The hospital as a multi-product firm: Measuring the effect of hospital competition using value-added, procedure-specific indicators of clinical quality*

Matthew Skellern

27th November 2016

Abstract

Most econometric studies of the effect of market structure on hospital quality equate outcomes with final patient 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 Patient Reported Outcome Measures (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 introduction of patient choice of hospital may have had a negative effect on elective surgical 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.

1 Introduction

A defining characteristic of both health care and education is that final outcomes are determined not just by the producer's (hospital's or school's) choice of inputs, but also by characteristics of the consumer (patient or student). The influence of (often poorly observed) consumer characteristics on performance means that final outcomes, such as post-operative health status or exam results, are a problematic indicator of provider quality. While the confounding effect of consumer characteristics has led over the last decade to an increasing focus, in the economics of education literature, on value-added measures of school quality, the health economics literature largely continues to equate hospital outcomes with final patient health status, not the change in health status resulting from treatment, for the understandable reason that value-added measures of health outcomes are much less widely available.

At the same time, standard economic models of hospital competition assume that hospitals produce a single type of output, and choose a single, hospital-wide quality level. The econometric literature on market structure and hospital outcomes largely works within this theoretical framework, by focusing on indicators of hospital performance – such as 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.

In this context, the present paper contributes to two new avenues of investigation concerning the relationship between market structure and hospital performance. First, it argues that the hospital should

*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, 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.

be conceptualised as a multi-product firm, and 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. Second, it makes use of a new value-added indicator of hospital performance that not only captures clinical quality at the level of the individual surgical procedure, but also measures clinical quality based not on final patient outcomes but on patients' health *gains* from surgery, to assess the impact of hospital competition on clinical quality in the English National Health Service (NHS). These new indicators of clinical quality, known as Patient Reported Outcome Measures or PROMs, take the form of two identical questionnaires – one capturing health status before treatment, the other capturing health status after treatment – that are used to construct a measure of health *gain* from treatment. In this respect, PROMs are a fundamental methodological advance on previous indicators of clinical quality, which equate hospital outcomes with post-treatment health status.

Since April 2009, PROMs questionnaires have been distributed to all patients undergoing one of four elective surgical procedures[1] – hip and knee replacement, groin hernia repair and varicose vein stripping – within the English NHS. These PROMs are merged 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.

Standard single-output-type economic models of fixed-price hospital competition (Gaynor 2006; Gaynor & Town 2012) 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. Two previous studies (Cooper et al. 2011; Gaynor et al. 2013) assessed the impact of the 2006 English NHS patient choice reforms on care quality by using mortality-based indicators of hospital-wide quality, and obtained results consistent with the basic theoretical prediction just outlined – hospitals in more competitive markets experienced larger improvements in quality than hospitals in less competitive markets.

By contrast, this paper, using value-added, elective-surgery-specific outcome measures, finds evidence that the hospital competition brought about by patient choice may have led to *lower* clinical quality. Although the estimates reported here are provisional, the best reading is that competition led to *lower* orthopaedic and varicose vein surgery quality, and had no effect on the quality of groin hernia repair surgery. These findings 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 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 one-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.

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

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

The predominant view of the 1990s Internal Market is that it was not successful at providing hospitals with incentives to improve clinical quality (Le Grand et al. 1998). A key reason for this failure was that hospitals were encouraged to compete on price as well as on quality, yet there was virtually no publicly available information during this era about the quality of care. This situation gave hospitals an incentive to compete on price *at the expense of* quality. Propper et al. (2004), using a cross-sectional estimation strategy, found that competition in the Internal Market led to higher mortality rates from acute myocardial infarction (AMI, or heart attack). Propper et al. (2008) used a difference-in-differences (DiD) estimator with hospital fixed effects to identify the effect of the Internal Market, finding that hospitals in the most competitive markets experienced smaller decreases in AMI mortality rates, as well as larger decreases in elective surgery waiting times, than other hospitals. 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).[2]

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, initially by encouraging care purchasers to 'selectively' enter into bulk contracts with providers.

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

[2] 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.

et al. (2011) use a DiD strategy to estimate the effect of the competition generated by patient choice on AMI mortality. By comparing changes in low and high competition areas, they find that hospital competition led to significant reductions in 30-day in-hospital AMI mortality. Gaynor et al. (2013) use a similar DiD-style estimation strategy, finding that the patient choice reforms led to larger reductions 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 possess more hospitals, and hence be more competitive, than other markets. They find that higher competition led to higher 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.[3] 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. (2013) 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 because clinical quality is a hospital-wide variable or has a hospital-wide component and/or (3) this led to quality improvements in other areas of hospital activity because of positive spillovers between elective surgery quality and quality in other areas.

Claim (1), which remains largely unexamined by the existing literature, is the major focus of this paper. One study (Feng et al. 2014) 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 uses seeks to draw causal conclusions by using instrumental variables estimation to address potential sources of endogeneity.

Claim (1) remains controversial. Many health policy analysts are skeptical that competition will improve hospital performance because patients cannot or do not choose a hospital on the basis of clinical quality (Jones & Mays 2009; Fotaki et al. 2008). 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.[4] 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, for most of this four-year period, visitors to the NHS Choices website would have been 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.[5] Notwithstanding these

[3] 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.

[4] 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).

[5] 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

deficiencies in official efforts to disseminate information about hospital quality, patients may have access to information about hospital quality via other channels, such as their General Practitioner. The patient choice reforms may have enabled patients and GPs to act on any private information 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 by generating competition between hospitals for market share is ultimately an empirical question – this is the question that this paper aims to shed some light on.

If claim (2) were true, then the existing econometric assessments of the patient choice reforms, by showing that competition had a positive effect on 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.

Claim (3) 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.[6]

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 complement 'objective' measures of health status (such as blood pressure or limb mobility) to improve their treatment of individual ailments, only recently have policymakers recognised their potential for use in policy evaluation and performance measurement.[7]

Despite it being long understood that health care is only an intermediate good, whose ultimate purpose is to produce a non-market good known as *health* (Grossman 1972; Becker 1964), health care providers around the world have until recently tended to measure their output in terms of the amount of *health care* produced (e.g. number of surgical procedures performed), rather than the amount of

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.

[6] 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.

