diagnoses)	(0.00143)	(0.0557)	(0.207)	(0.00121)
IMD Income	-0.0106	-1.112***	-0.578	-0.00855
Deprivation score	(0.00691)	(0.271)	(0.736)	(0.00572)
Dummy: part of a specialist	-0.00274	0.184	-	-0.00346
hospital trust	(0.00593)	(0.293)	-	(0.00485)
Dummy: part of a teaching	-0.00354	-0.194	0.187	-0.00182
hospital trust	(0.00445)	(0.194)	(0.281)	(0.00296)
Dummy: part of a university	-0.000770	0.0279	0.563**	0.000534
hospital trust (non-teaching)	(0.00284)	(0.130)	(0.256)	(0.00234)
Dummy: private hospital	-0.00871	-0.360	-	-0.0107*
	(0.00712)	(0.221)	-	(0.00650)
Dummy: patient	-	-	-0.251	0.00256
treated as a day case	-	-	(0.174)	(0.00165)
Dummy: knee replacement	-0.117***	-5.184***	-	-0.117***
	(0.00139)	(0.0650)	-	(0.00136)
Dummy: groin hernia repair	-	-	-	-0.345***
	-	-	-	(0.00225)
Dummy: varicose vein	-	-	-	-0.332***
stripping	-	-	-	(0.00275)
Dummy: revision to hip	-0.140***	-7.437***	-	-0.140***
replacement	(0.00361)	(0.155)	-	(0.00361)
Dummy: revision to knee	-0.0565***	-4.279***	-	-0.0564***
replacement	(0.00455)	(0.150)	-	(0.00456)
Observations	195,314	214,904	22,327	276,980
R-squared	0.0684	0.101	0.0166	0.283

Table reports second-stage estimates from instrumental variables estimation of the effect of hospital competition on casemix-adjusted health gain from elective surgery, where competition intensity is instrumented by its average level in the three financial years before the introduction of patient choice of hospital (2002/3 to 2004/5). The coefficients on (instrumented) competition intensity give the treatment effects of interest. A constant term is included, but its coefficient is not reported. Standard errors clustered at the hospital (site) level are reported in parentheses. Statistical significance is reported as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Competition intensity is measured as the negative log of an MSOA-centred, hospital-level HHI based on predicted patient choices and averaged over six high-volume elective surgical procedures – see text for further information. See Table 12 notes for further information, all of which also applies unless it contradicts information provided here.

Table 15: Impact of hospital competition (negative log of predicted patient choice HHI) on casemix-adjusted health gain from hip and knee replacement surgery

	(1)	(2)	(3)	(4)
Procedure	Hip Replacement		Knee Replacement	
Casemix-Adjusted Outcome	EQ-5D health gain	Oxford Score health gain	EQ-5D health gain	Oxford Score health gain
Competition intensity	-0.00804	-0.514**	-0.00869*	-0.494**
(negative log HHI)	(0.00508)	(0.233)	(0.00499)	(0.221)

Table reports second-stage estimates from instrumental variables estimation of the effect of hospital competition on casemix-adjusted health gain from hip and knee replacement surgery, where competition intensity is instrumented by its average level in the three financial years before the introduction of patient choice of hospital (2002/3 to 2004/5). Standard errors clustered at the hospital (site) level are reported in parentheses. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1. Competition intensity is measured as the negative log of an MSOA-centred, hospital-level HHI based on predicted patient choices and averaged over six high-volume elective surgical procedures – see text for further information. See Table 12 notes for further information, all of which also applies unless it contradicts information provided here.

Table 16: Impact of hospital competition (negative log of predicted patient choice HHI) on casemix-adjusted health gain from surgery: Other outcome measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Procedure	Groin Hernia	Varicose Vein	Orthopaedic Surgery	Groin Hernia	Varicose Vein	All PROMs	Hip Replacement	Knee Replacement
Casemix-Adjusted Outcome	EQ–5D Index Score		EQ – Visual Analogue Score					
Competition intensity	-0.00130	-0.00227	-0.235	0.0171	-0.208	-0.158	-0.152	-0.374
(negative log HHI)	(0.00229)	(0.00515)	(0.249)	(0.184)	(0.349)	(0.184)	(0.299)	(0.292)

Table reports second-stage estimates from instrumental variables estimation of the effect of hospital competition on casemix-adjusted health gain from elective surgery as captured by alternative PROMs indicators, where competition intensity is instrumented by its average level in the three financial years before the introduction of patient choice of hospital (2002/3 to 2004/5). Standard errors clustered at the hospital (site) level are reported in parentheses. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1. Competition intensity is measured as the negative log of an MSOA-centred, hospital-level HHI based on predicted patient choices and averaged over six high-volume elective surgical procedures – see text for further information. See Table 12 notes for further information, all of which also applies unless it contradicts information provided here.