[7] 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.

health produced (Appleby and Devlin 2010). Initial attempts in the 1980s and 1990s to measure health *outcomes* rather than *outputs* tended to equate health with the absence of sickness, focusing on outcomes such as mortality rates, readmission rates, and complication rates, because these measures could often be derived from administrative data sets.[8] 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, surgical outcomes vary greatly. Mortality-based indicators of hospital performance do not directly capture this variation.[9]

In recognition of the limitations of existing measures of elective surgery quality, in April 2009, the NHS, after conducting a review (Smith et al. 2005) and pilot programme (Browne et al. 2007), started collecting PROMs for four surgical procedures – hip replacement, knee replacement, groin hernia repair, and varicose vein treatment. Patients undergoing any of the four procedures are asked to fill out the generic EQ-5D survey of health-related quality of life (EuroQol Group 1990) both before surgery (either at their pre-surgical assessment, or on admission for surgery), and after surgery (three or six months post-operatively, depending on the procedure undertaken). At the same time, patients for all but one of the procedures (groin hernia repair) are asked to complete 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 centrepiece of the NHS PROMs programme, the EQ-5D survey of health-related quality of life, is very widely used in the UK and Europe. The UK's National Institute for Health and Care Excellence (NICE), which is responsible for approving 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 are asked to choose from one of three options – 1 (no problems), 2 (some problems), or 3 (extreme problems) – giving $3^5 = 243$ possible permutations of response.[10] These 243 possible response profiles are then aggregated with weights obtained from population-level surveys, generating a utility metric of health states, with 1 representing perfect health and 0 representing death.[11] The result is a utility measure that can be interpreted cardinally – for example, the UK value for health state 12331 is 0.07, meaning that 1 year lived in that state is equivalent to 0.07 of a year in perfect health.

[8] Efforts to more effectively measure health *outcomes* – or a health care system's production of *health* – are often spoken about in the policy literature in terms of attempts to develop improved measures of the *quality* of health care provided. The difference is that, while health outcomes can (at least in theory) be directly measured, measuring the quality of health care requires an understanding of the production process whereby health care of a certain *quality* is combined with other inputs (such as time and human capital) to produce health *outcomes*. Backing out health care quality from health outcomes therefore requires one to control for the many other factors that influence health outcomes – this is part of the challenge of casemix adjustment.

[9] 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.

[10] 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.

[11] 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 (HSCIC 2013a, p.30), 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.

The procedure-specific PROMs – OHS, OKS and AVVQ – generally have a greater capacity than the EQ-5D to detect changes in health status resulting from surgery, as they ask condition-specific questions.[12] The OHS and OKS consist of 12 multiple-choice questions each of which confers between 0 and 4 points, resulting in an overall score between 0 (worst) and 48 (best).[13] The AVVQ consists of 13 multiple-choice questions each of which confers a certain number of points, resulting in an overall score between 0 (best) and 100 (worst).[14] In this paper, the scale of the AVVQ is reversed so that, like all other measures examined, higher scores denote better health. Table 2 reports summary statistics for all the outcome variables examined in this paper.

The EQ-5D, OHS, OKS and AVVQ have been extensively validated using standard psychometric tests as tools for capturing health gains from the four surgical procedures included in the NHS PROMs programme (Smith et al. 2005). Table 3 reports the effect sizes – a measure of a PROM's responsiveness to an intervention, equal to the average health gain divided by standard deviation of pre-operative score – of the PROMs studied in this paper. By convention, 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). Table 3 shows that the procedure-specific outcome variables – the OHS, OKS, and AVVQ – do an excellent job of capturing variation in health status, with effect sizes ranging from 0.71 to 2.32. The EQ-5D index score performs moderately well for minor surgical procedures (0.38-0.39) and very well for major surgical procedures (0.93-1.25). By contrast, the EQ-VAS performs moderately well in relation to hip replacement surgery, but poorly in relation to all other procedures. Overall, these effect sizes show that the PROMs collected by the English NHS contain meaningful variation that should be capable of detecting changes in health gain resulting from differential exposure to competition. Given the effect sizes reported in Table 3, this paper designates the procedure-specific PROMs and the EQ-5D as the main outcome variables of interest. Estimates using the EQ-VAS as an outcome variable are reported for robustness.

A good outcome measure should be correlated with other measures that are known to capture outcomes: 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.[15] Table 4 shows that all such correlations are above 0.2 except for that between the AVVQ and the EQ-VAS for varicose veins. There is a particularly strong correlation between the EQ-5D and the Oxford Hip and Knee Scores (0.62 and 0.63 respectively). These correlations provide evidence that the PROMs studied in this paper are capturing a coherent, underlying concept of health gain from surgery.

By asking patients an identical set of questions before and after surgery, PROMs constitute a 'value-added' measure 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; Koedel et al. 2015; Gibbons et al. 2013a; 2013b). As such, they constitute a fundamental methodological advance on previous methods of measuring the impact of health care interventions, in that the pre-surgical questionnaire provides a baseline measure of the patient's health status, thus greatly ameliorating one of the most challenging problems in the literature on hospital competition and quality – namely, disentangling the contribution of the hospital to the patient's post-surgical health status from aspects of pre-surgical health status, or other patient characteristics, that also affect the patient's post-intervention health status, but that are often not observable, or imperfectly observed.

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 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

[12] 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 patient utility, but rather as clinically relevant measures of health gain from surgery.

[13] 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 (HSCIC 2013a).

[14] Due to rounding of the weights used for each question, the maximum AVVQ score is actually 99.658 (HSCIC 2013a).

[15] There is no relationship between hospital mortality ratees and average adjusted PROMs health gains. This is presumably a reflection of the fact that death is an extremely rare outcome of all four PROMs procedures.

impact of hospital competition using mortality-based quality indicators, would have effectively also been capturing 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. 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'[16] casemix-adjusted mortality ratios and average casemix-adjusted PROMs health gains from elective surgery.[17] There does not appear to be any relationship between these variables. Table 5 reports the corresponding correlations for both raw (unadjusted) and risk-adjusted PROMs health gains. If anything, there appears to be a small *positive* correlation between hospital trusts' standardised mortality rates and unadjusted PROMs health gains for orthopaedic surgery. When casemix-adjusted PROMs health gains are used, health gains for several elective procedures appear to be correlated, but there continues to be essentially no correlation between PROMs health gains and standardised mortality. Bivariate regressions of the log of trusts' standardised mortality ratio on the log of average adjusted EQ-5D health gain show no statistically significant relationship between these variables.

I next present data on the relationship between PROMs health gains and AMI mortality at the hospital site level.[18] Whereas there are questions about the capacity of standardised mortality indicators 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 quality of different treatments within a given hospital are correlated with each other, there should be a 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. Again, there appears to be no relationship between these dimensions of hospital performance. Simple bivariate regressions of the log of hospitals' AMI mortality on the log of average EQ-5D health gain show no significant relationship at 5 per cent level for three PROMs procedures, and a marginally significant relationship (p-value 4.9%) for the fourth.[19] Thus, there appears to be little or no relationship between the quality of a hospital's elective surgery and its AMI mortality rate.

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

[16] 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.

[17] From 2010/11 onwards, my measure of risk-adjusted hospital mortality is the official NHS Standardised Hospital Mortality Indicator or SHMI (HSCIC 2013b). 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.

[18] 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.

[19] A 1 per cent increase in adjusted PROMs health gain for knee replacement surgery is associated with a 0.21 per cent decrease in logged AMI mortality. However, given that this relationship is only marginally significant at the 5 per cent level (t-statistic = 1.97), and given also that this is the only one of the 16 bivariate regressions reported in Section Three that shows a significant relationship between mortality and PROMs health gain, a fairly low weight should arguably be given to this finding of significance.

change in AMI mortality.[20] Neither table indicates a correlation between first differenced mortality rates and first differenced adjusted PROMs health gains, and simple bivariate regressions in logs confirm no statistically significant relationships at the 5 per cent level.[21]

The apparent lack of any 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).[22] 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 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 offer a clear prediction – 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 ($x = 2$) – with associated quality levels z_{j1} and z_{j2}.[23] 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

[20] 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, we define a financial year as running from 1 April until 31 March. All references to years in this paper refer to financial years.

[21] 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.

[22] 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.

[23] 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.

competitors. Increased hospital competition is represented in the model as an increase in N.[24] 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; if $\frac{\partial c_x}{\partial z_{-x}} > 0$ and $\frac{\partial^2 c_x}{\partial z_x \partial z_{-x}} > 0$, they are cost substitutes.[25]

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.[26] 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:

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

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

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

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 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 ($\frac{\partial^2 c_1}{\partial z_1 \partial z_2} < 0$), 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 ($\frac{\partial^2 c_1}{\partial z_1 \partial z_2} > 0$), the increase in z_1 leads to higher marginal costs, implying a *decrease* in z_2. Thus, increased competition leads to higher emergency

[24] 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).

[25] 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.

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

care quality if the two outputs are cost complements, but lower emergency care quality if the two outputs are cost substitutes.[27]

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:

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

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 :

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

These first order conditions imply that increasing competition in elective surgery leads to higher emergency care quality but, perversely, to *lower* elective surgery quality if the two types of hospital output are cost substitutes.[28] This possibility provides a further argument in support of looking at the effect of the 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

PROMs surveys have only been collected within the English NHS since April 2009 – *after* the introduction of patient choice of hospital for elective surgery. The lack of pre-reform PROMs data means that it is not possible to employ a DiD style estimation strategy of the kind used by 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

[27] 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.

[28] 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.

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[29] at hospital site j and year t:

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

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 ε_{ijpt} is an error term.[30] $\boldsymbol{X}_{ijpt}$ 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 5.6. To account for serial correlation, 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.[31] 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.[32] A separate HHI is calculated for each financial year for each of six high-volume elective surgical procedures: the four PROMs procedures (hip and knee replacement, groin hernia repair, and varicose vein stripping), plus knee arthroscopy and cataract repair, using a definition of market size that is discussed 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 type of endogeneity in the literature on hospital competition and quality is casemix bias – that is, omitted variable bias due to heterogeneity in unobserved patient characteristics and health status. A hospital with poor average health outcomes may be providing poor quality of care, or its patients may simply start out sicker, on average, than those who attend other hospitals. Unobserved casemix differences can lead to bias via (at least) two channels: the first via the confounding role of patient choices (selection bias), and the second via association between unobserved patient health status and the geographical correlates of competition intensity such as poverty and urbanness. A third potential type of endogeneity is reverse causality arising from the influence of patient choices on one's measure of competition intensity. The next section outlines how I use instrumental variables to address the two types of bias arising from patient choices (selection bias and reverse causality). After that, I show how using value-added measures of health outcomes can address endogeneity due to the geographically-based correlation between unobserved patient health status and competition intensity.

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

When patients (in consultation with their primary care physician) choose a hospital, their choices may be systematically influenced by patient-level characteristics that influence outcomes. For example, if sicker

[29] Hip and knee replacement observations are analysed together because they are both performed by a hospital's orthopaedic department. In all orthopaedic surgery regressions, a dummy variable for knee replacement is included, to capture level differences in health gains between the two surgical procedures. Estimates from running regressions separately for hip and knee replacement are available on the author's website at: http://personal.lse.ac.uk/skellern.

[30] Total and procedure-specific hospital admissions may be influenced by hospital quality after the introduction of patient choice; this issue is discussed in Section 5.6.

[31] The competition indices and hospital-level averages used in this paper are calculated using all observations in the dataset. However, when running the regressions, any patients that attended a hospital that treated fewer than 50 patients for the procedure and year in question are omitted. Also omitted are a small number of additional observations that satisfied this restriction, but for whom there were fewer than 5 PROMs surveys for the hospital and year in question.

[32] 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.

patients select into attending higher-quality hospitals because they have more at stake from surgery, then the observed distribution of hospital quality will be distorted relative to the true distribution. In classical settings, where post-treatment health status is the outcome variable, this form of patient selection would lead to a compression or reversal of the distribution of observed hospital qualities 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. This problem of selection bias can be understood as a form of omitted variable bias because, if all relevant patient characteristics were observable, the problem could be eliminated by controlling for these patient characteristics in one's regressions.