Table 17: Impact of hospital competition (negative log of predicted patient choice HHI) on casemix-adjusted health gain from surgery: Alternative competition indices

	(1)	(2)	(3)	(4)
Procedure	Orthopaedic Surgery		Varicose Vein Stripping	All PROMs Procedures
Casemix-Adjusted Outcome	EQ-5D health gain	Oxford Score health gain	Aberdeen Questionnaire health gain	EQ-5D health gain
(1) MSOA-centred, hospital-level	-0.0221*	-1.470***	-0.0719	-0.0143
HHI (actual patient choices)	(0.0117)	(0.497)	(0.974)	(0.00893)
(2) GP-centred, hospital-level	-0.00546	-0.486**	-0.445	-0.00300
95% radius HHI	(0.00426)	(0.198)	(0.679)	(0.00355)
(3) Competition = number of	-0.000355	-0.0272	-0.0534**	-0.000219
competitors within 30km	(0.000375)	(0.0181)	(0.0227)	(0.000305)

Table reports robustness tests of the second-stage estimates reported in Table 14 using alternative indicators of competition intensity. Standard errors clustered at the hospital (site) level are reported in parentheses. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1. See Table 12 notes for further information, all of which also applies unless it contradicts information provided here. Row (1) uses an HHI based on actual patient choices rather than predicted patient choices. Row (2) uses a GP-centred, hospital-level 95% radius HHI. Row (3) defines competition as being equal to the number of competitor hospitals within a 30km radius.

Table 18: Impact of hospital competition (negative log of predicted patient choice HHI) on casemix-adjusted health gain from surgery: Robustness tests

	(1)	(2)	(3)	(4)
Procedure	Orthopaedic Surgery		Varicose Vein Stripping	All PROMs Procedures
Casemix-Adjusted Outcome	EQ-5D health gain	Oxford Score health gain	Aberdeen Questionnaire health gain	EQ-5D health gain
(1) Omit all London observations	-0.00864*	-0.466**	-0.686**	-0.00633*
	(0.00445)	(0.203)	(0.310)	(0.00329)
(2) Include control for Foundation	-0.00746*	-0.460**	-0.661**	-0.00564*
Trust status	(0.00413)	(0.188)	(0.287)	(0.00307)
(3) Restrict to NHS hospitals	-0.00739*	-0.456**	-0.690**	-0.00543*
	(0.00416)	(0.186)	(0.277)	(0.00307)
(4) Don't interact year and region	-0.00675	-0.424**	-0.664**	-0.00493
of England controls	(0.00422)	(0.189)	(0.282)	(0.00312)
(5) No urban status control,	-0.00875**	-0.546***	-0.700***	-0.00683***
no population density control	(0.00357)	(0.160)	(0.225)	(0.00262)
(6) No population density control	-0.00732**	-0.461***	-0.716***	-0.00566**
	(0.00359)	(0.161)	(0.231)	(0.00263)
(7) No urban status control	-0.00862**	-0.521***	-0.676**	-0.00645**
	(0.00415)	(0.187)	(0.274)	(0.00306)
(8) Population density based on	-0.00770*	-0.464**	-0.701**	-0.00561*
current patient flows	(0.00421)	(0.189)	(0.282)	(0.00314)
(9) Instrument all scale effects	-0.00535	-0.391**	-0.626**	-0.00470
variables	(0.00443)	(0.188)	(0.287)	(0.00322)
(10) Omit hospital scale effects	-0.00425	-0.294	-0.686**	-0.00337
variables	(0.00420)	(0.183)	(0.273)	(0.00318)
(11) All non-binary variables	-0.00418*	-0.00697**	-0.00669**	-0.00349*
logged	(0.00251)	(0.00330)	(0.00296)	(0.00201)
(12) Include hospital fixed effects	0.000796	0.131	0.521	0.000205
	(0.00262)	(0.0969)	(0.515)	(0.00196)