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 constructing a measure of competition intensity is how to define the size of the market within which a given hospital operates. A commonly used approach is to centre markets on hospitals themselves and to assume that hospital j's market includes all hospitals within the radius required to encompass the home address of a certain percentage (such as 75% or 95%) of j's patients. The problem with these 'variable radius' methods of defining market size is that, when patients can choose which hospital they attend, percentiles of patient distance travelled will in general be 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. 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 (including two instrumentation 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' neighbourhood 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.[33] Competition measures that centre hospital markets on an exogenously defined statistical unit corresponding to a 'neighbourhood' will still be influenced by patient choices, but are unlikely to be biased by reverse causality because it is hard to think of reasons why the direction of any such influence would be systematic, as it would be with a variable radius definition of market size if, for instance, patients travel further to attend higher quality hospitals. To calculate this index, for each of the six elective surgical procedure listed above, I first calculate an MSOA-level HHI equal to the sum of each provider's squared market shares in the MSOA. I then create a procedure-specific, hospital-year-level HHI equal to the weighted average of the HHIs of all the MSOAs that it serves. Finally, a single HHI for each hospital and year is then calculated as a weighted average of the procedure-specific HHIs.

Secondly, as well as using current-period measures of competition intensity, this paper also 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.[34] 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 even when a neighbourhood-centred definition of hospital markets is used.

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 alternative j be an additive function of a systematic component $\boldsymbol{x}_i'\boldsymbol{\theta}_j + \boldsymbol{z}_{ij}'\boldsymbol{\gamma}$ and a random component ε_{ij}. The vector $\boldsymbol{x}_i$ – with 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.

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

[33] 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.

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

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:

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

If it is assumed that ε_{ij} is distributed standard Type 1 extreme value with cumulative distribution function $F(\varepsilon) = e^{-e^{-\varepsilon}}$, then it can be shown (Maddala 1983) that:

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

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 fo each individual, are used to calculate predicted market shares for each hospital. An MSOA-level HHI is then calculated, and a hospital-level HHI is then calculated as a weighted average of the MSOA-level HHIs in the manner described earlier. Further details are provided in Appendix 2.

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.[35] 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 that the HHI is influenced by hospital quality even when instrumented by its average pre-reform level.

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 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, 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.

[35] 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. All variables used to predict patient choice of hospital in the first stage are included in the second stage regression of elective surgery quality on intensity of competition. As computational feasibility limits the number of variables that can be included in the conditional logit model in the first stage, this means that my second stage regressions using a predicted patient choice HHI include a smaller number of control variables than those using an HHI based on actual patient choices. Partly for this reason, in reporting our main results we continue to include estimates using actual patient choice HHIs as well as those using predicted patient choice HHIs.

Performing instrumental variables estimation in two separate stages will lead to incorrect standard errors in the second stage, as the standard errors for 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.

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. It is to this problem that we now turn.

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-defined nature of hospital markets – yet there are many correlates of geography, including, most importantly, patient health status, that may also influence outcomes.[36] These correlates of geography will lead to omitted variable bias if not adequately controlled for. For instance, 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.

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. 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. Therefore 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 be able to identify a causal effect. In the present study, it is *theoretically* possible to estimate the causal effect of competition on elective surgery quality in this fashion – but 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 intensity at the hospital level to have generated sufficient variation in within-hospital 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 treatment* rather than with *post-treatment health status*) provides a much more powerful way of controlling for the correlates of geography. 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 would arise if hospital fixed effects were 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 replace with another, more subtle source of potential bias if unobserved components of pre-treatment health status

[36] 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).

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.

It is certainly the case that *observably* sicker patients will have higher health gain from surgery as measured by PROMs. If for no other reason, this is guaranteed by ceiling effects arising from the bounded nature of PROMs scores – a patient who reports close to perfect health before treatment is simply 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, it is possible to remove this influence of pre-treatment health status on outcomes.[37] 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 for the EQ-5D and Oxford Hip Score PROMs. Panels (A) and (C) show how unadjusted health gains from surgery are a direct function of pre-treatment health status, while panels (B) and (D) show how the risk adjustment process eliminates this relationship.[38]

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 it is implausible that aspects of pre-treatment health status relevant to post-treatment health status, as captured by PROMs surveys, would not be captured by the pre-treatment survey, given that it asks exactly the same set of questions. Fundamentally, this is an argument that surveying patients on pre-treatment health status, using an identical questionnaire to that used to capture post-treatment health status, entails a radical redefinition of the boundary between 'unobserved' and 'observed' health status, in favour of the latter. Therefore, estimating the effect of market structure on hospital performance using value-added outcome measures such as risk-adjusted PROMs health gains not only eliminates the problem of casemix bias due to the unobserved correlates of geography, but represents an advance on conventional difference-in-differences methods by enabling one to control for pre-treatment health status at the individual patient level rather than the hospital level, and by eliminating any bias that arises when hospital fixed effects are employed to control for any time-varying unobserved components of casemix.

5.4 Robustness test: GP-centred markets

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 rather than hospitals. 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

[37] When publishing hospital-level PROMs scores, the Health and Social Care Information Centre (HSCIC) provides casemix-adjusted health gains from surgery, in addition to unadjusted outcomes. The HSCIC 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, and provides a list of the variables used.

[38] It should be clarified that 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.

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.

In addition to the GP-level HHI just described, a second alternative measure of competition intensity is constructed by performing a final stage of aggregation not undertaken by Cooper et al. (2011), calculating a hospital-level HHI that is equal to the weighted sum of the GP-level HHIs of its patients' GP surgeries. While both the GP-level, GP-centred HHI and the hospital-level, GP-centred HHI are valid ways of measuring competition intensity, the latter has the intuitively appealing property that treatment intensity is the same for every patient attending a given hospital in a given year.

5.5 Correlation between competition intensity measures

Table 9 reports the correlation between the competition indices used in this paper and a dummy variable indicating that the patient lives in an urban area. In addition to the competition indices already described, I also calculate a measure similar to that used by Propper et al. (2004), in which intensity of competition experienced by hospital j is equal to the number of hospitals within 30km of j. 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.

Several conclusions can be drawn from Table 9. First, urban status is essentially uncorrelated with the HHIs based on actual patient choices, suggesting that these competition measures do an excellent job of controlling for rural-urban differences. This lack of correlation suggests that, to the extent that concerns about the unobserved correlates of geography are mainly concerns about rural-urban differences, the HHIs based on actual patient choices may already be delivering variation in competition intensity that is free from this form of omitted variable bias, even without resort to other strategies such as value-added outcome measures or hospital fixed effects.