Table reports robustness tests of the second-stage estimates reported in Table 14, from instrumental variables estimation of the effect of hospital competition on casemix-adjusted health gain from elective surgery, where competition intensity is instrumented by its average level in the three financial years before the introduction of patient choice of hospital (2002/3 to 2004/5). Standard errors clustered at the hospital (site) level are reported in parentheses. Statistical significance is reported as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. See Table 12 notes for further information, all of which also applies unless it contradicts information provided here. Row (1) omits all observations from London hospitals. Row (2) includes a control for Foundation Trust status. Row (3) omits all observations in which the patient attended a private hospital. Row (4) includes separate year and region of England dummies, as opposed to interacting these dummies, as is the case in the main specification. Row (5) omits controls for patient's urban status and hospital catchment area population density; Row (6) omits only the population density control, while Row (7) omits only the urban status control. Row (8) uses a population density control based on current patient flows, as opposed to patient flows in the three years before the introduction of patient choice (2002/3-2004/5). Row (9) instruments total trust admissions and procedure-specific hospital site admissions with their pre-reform average values, as opposed to using pre-reform average values as proxies for current-period values. To address estimation problems (high collinearity between instruments), the estimates reported in Row (9) omit the quadratic procedure-specific and trust-wide scale effects controls. Row (10) omits hospital scale effects controls (total trust admissions, procedure-specific hospital site admissions, and their respective quadratic terms) altogether. Row (11) enters any non-binary variables in logs, while Row (12) includes hospital (site) fixed effects.

Table 19: Impact of hospital competition (negative log of predicted patient choice HHI) on casemix-adjusted health gain from surgery: Time trends

	(1)	(2)	(3)	(4)
Procedure	Orthopaedic Surgery		Varicose Vein Stripping	All PROMs Procedures
Casemix-Adjusted Outcome	EQ-5D health gain	Oxford Score health gain	Aberdeen Questionnaire health gain	EQ-5D health gain
Competition intensity (negative log HHI)	-0.0100* (0.00597)	-0.513** (0.236)	-0.407 (0.341)	-0.00755* (0.00428)
Competition intensity interacted with time trend	0.00161 (0.00204)	0.0357 (0.0734)	-0.219 (0.158)	0.00132 (0.00156)

Table reports robustness tests of the second-stage estimates reported in Table 14 when an additional term is included interacting current-period competition intensity with a linear time trend from 2009/10 (t = 0) to 2012/13 (t = 3). This interaction term, as well as current period competition intensity, are instrumented by pre-reform average competition intensity, as well as by an interaction term between pre-reform average competition intensity and a linear time trend from 2009/10 to 2012/13. The coefficient on both excluded variables is reported. Standard errors clustered at the hospital (site) level are reported in parentheses. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1. See Table 12 notes for further information, all of which also applies unless it contradicts information provided here.

Table A1: Minimum and maximum values of post-operative health status scores

Outcome variable	Adjusted min	Scale min	Observed min	Observed max	Scale max	Adjusted max
EQ-5D Index Score	-0.594	-0.594	-0.594	1	1	2.323
EQ-5D Visual Analogue Score	0	0	0	100	100	183.734
Oxford Hip Score	0	0	0	48	48	77.678
Oxford Knee Score	0	0	0	48	48	101.483
Aberdeen Varicose Vein Questionnaire	14.387	0.342	13.362	100	100	152.908

Table reports minimum and maximum Q2 (post-operative) values of PROMs studied in this paper, before and after casemix adjustment. The “Scale min” and “Scale max” columns report the minimum and maximum possible values of each PROM before adjustment. Unlike the other PROMs used in this paper, for the Aberdeen Varicose Vein Questionnaire (AVVQ), a higher score denotes worse health status. In this paper, including in Table A1, the AVVQ score is reversed, so that 0 denotes the worst possible health state, and 100 denotes perfect health. More precisely, the worst possible health state is 0.342, due to rounding of the weights used for each question.

Table A2: Coefficients from final stage of risk adjustment of post-operative PROMs scores: Orthopaedic surgery