Secondly, there is a moderate correlation between the predicted patient choice HHI, urban status, and the fixed-radius measure. This reflects the influence of urban-rural differences on the fixed radius measure, and the use of distance-to-hospital-based variables as the main means of satisfying the exclusion restriction when calculating the predicted patient choice HHI. The fact that predicted patient choice HHIs are correlated with urbanness in a way that actual patient choice HHIs are not suggests that the success of predicted patient choice HHIs at controlling for reverse causality and patient selection comes at the expense of increasing the risk of omitted variable bias arising from the unobserved correlates of urbanness. However, as argued in Section 2.2, the use of a risk-adjusted, value-added indicator of hospital performance addresses this risk, and arguably in a more convincing way than the hitherto dominant approach of including hospital fixed effects in one's model.

Thirdly, the level to which a competition measure is aggregated matters a lot. The hospital-level, GP-centred HHI is just a weighted average of the GP-level, GP-centred HHIs of its patients' GP surgeries, yet the correlation between these two indices is only 0.448. By contrast, the neighbourhood-centred HHI based on actual patient choices and the GP-centred, hospital-level HHI have a correlation above 0.8. This high correlation appears related to the fact that both indices are calculated at the hospital level. This high correlation suggests that these two competition indices are capturing a similar underlying concept, even though they are constructed in very different ways.

Fourthly, and in keeping with the existing literature (e.g. Cooper et al. 2011), the neighbourhood-centred HHI based on predicted patient choices have a correlation of around 0.4 with the hospital-level HHIs based on actual patient choices. Somewhat puzzlingly, the neighbourhood-centred HHI based on predicted patient choices has a higher correlation with the GP-centred, hospital-level HHI than with the neighbourhood-centred HHI based on actual choices.

5.6 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 (approximately 125 million observations per year), and the PROMs survey responses by individual patients. NHS patients can be treated either in NHS (public) hospitals, or in private hospitals that are registered to accept NHS patients.[39] No distinction is made between NHS and private providers when calculating competition indices or when running regressions, except to include a dummy variable in the latter indicating whether a hospital is privately run. The HES dataset used in this study encompasses eleven full 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.[40] Finally, the dataset includes all non-elective AMI admissions. In total there are 8.6 million observations in the dataset.

NHS hospital trusts often consist of multiple hospital sites which can be located up to 100km from each other. Individual hospital sites within a hospital trust are for many purposes run independently, although their finances are managed at the trust level. As individual hospital sites within a given trust can act as effective competitors with each other, analysis is conducted at the hospital site level rather than trust level.[41]

The main part of the dataset consists of all admissions for PROMs procedures from 2009/10, the first year in which PROMs surveys were collected, through to the end of 2012/13. There are 1,261,134 observations in this main part of the dataset, although not all of these patients will have been eligible to receive a PROMs survey, as 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. Appendix 3 provides the procedure and diagnosis codes used in constructing the dataset.

The PROMs dataset contains 673,584 survey responses, of which 485,711 – or 72.1 per cent – contain the epikey field which allows the record to be linked to HES.[42] Of these, 468,578 – or 96.5 per cent – were successfully matched to the HES dataset. When calculating measures of competition intensity, this paper makes use of the full HES dataset. However, the casemix adjustment and main regressions use only the PROMs records that could be linked to HES. Table 10 breaks down the number of survey responses by procedure and linkage status.

The regressions performed in this study include a rich set of patient-, hospital- and region-level controls. Patient level controls include dummies for gender crossed with age (in five-year intervals), and dummies for the day and month of the operation date (to control for day-of-week and seasonality effects). They also include a dummy indicating whether the patient lives in an urban area, a dummy indicating whether the patient was treated as a day case, two dummies capturing 'low' and 'high' severity (respectively, just one diagnosis field completed, and three or more diagnosis fields completed), and a control for 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 for the Index of Multiple Deprivations income deprivation score, which measures the percentage of households that are income deprived in the patient's Lower Super Output Area (LSOA) of residence, is included. As poverty is associated with poor health, including this last variable may control for unobserved dimensions of health status that influence outcomes.

At the hospital level, dummies are included that indicate whether the patient's site of treatment is part of a specialist trust, a teaching trust, a university trust, a standard acute trust, or a private

[39] 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.

[40] 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.

[41] 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.

[42] This number also omits around 4,000 observations that appear to be duplicates.

provider.[43] Care quality tends to be higher, and costs lower, in larger health care markets, and in larger hospitals. These volume (or scale) effects can lead to upward-biased estimates of the effect of quality on competition if areas with larger markets and hospitals also have higher competition intensity. For this reason, a control for population density in the hospital's catchment area, as well as year-specific controls for the site's total number of patients for the procedure in question, are included, as well as for the trust's total number of admissions for all causes. A quadratic term is included for both total and procedure-specific admissions, to control for possible non-linearities in scale effects. As total and procedure-specific admissions may be influenced by hospital quality in the period after the introduction of patient choice, in the baseline specification lagged values of these variables (specifically, their average values over the three years from 2002/3 to 2004/5) are used in place of current-period values.

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 possible other region-level trends in correlates of health outcomes. Table 11 provides average values for key control variables used in this paper.[44]

6 Results

6.1 Main estimates

Table 12 reports estimates using the paper's main competition indices – neighbourhood-centred HHIs based on actual or predicted patient choices. Columns (1) and (2) report cross-sectional estimates using current-period (financial year) HHI as the treatment intensity variable, and current-period values of total and procedure-specific admissions (and their respective quadratics) as controls. Columns (3) to (6) report estimates using the paper's preferred specification, in which current-period HHI is instrumented by its pre-reform average value, and pre-reform averages are used as proxies for total and procedure-specific controls – Columns (3) and (4) report first stage estimates (the coefficient on the excluded variable, pre-reform average HHI, when regressed on current-period HHI), while Columns (5) and (6) report second stage estimates (the coefficient on the fitted value of current-period HHI when regressed on PROMs health gain). The first stage estimates reported in Columns (3) and (4) are very strong in all cases. This also means that the cross-sectional estimates of treatment effects tend to be quite similar to the second stage estimates from IV estimation. For this reason, throughout Section 2.2, cross-sectional estimates are not discussed unless they deviate notably from the second stage estimates which are of greater interest.