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Procedure	Hip Replacement – Primary		Hip Replacement – Revision		Knee Replacement – Primary		Knee Replacement – Revision	
Casemix-Adjusted Outcome	EQ-5D	Oxford Score	EQ-5D	Oxford Score	EQ-5D	Oxford Score	EQ-5D	Oxford Score
Pre-treatment (Q1) score	0.446**	0.360**	0.188**	0.503**	0.634**	0.517**	0.850**	0.720**
	(0.0214)	(0.0127)	(0.00888)	(0.0365)	(0.0278)	(0.0152)	(0.102)	(0.0572)
Pre-treatment (Q1) score squared	-0.180**	-0.00472**	-	-0.00441**	-0.256**	-0.00600**	-0.327**	-0.00772**
	(0.0112)	(0.000277)	-	(0.000706)	(0.0143)	(0.000318)	(0.0532)	(0.00139)
Dummy: Female	-0.0157**	-1.017**	-	-0.644**	-0.0100**	-0.745**	0.0221**	0.508**
	(0.00122)	(0.0462)	-	(0.169)	(0.00148)	(0.0508)	(0.00757)	(0.250)
Dummy: Q1 assisted	0.0123**	-	-	-	0.0056**	0.218**	-	-
	(0.00202)	-	-	-	(0.00191)	(0.0657)	-	-
Dummy: Q2 assisted	-0.0570**	-1.234**	-0.0500**	-0.440	-0.0546**	-1.154**	-0.0945**	-2.385**
	(0.00305)	(0.0923)	(0.0110)	(0.294)	(0.00284)	(0.106)	(0.0127)	(0.428)
Dummy: Living alone at time of Q1	-	-	-	-	-0.0036**	-	-	-
	-	-	-	-	(0.00145)	-	-	-
Dummy: Living alone at time of Q2	-	-	-	-		-	-	-
	-	-	-	-		-	-	-
Dummy: Disabled at time of Q1 survey	0.00599**	0.428**		0.958**	0.00875*	1.076**	0.0178*	1.054**
	(0.00140)	(0.0517)		(0.236)	(0.00151)	(0.0565)	(0.00918)	(0.318)
Dummy: Disabled at time of Q2 survey	-0.248**	-8.838**	-0.233**	-9.399**	-0.223**	-8.873**	-0.239**	-9.439**
	(0.00164)	(0.0676)	(0.00523)	(0.220)	(0.00177)	(0.0692)	(0.00728)	(0.252)
Dummy: Previous surgery on body part	-	-0.998**	-0.0436**	-1.250**	-	-0.887**	-	-
	-	(0.147)	(0.0105)	(0.369)	-	(0.126)	-	-
Dummy: Heart disease	-	-0.196**	-	-	-0.0071**	-0.378**	-	-0.757*
	-	(0.0824)	-	-	(0.00225)	(0.0700)	-	(0.399)
Dummy: High blood pressure	-	-	-	-	0.00337*	0.163**	-	-
	-	-	-	-	(0.00129)	(0.0431)	-	-
Dummy: Stroke	-	-	-	-	-0.0140**	-0.538**	-	-
	-	-	-	-	(0.00566)	(0.206)	-	-
Dummy: Poor circulation	-0.0448**	-2.346**	-0.0420**	-2.212**	-0.0425**	-2.207**	-0.0502**	-1.764**
	(0.00303)	(0.104)	(0.0127)	(0.357)	(0.00281)	(0.0976)	(0.0121)	(0.394)
Dummy: Lung disease	-0.00770**	-	-	-	-	-	-	-0.663*
	(0.00258)	-	-	-	-	-	-	(0.400)
Dummy: Diabetes	-	-0.435**	-	-	-0.0067**	-0.713**	-	-
	-	(0.0876)	-	-	(0.00216)	(0.0728)	-	-
Dummy: Kidney disease	0.00926*	-	-	-	-	-	-	-
	(0.00538)	-	-	-	-	-	-	-
Dummy: Nervous system diseases	-	-	-	-	-0.0278**	-	-0.0885**	-
	-	-	-	-	(0.00831)	-	(0.0309)	-
Dummy: Liver disease	-	-	-	-2.764**	-	-	-	-
	-	-	-	(1.124)	-	-	-	-
Dummy: Cancer	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-
Dummy: Depression	-0.102**	-1.816**	-0.124**	-2.855**	-0.0965**	-1.534**	-0.105**	-1.683**
	(0.00316)	(0.101)	(0.0120)	(0.384)	(0.00280)	(0.0901)	(0.0144)	(0.375)
Dummy: Arthritis	-0.0142**	-0.189**	-0.0156**	-0.657**	-0.0083**	0.0944**	-	0.587**
	(0.00120)	(0.0422)	(0.00568)	(0.171)	(0.00140)	(0.0478)	-	(0.245)
Age	-	0.0772**	-	-	0.0103**	0.402**	0.00422**	0.389**
	-	(0.0171)	-	-	(0.00079	(0.0328)	(0.000393)	(0.122)
Age Squared	-	-0.000710**	-	-	-5.7e-5**	-0.00244**	-	-0.00183**
	-	(0.000134)	-	-	(5.7e-6)	(0.000237)	-	(0.000912)
Dummy: Mixed ethnicity	-	-	-	-	-	-	-0.166**	-4.636**
	-	-	-	-	-	-	(0.0764)	(1.870)