Columns (5) and (6) indicate a negative effect of competition on orthopaedic surgery quality significant at the 5 per cent level in three out of the four competition index – outcome variable combinations. The two significant estimates for the OHS/OKS are very similar in magnitude when normalised by the standard deviation of each competition index – a one standard deviation increase in competition intensity leads to a decrease in health gain from orthopaedic surgery of between 0.256 and 0.394 points out of a maximum of 48. These two point estimates easily lie within each others' 95 per cent confidence intervals. The estimates for groin hernia repair are not statistically significant. When an HHI based on predicted patient choices is used, competition has a negative effect on the quality of varicose vein surgery significant at the 1 per cent level for the AVVQ and the 10 per cent level for the EQ-5D. Finally, using the EQ-5D PROM and pooling all procedures together, with procedure dummies to control for level differences in health gains, competition leads to lower health gains from elective surgery significant at the 10 per cent level, irrespective of which competition index is used.

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 leads to a decrease in health gain of 0.00632 from orthopaedic surgery, and of 0.00644 from varicose vein surgery. An average

[43] A dummy denoting whether a hospital is a Foundation Trust (FT) is not included, as granting of FT status is endogenous to hospital performance. By contrast, all of the hospital type dummies included in the regressions reflect historically determined (and therefore exogenous) hospital characteristics. Given the absence of an FT dummy, the dummy denoting 'standard acute hospital' encompasses both NHS Foundation Trusts and NHS acute trusts that are not Foundation Trusts.

[44] For the regressions using predicted patient flow HHIs, a more limited subset of controls is used, as any control variables included in the second stage should also be included in the first stage – but for tractability reasons only a limited number of predictors of patient choice can be included in the conditional logit model.

orthopaedic surgery patient experiences an increase in health status from 0.376 before surgery to 0.733 after surgery. If their hospital experienced a one standard deviation increase in competition intensity, their post-surgical health status would be only 0.727. Thus, they would be indifferent between one year in their post-operative heath state and 0.727 years in perfect health, as opposed to 0.733 years before the increase in competition.

In like manner, an average varicose vein patient experiences an increase in health status from 0.747 before surgery to 0.852 after surgery. If their hospital experienced a one standard deviation increase in competition intensity, their post-surgical health status would be only 0.846. Thus, they would be indifferent between one year in their post-operative heath state and 0.846 years in perfect health, as opposed to 0.852 years before the increase in competition. These impacts of competition on health gains from elective surgery are small, but nonetheless distinguishable from zero.

In summary, the headline estimates suggest a negative effect of competition on the quality of both orthopaedic surgery and varicose vein surgery significant at the 5 per cent – at times the 1 per cent – level. There is no evidence of a positive or negative effect of competition on quality of groin hernia repair surgery. Pooling all procedures, there appears to be a negative effect on overall elective surgery quality significant at the 10 per cent level.

6.2 Robustness tests

Table 13, which is laid out identically to Table 12, reports estimates using the EQ-VAS PROM. As expected given the EQ-VAS's smaller effect sizes reported in Table 9, these estimates are less statistically significant than those reported in Table 12. Nonetheless, one of the two instrumented competition indices indicates a negative effect of competition on EQ-VAS health gains for orthopaedic surgery, varicose vein stripping surgery, and across all procedures. Cross-sectional estimates using an HHI based on actual patient choices indicate a negative effect of competition on EQ-VAS health gains significant at the 5 per cent level for both orthopaedic surgery and across all procedures.

Table 14 reports second-stage estimates using alternative specifications but the same outcome variables and competition indices as in Table 12, and continuing to instrument current-period competition intensity by its average pre-reform value. In the main specification reported in Table 12, dummies for region of England crossed with year are included. One might be concerned that these dummies soak up too much variation, and that a more parsimonious set of controls might yield more statistically significant treatment effects. Column (1) reports estimates when separate year dummies and region of England dummies are included. The results are largely unchanged. One possible explanation for the insignificance of many of the estimates is that including controls for urban status and catchment area population density 'over-controls' for variables that may be highly collinear with competition intensity. Column (2) reports estimates when these variables are not included. The results are largely unchanged. 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. Column (3) reports estimates when all London hospitals are excluded. The results are largely unchanged.

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. Column (4) reports estimates using an alternative approach, namely instrumenting current-period values of these variables with their pre-reform averages. They are very similar to the Table 12 results. Column (5) reports estimates when total and procedure-level admissions, and their respective quadratics, are omitted from the regressions altogether. The estimates for orthopaedic surgery using the EQ-5D PROM are no longer significant, and the estimates using the OHS/OKS are now only significant at the 10 per cent level. Also, there is now a significant negative effect of competition on groin hernia surgery quality. These changes relative to the headline estimates confirm that hospital scale effects are an important driver of outcomes, and need to be controlled for. Column (6) reports estimates using a log specification, in which any non-dummy variable is replaced with its logged value, with variables scaled up where necessary to avoid taking logs of zero or a negative number. The results are qualitatively unchanged.

Finally, column (7) reports estimates using procedure-specific competition indices, instead of indices constructed by taking the weighted average of procedure-specific HHIs for six high-volume elective

surgical procedures. The use of weighted average HHIs captures the idea that hospital managers do not perceive themselves as being exposed to different levels of competition intensity for different elective procedures, but rather perceive a single level of competition intensity within their sphere of operation, which can be captured by calculating the average competition intensity across a range of surgical procedures. By contrast, the estimates reported in column (7) are motivated by the idea that hospitals do experience competitive pressure that is differentiated by surgical procedure. The results are little changed by this alternative specification. This is unsurprising, as the procedure-level HHIs for a given hospital/year combination are generally highly correlated.

In three of the seven alternative specifications reported in Table 14, there is a *positive* effect of competition on varicose vein surgery quality as captured by the EQ-5D PROM when using an HHI based on actual patient choices – but all three are only significant at the 10 per cent level. Section 7 considers the overall weight that should be given to these findings.

Table 15 reports estimates using alternative competition indices based on GP-centred markets. The table is laid out identically to Table 12, with Columns (1), (3) and (5) reporting estimates using a hospital-level, GP-centred HHI, and Columns (2), (4) and (6) reporting estimates using a GP-level, GP-centred HHI. Columns (1) and (2) report estimates using a current-period competition index, while Columns (3) to (6) report estimates when instrumenting current-period competition with its average pre-reform value. As with the neighbourhood-centred competition measures, the first stage is very strong in all cases. The estimates using the hospital-level, GP-centred competition index are similar to those using the neighbourhood-centred HHIs, albeit less statistically significant. The second stage of the IV estimation indicates a negative effect of competition on orthopaedic surgery quality as captured by OHS/OKS health gains, significant at the 5 per cent level. No other IV estimates are significant, but cross-sectional estimates indicate a negative effect of competition on varicose vein surgery as captured by AVVQ health gains, significant at the 5 per cent level. The cross sectional estimates also indicate a positive effect of competition on groin hernia repair quality significant at the 10 per cent level – but little weight should be given to this estimate, as it is not reflected in any of the other specifications. Pooling all procedures, the cross-sectional estimates also indicate a negative effect of competition on overall elective surgery quality significant at the 10 per cent level. Overall, the results of the regressions using the hospital-level, GP-centred HHI provide support the headline results using neighbourhood-centred HHIs reported in Table 12.[45]

The estimates using the GP-level, GP-centred HHI are, on the other hand, quite different. None of the estimates using current-period HHI or the pre-reform HHI instrument are significant, with the exception of the regression of EQ-5D health gain pooling across all procedures on current-period HHI, which indicates a negative effect of competition on overall elective surgery quality significant at the 10 per cent level.[46]

Finally, Table 16 reports two additional robustness tests. Column (1) reports estimates when intensity of competition experienced by hospital j is set equal to the number of hospitals within a 30km radius of j. Using this index, competition continues to have a negative effect on varicose vein surgery quality as captured by the AVVQ, significant at the 5 per cent level. The point estimates of the effect of competition on orthopaedic surgery continues to be negative, and is close to being statistically significant at conventional levels (p-value for OHS/OKS 11.2 per cent). Column (2) reports estimates using the MSOA-centred, hospital-level predicted patient choice HHI used for our main estimates, but including a hospital fixed effect. As discussed in Section 5.3, the predicted HHI measure should in theory allow identification of the causal effect of competition in a model with hospital fixed effects (that is, using only within-hospital variation in year-on-year competition intensity) – but the four years encompassed by this paper's dataset are unlikely to contain sufficient exogenous variation in within-hospital competition intensity to allow for statistically significant estimates to be obtained. The average within-hospital

[45] The full set of alternative specifications reported in Table 13 was also run using the GP-centred, hospital-level HHI. They show that the estimates for this competition index reported in Table 15 are robust to these alternative specifications. These estimates are available in a Web Appendix available from the author's website: http://personal.lse.ac.uk/skellern.

[46] Estimates using a range of other competition indices are provided in a Web Appendix available from the author's website: http://personal.lse.ac.uk/skellern. This Web Appendix also reports estimates using an alternative instrument for competition intensity based on the historically determined location of NHS hospitals, as used by Cooper et al. (2011). This instrument yields different estimates to those reported in this paper. I do not report them here because, as discussed in the Web Appendix, the results are sensitive to specification, and there is reason to believe that, in the dataset used for this paper, this historical location instrument is not exogenous with respect to rural/urban differences.

standard deviation in the predicted HHI is only 0.1735, as compared with a between-hospital standard deviation in average predicted HHI of 0.8506. To this extent, it is not surprising that the estimates with predicted HHIs and hospital fixed effects do not find any significant effects of competition on elective surgery quality. That said, when hospital fixed effects are included in the regressions, the signs of the estimates continue to be largely consistent with this paper's findings of a negative effect of competition on orthopaedic surgery and varicose vein surgery quality, and are close to being statistically significant at conventional levels for the EQ-5D in relation to orthopaedic surgery (p-value 12.1 per cent).

7 Discussion and conclusions

7.1 Discussion

Overall, the evidence seems to indicate that hospital competition had a negative impact on quality of orthopaedic surgery and varicose vein surgery, and no impact on the quality of groin hernia repair. For orthopaedic surgery, the estimates indicate that competition had a negative effect on quality as captured by the OHS/OKS, and also, in some cases, as captured by the EQ-5D and EQ-VAS. The findings in relation to orthopaedic surgery are robust to a wide range of specifications.[47]

A negative effect of competition on varicose vein surgery quality is found when using a neighbourhood-centred HHI based on predicted patient choices for all three PROMs (EQ-5D, AVVQ and EQ-VAS), using both cross-sectional and instrumental variables estimators. There is some other evidence in support of this finding when using other competition indices. Notwithstanding these findings, caution is needed, for three reasons. The first is that three of the robustness tests using the neighbourhood-centred HHI based on actual patient choices indicate a *positive* effect of competition on varicose vein surgery quality (as captured by the EQ-5D PROM) significant at the 10 per cent level. I am inclined to disregard these findings, given that they are never replicated using one of the headline specifications, and given also the much stronger evidence (using both the EQ-5D and the AVVQ) indicating a negative effect of clinical quality for this surgical procedure. A second reason for caution is that there are only 41,734 varicose veins observations in the dataset, in contrast with 324,173 orthopaedic surgery observations. While this relatively small sample makes the finding of statistical significance in a sense even more noteworthy, on the other hand, it does also mean that the policy significance of this finding is somewhat less than if a similarly robust finding had been made in relation to orthopaedic surgery. A third reason for caution is that varicose vein surgery is also conducted on an outpatient basis, and all such outpatient procedures are not included in the dataset – although it is hard to think of reasons why this selection of the sample of varicose vein patients might have led to biased estimates.

There does not appear to have been any effect of competition on the quality of groin hernia repair surgery as captured by PROMs. The lack of any significant findings (bar one positive estimate at the 10 per cent level in the cross-sectional estimates, and one negative estimate at the 5 per cent level in the robustness tests) likely reflects, in part, the lack of a procedure-specific PROM, and the relatively poor ability of the EQ-5D to capture health gains from this surgical intervention (as reflected in the low effect size reported in Table 9).