Dummy: Asian	-0.0563**	-3.697**	-	-3.659**	-0.0212**	-2.166**	-	-
ethnicity	(0.0137)	(0.492)	-	(1.018)	(0.00605)	(0.197)	-	-
Dummy: Black	-0.0311**	-2.829**	-	-	-0.0271**	-2.509**	-	-
ethnicity	(0.0118)	(0.412)	-	-	(0.00743)	(0.300)	-	-
Dummy: Other	-	-1.072**	-	-	-	-1.141**	-	-
ethnicity	-	(0.382)	-	-	-	(0.453)	-	-
Dummy: Unknown	0.00765**	0.423**	-	-	0.0107**	0.253**	-	-
ethnicity	(0.00262)	(0.0838)	-	-	(0.00297)	(0.0958)	-	-
Charlson score	-	0.0276**	-0.00164	-	-0.0021**	-	-	-
	-	(0.0121)	(0.00132)	-	(0.00032)	-	-	-
Dummy: Day case	-	-	0.209**	-	-	-	-	-
	-	-	(0.0221)	-	-	-	-	-
Dummy: One HES	-0.00480**	-0.181**	-	-	-0.0033**	-0.0686	-	-
comorbidity	(0.00187)	(0.0652)	-	-	(0.00159)	(0.0531)	-	-
Dummy: Two HES	-0.0114**	-0.464**	-	-	-	-	-	-
comorbidities	(0.00182)	(0.0662)	-	-	-	-	-	-
Dummy: Three plus	-0.0236**	-0.887**	-	-	-0.0120**	-0.464**	-	-
HES comorbidities	(0.00191)	(0.0709)	-	-	(0.00170)	(0.0554)	-	-
Dummy: Self-	-	-	-	-	-0.0195**	-0.945**	-	-
discharge	-	-	-	-	(0.00847)	(0.305)	-	-
IMD overall score	-7.44e-4**	-0.0420**	-0.00122**	-0.0529**	-7.7e-4**	-0.0412**	-0.00128**	-0.0665**
	(5.79e-5)	(0.00201)	(0.000212)	(0.00761)	(5.0e-5)	(0.00185)	(0.000268)	(0.00971)
Dummy: Malunion	-0.0722**	-2.707**	-	-	-	-	-	-
of fracture	(0.0176)	(0.584)	-	-	-	-	-	-
Dummy: Rheumatic	-0.0754**	-	-	-	-	3.114**	-	-
arthritis	(0.0129)	-	-	-	-	(0.260)	-	-
Dummy: General	-	-	-	-	-0.0138**	-	-	-
arthritis	-	-	-	-	(0.00325)	-	-	-
Dummy: Mechanical	-	-	0.0442**	1.928**	-0.172**	-	0.0263**	1.412**
complications	-	-	(0.00594)	(0.225)	(0.0671)	-	(0.00789)	(0.221)
Dummy:	0.0176**	0.349**	-	1.401**	-	-	-	4.301**
Coxarthrosis	(0.00214)	(0.0807)	-	(0.449)	-	-	-	(1.552)
Dummy: Arthrosis	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-
Dummy: Unspecified	-	-	0.173**	7.936**	-	-	-	-4.201**
joint disorder	-	-	(0.0220)	(1.751)	-	-	-	(0.353)
Dummy: Fracture	-	-			-	-	-	-
	-	-			-	-	-	-
Dummy: Symptoms	-0.00967**	-0.430**	-0.0338**	-1.493**	-	-	-0.0542**	-3.124**
period 1-5 years	(0.00145)	(0.0534)	(0.00721)	(0.237)	-	-	(0.0113)	(0.401)
Dummy: Symptoms	-0.0193**	-0.780**	-0.0588**	-2.485**	-	0.557**	-0.0597**	-3.419**
period 6-10 years	(0.00244)	(0.0919)	(0.0112)	(0.323)	-	(0.0575)	(0.0125)	(0.404)
Dummy: Symptoms	-0.0238**	-0.876**	-0.0415**	-2.025**	0.00693*	0.594**	-0.0677**	-3.417**
period 10 plus years	(0.00292)	(0.106)	(0.00822)	(0.268)	(0.00155)	(0.0588)	(0.0126)	(0.447)
Observations	115,816	127,831	9,317	10,392	126,165	137,919	6,192	6,771
R-squared	0.373	0.368	0.358	0.417	0.349	0.370	0.376	0.428

Table reports coefficients from the second stage of risk-adjustment of post-treatment PROMs scores as outlined in Appendix 1. The second stage regressions only include variables found to be statistically significant at the 5 per cent level in the first stage of the risk adjustment exercise; coefficients from first stage regressions available on request. Regressions include hospital site fixed effects and a constant, the coefficient on which is not reported. Standard errors clustered at the hospital (site) level are reported in parentheses. Statistical significance is reported as follows: ** $p<0.05$, * $p<0.1$. Variables in italics are derived from PROMs dataset, while variables not in italics are derived from HES. Variables are listed in the table even when not included in any of the reported regressions, in order to indicate that they were included in the first stage but found to have a statistically insignificant effect. The list of variables included in the risk adjustment exercise is identical to that used by NHS for risk-adjustment of Trust-level average PROMs scores (NHS England 2013b). An identical risk adjustment exercise is conducted for the EQ-VAS outcome variable; coefficients available on request.