Two questions naturally arise in response to the above findings. Why does competition appear to have had a negative effect on the quality of orthopaedic surgery and varicose vein surgery? 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 finding of a negative effect of competition on the quality of orthopaedic surgery

[47] Estimates for hip and knee replacement separately are reported in a Web Appendix available from the author's website: http://personal.lse.ac.uk/skellern. These estimates show that the statistical significance of this paper's orthopaedic surgery estimates are driven mainly by knee replacement observations, the estimates for which indicate a highly statistically significant negative effect of competition on quality. The estimates for hip replacement are, while always negative in sign, generally around one third the magnitude of those for knee replacement, and are never statistically significant at conventional levels. Feng et al. (2014), using English NHS hip replacement PROMs data from 2010/11, find no statistically significant association between intensity of competition and health gains from surgery as captured by PROMs. The results reported in the Web Appendix therefore confirm that, when I estimate the causal effect of competition using only hip replacement patients, I obtain similar results to those of Feng et al. (2014).

and varicose vein surgery. 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 supposedly relevant information about 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 that could explain both the findings of the existing literature, and the findings of this paper, 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 their mortality rates instead of on – and perhaps at the expense of – improving their 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 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 of correlation between a hospital's emergency care quality (as captured by mortality rates) and its elective surgery quality (as captured by PROMs health gains), and put forward a theoretical model that takes into account interactions between production in emergency care and elective surgery.

In contrast to the existing literature, this paper finds that, when value-added, elective-surgery-specific outcome measures are used – Patient Reported Outcome Measures of health gain from four high-volume elective surgical procedures – 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. Although a number of caveats to these findings are noted, the very fact that they deviate substantially from those of the existing literature suggests the merit of looking again at this important policy reform using alternative outcome measures.[48]

This paper's findings call into question a common interpretation of the existing econometric literature on the English NHS patient choice reforms (Cooper et al. 2011; Gaynor et al. 2013; Bloom et al.

[48] This paper's conceptualisation of the hospital as a multi-product firm is echoed by a similar recent study using US Medicare data (Colla et al. 2016), which examines the effect of hospital competition on outcome measures specific to heart attack, orthopaedic surgery, and dementia patients. Colla et al. (2016) examine the hypothesis that hospital responses to competition may be differentiated by output type depending on differences in profitability and quality elasticities of demand. They do not, however, make use of value-added indicators of hospital quality, nor do they consider the possibility that hospital responses to competition across areas of activity may be linked via complementarities and substitutibilities between output types in production. In spite of these differences with the present study, they find, as I do, that the impacts of hospital competition are differentiated by output type, and that AMI mortality is not a good proxy for overall hospital quality.

2015) – namely that it shows that hospital competition leads to across-the-board improvements in care quality. However, this paper's findings do *not* not offer support for one commonly heard criticism of this literature – 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 had been suggested by the existing literature, in the sense that hospital responses to the competitive environment may not be the same across all areas of activity, but may rather be differentiated by area of activity, depending on the observability of performance in each area, the extent to which performance is being 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 different and more complex 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.

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 and note 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 findings reported in the present study suggest that a similar point can be made about the commencement, in 2006, of a new era of hospital competition driven by patient choice of hospital for elective surgery. To be sure, vastly larger amounts of data about hospital performance were 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 national collection of PROMs surveys in 2009/10, several years after the inauguration of a new era of provider competition whose success was premised on the existence of meaningful signals of clinical quality. Even then, PROMs were published on the back alley of an NHS website of data sources, and labelled as "experimental statistics". Given this situation, it is perhaps not surprising that NHS hospitals appear to have responded to a 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, even though these lower-profile areas had a direct impact on health outcomes in elective surgery.

References

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.

Becker, Gary (1964), *Human Capital*, Columbia University Press, New York.

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.

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] (2008), *Guidance on the routine collection of Patient Reported Outcome Measures (PROMs)*, 8 December, Department of Health, London, available from

http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_092625.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.

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), 'Competition in Health Care Markets, Handbook of Health Economics' Vol. 2 Chapter 9, 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.

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.

HSCIC [Health and Social Care Information Centre] (2013a), 'Provisional Monthly Patient Reported Outcome Measures (PROMs) in England: A guide to PROMs methodology', Version 5, 18 March, Health and Social Care Information Centre.

HSCIC [Health and Social Care Information Centre] (2013b), '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.

HSCIC [Health and Social Care Information Centre] (2011), *Finalised Patient Reported Outcome Measures (PROMs) in England, April 2009 – March 2010: Pre- and post-operative data: Experimental statistics*, August, available from
https://catalogue.ic.nhs.uk/publications/hospital/proms/proms-eng-apr-09-mar-10-final/fina-prom-data-2009-10-rep.pdf, accessed 16 August 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.1104-1110.

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.

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

Le Grand Julian, Nicholas Mays and Jo-Ann Mulligan (1998), *Learning from the NHS Internal Market*, King's Fund, London.

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.

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, 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, 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

A1.1 The casemix adjustment methodology

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. This methodology is somewhat simpler than the NHS methodology, as aggregation to the provider level is not required. The current NHS casemix adjustment methodology (DH 2012a; b; c; d; e) involves the estimation of a GLS model with provider fixed effects:

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

In this equation, $Q2_i$ and $Q1_i$ denote the post-operative and pre-operative survey score respectively 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;[49] u_j is a hospital fixed effect; and ε_{ij} is an error term. 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 below), and then re-running the regression using only those variables which are significant at the 5 per cent level.

The first step to deriving *adjusted post-operative health status* and *adjusted health gain from surgery* from this regression is to calculate $\widehat{Q2_i}$, the fitted $Q2_i$, which is defined in the usual way:

$$\widehat{Q2_i} = \widehat{\alpha} + \widehat{\beta_1} Q1_i + \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}_2} + \boldsymbol{z}_{ij}'\widehat{\boldsymbol{\beta}_3} + \widehat{u}_j \tag{6}$$

Secondly, the 'predicted' $Q2_i$, $\widetilde{Q2_i}$, is then 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 $\overline{\overline{u}}$, the (weighted) mean value of the hospital fixed effect.[50] That is:

$$\widetilde{Q2_i} = \widehat{Q2_i} - \widehat{u}_j + \overline{\overline{u}} = \widehat{\alpha} + \widehat{\beta_1} Q1_i + \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}_2} + \boldsymbol{