Table A3: Coefficients from final stage of risk adjustment of post-operative PROMs scores: Groin hernia repair and varicose vein stripping

	(1)	(2)	(3)
Procedure	Groin Hernia Repair	Varicose Vein Stripping	
Casemix-Adjusted Outcome	EQ-5D	EQ-5D	Aberdeen Varicose Vein Questionnaire
Pre-treatment (Q1) score	0.237**	0.432**	0.674**
	(0.00439)	(0.0437)	(0.0650)
Pre-treatment (Q1) score squared	-	-0.0506**	-0.00117**
	-	(0.0166)	(0.000419)
Dummy: Female	-0.0224**	-0.0179**	-1.346**
	(0.00237)	(0.00227)	(0.110)
Dummy: Q1 assisted	0.0109**	-	-
	(0.00246)	-	-
Dummy: Q2 assisted	-0.0439**	-0.0319**	-0.774**
	(0.00403)	(0.00814)	(0.347)
Dummy: Living alone at time of Q1 survey	-0.00653**	-	-
	(0.00168)	-	-
Dummy: Living alone at time of Q2 survey	-	-0.0128**	-0.393**
	-	(0.00304)	(0.161)
Dummy: Disabled at time of Q1 survey	-0.0466**	-0.0551**	-
	(0.00304)	(0.00605)	-
Dummy: Disabled at time of Q2 survey	-0.172**	-0.190**	-3.485**
	(0.00320)	(0.00659)	(0.246)
Dummy: Previous surgery on body part	0.00245**	-0.0211**	-2.470**
	(0.00118)	(0.00224)	(0.122)
Dummy: Heart disease	-	-	-
	-	-	-
Dummy: High blood pressure	-	-	-
	-	-	-
Dummy: Stroke	0.0182**	-	-
	(0.00567)	-	-
Dummy: Poor circulation	-0.0443**	-0.0421**	-1.420**
	(0.00395)	(0.00360)	(0.170)
Dummy: Lung disease	-	-	-
	-	-	-
Dummy: Diabetes	0.00921**	-	-
	(0.00288)	-	-
Dummy: Kidney disease	-	-	-
	-	-	-
Dummy: Nervous system diseases	-0.0230**	-	-
	(0.00754)	-	-
Dummy: Liver disease	0.0310**	0.0325**	-1.843*
	(0.0109)	(0.0157)	(0.970)
Dummy: Cancer	-	-	-
	-	-	-
Dummy: Depression	-0.0856**	-0.0775**	-0.666**
	(0.00397)	(0.00551)	(0.224)
Dummy: Arthritis	-0.0358**	-0.0279**	-
	(0.00172)	(0.00347)	-
Age	-	-	-
	-	-	-
Age Squared	-	5.22e-06**	-
	-	(9.76e-07)	-
Dummy: Mixed ethnicity	-	-	-
	-	-	-
Dummy: Asian ethnicity	-0.0503**	-0.0415**	-2.849**
	(0.00552)	(0.0104)	(0.492)

Dummy: Black ethnicity	-0.0263**	-	-
	(0.00787)	-	-
Dummy: Other ethnicity	-0.0240**	-0.0368**	-2.500**
	(0.00858)	(0.0120)	(0.526)
Dummy: Unknown ethnicity	0.00236	-	-
	(0.00171)	-	-
Charlson score	-0.000827*	-	-
	(0.000424)	-	-
Dummy: Day case	0.00412**	-	-
	(0.00151)	-	-
Dummy: One HES comorbidity	-0.000339	-0.0106**	-
	(0.00126)	(0.00278)	-
Dummy: Two HES comorbidities	-0.000459	-0.0220**	-
	(0.00161)	(0.00434)	-
Dummy: Three plus HES comorbidities	-0.00491**	-0.0245**	-0.255
	(0.00191)	(0.00446)	(0.248)
Dummy: Self-discharge	-	-	-
	-	-	-
IMD overall score	-0.000556**	-0.000601**	-0.0164**
	(5.30e-05)	(9.73e-05)	(0.00465)
Dummy: Symptoms period 1-5 years	-	-	-
	-	-	-
Dummy: Symptoms period 6-10 years	-	-	-
	-	-	-
Dummy: Symptoms period 10 plus years	-	-	-
	-	-	-
Dummy: Symptoms period 1 plus years	-0.00629**	-	-
	(0.00110)	-	-
Observations	79,466	25,935	27,426
R-squared	0.351	0.423	0.403

Table reports coefficients from the second stage of risk-adjustment of post-treatment PROMs scores as outlined in Appendix 1. The second stage regressions only include variables found to be statistically significant at the 5 per cent level in the first stage of the risk adjustment exercise; coefficients from first stage regressions available on request. Regressions include hospital site fixed effects and a constant, the coefficient on which is not reported. Standard errors clustered at the hospital (site) level are reported in parentheses. Statistical significance is reported as follows: ** p<0.05, * p<0.1. Variables in italics are derived from PROMs dataset, while variables not in italics are derived from HES. Variables are listed in the table even when not included in any of the reported regressions, in order to indicate that they were included in the first stage but found to have a statistically insignificant effect. The list of variables included in the risk adjustment exercise is identical to that used by NHS for risk-adjustment of Trust-level average PROMs scores (NHS England 2013b). An identical risk adjustment exercise is conducted for the EQ-VAS outcome variable; coefficients available on request. The first stage risk adjustment regressions for groin hernia repair include a dummy variable indicating a symptoms period of 1 plus years and omitting the other symptoms period dummies, while for varicose vein stripping the other three symptoms period dummies are included and the 1 plus years dummy is omitted.

Table A4: Trust-Level Regressions of Log of Standardised Hospital Mortality on Log of Average EQ-5D Health Gain

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Variable	Log of Trusts' Standardised Hospital Mortality							
Explanatory Variable Casemix-Adjusted?	No				Yes			
Procedure On Which Explanatory Variable Is Based	Hip	Knee	Groin	Vein	Hip	Knee	Groin	Vein
Log of Trusts' Average EQ-5D Health Gain	0.191*** (0.0362)	0.0903** (0.0298)	0.0230** (0.00850)	0.00591 (0.00792)	-0.0121 (0.0531)	0.0853** (0.0383)	-0.00510 (0.0151)	0.0344** (0.0146)
Observations	554	554	568	452	533	542	532	256
R-squared	0.048	0.016	0.013	0.001	0.000	0.009	0.000	0.021

Table reports coefficients from univariate regression of log of Hospital Trusts' Standardised Hospital Mortality on log of Hospital Trusts' Unadjusted or Casemix-Adjusted Average EQ-5D Health Gain. See Figure 1 notes for SHM sources. Average EQ-5D health gains from hip and knee replacement in 2012/13 exclude data from patients undergoing revisions to these procedures; see Figure 1 notes for more. A constant is included but not reported. Standard errors are reported in parentheses. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Site-Level Regressions of Log of AMI Mortality on Log of Average EQ-5D Health Gain

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Variable	Log of Sites' AMI Mortality							
Explanatory Variable Casemix-Adjusted?	No				Yes			
Procedure On Which Explanatory Variable Is Based	Hip	Knee	Groin	Vein	Hip	Knee	Groin	Vein
Log of Sites' Average EQ-5D Health Gain	-0.0570 (0.0965)	-0.0242 (0.0883)	0.0172 (0.0321)	0.0402 (0.0313)	-0.00833 (0.133)	0.0900 (0.124)	0.0396 (0.0517)	-0.0378 (0.0369)
Observations	705	696	818	611	703	698	831	635
R-squared	0.000	0.000	0.000	0.003	0.000	0.001	0.001	0.002

Table reports coefficients from univariate regression of log of Hospital Sites' AMI Mortality on log of Hospital Sites' Unadjusted or Casemix-Adjusted Average EQ-5D Health Gain. See Figure 2 notes for AMI sample restrictions. Average PROMs health gains from hip and knee replacements include both primary procedures and revisions. A constant is included but not reported. Standard errors are reported in parentheses. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Trust-Level Regressions of Log of First Differenced Standardised Hospital Mortality on Log of First Differenced Average EQ-5D Health Gain

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Variable	Log of Trusts' First Differenced Standardised Hospital Mortality							
Explanatory Variable Casemix-Adjusted?	No				Yes			
Procedure On Which Explanatory Variable Is Based	Hip	Knee	Groin	Vein	Hip	Knee	Groin	Vein
Log of Trusts' First Differenced Average EQ-5D Health Gain	0.0498 (0.0879)	0.0934 (0.0968)	-0.101 (0.0978)	0.145 (0.114)	0.0995 (0.118)	-0.0149 (0.0910)	-0.0806 (0.102)	0.158 (0.217)
Observations	133	123	104	83	119	132	95	40
R-squared	0.002	0.008	0.010	0.020	0.006	0.000	0.007	0.014

Table reports coefficients from univariate regression of log of Hospital Trusts' First Differenced Standardised Hospital Mortality on log of Hospital Trusts' First Differenced Unadjusted or Casemix-Adjusted Average EQ-5D Health Gain. See Figure 1 notes for SHM sources. Average EQ-5D health gains from hip and knee replacement in 2012/13 exclude data from patients undergoing revisions to these procedures; see Figure 1 notes for more. A constant is included but not reported. Standard errors are reported in parentheses. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Site-Level Regressions of Log of First Differenced AMI Mortality on Log of First Differenced Average EQ-5D Health Gain

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome Variable	Log of Sites' First Differenced AMI Mortality							
Explanatory Variable Casemix-Adjusted?	No				Yes			
Procedure On Which Explanatory Variable Is Based	Hip	Knee	Groin	Vein	Hip	Knee	Groin	Vein
Log of Sites' First Differenced	0.0403	-0.0475	0.0445	0.0869	0.106	0.0840	0.101	0.121*
Average EQ-5D Health Gain	(0.0785)	(0.0913)	(0.0803)	(0.0948)	(0.0828)	(0.0996)	(0.0762)	(0.0681)
Observations	150	132	155	121	145	139	169	128
R-squared	0.002	0.002	0.002	0.007	0.011	0.005	0.010	0.025

Table reports coefficients from univariate regression of log of Hospital Sites' First Differenced Standardised Hospital Mortality on log of Hospital Sites' First Differenced Unadjusted or Casemix-Adjusted Average EQ-5D Health Gain. See Figure 2 notes for AMI sample restrictions. Average PROMs health gains from hip and knee replacements include both primary procedures and revisions. A constant is included but not reported. Standard errors are reported in parentheses. Statistical significance is reported as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1483	Davide Cantoni Jeremiah Dittmar Noam Yuchtman	Reallocation and Secularization: The Economic Consequences of the Protestant Reformation
1482	David Autor David Dorn Lawrence F. Katz Christina Patterson John Van Reenen	The Fall of the Labor Share and the Rise of Superstar Firms
1481	Irene Sanchez Arjona Ester Faia Gianmarco Ottaviano	International Expansion and Riskiness of Banks
1480	Sascha O. Becker Thiemo Fetzer Dennis Novy	Who voted for Brexit? A Comprehensive District-Level Analysis
1479	Philippe Aghion Nicholas Bloom Brian Lucking Raffaella Sadun John Van Reenen	Turbulence, Firm Decentralization and Growth in Bad Times
1478	Swati Dhingra Hanwei Huang Gianmarco I. P. Ottaviano João Paulo Pessoa Thomas Sampson John Van Reenen	The Costs and Benefits of Leaving the EU: Trade Effects
1477	Marco Bertoni Stephen Gibbons Olmo Silva	What’s in a name? Expectations, heuristics and choice during a period of radical school reform
1476	David Autor David Dorn Lawrence F. Katz Christina Patterson John Van Reenen	Concentrating on the Fall of the Labor Share
1475	Ruben Durante Paolo Pinotti Andrea Tesei	The Political Legacy of Entertainment TV

1474	Jan-Emmanuel De Neve George Ward	Happiness at Work
1473	Diego Battiston Jordi Blanes i Vidal Tom Kirchmaier	Is Distance Dead? Face-to-Face Communication and Productivity in Teams
1472	Sarah Flèche	Teacher Quality, Test Scores and Non-Cognitive Skills: Evidence from Primary School Teachers in the UK
1471	Ester Faia Gianmarco Ottaviano	Global Banking: Risk Taking and Competition
1470	Nicholas Bloom Erik Brynjolfsson Lucia Foster Ron Jarmin Megha Patnaik Itay Saporta-Eksten John Van Reenen	What Drives Differences in Management?
1469	Kalina Manova Zhihong Yu	Multi-Product Firms and Product Quality
1468	Jo Blanden Kirstine Hansen Sandra McNally	Quality in Early Years Settings and Children's School Achievement
1467	Joan Costa-Font Sarah Flèche	Parental Sleep and Employment: Evidence from a British Cohort Study
1466	Jo Blanden Stephen Machin	Home Ownership and Social Mobility
1465	Decio Coviello Stefano Gagliarducci	Tenure in Office and Public Procurement

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE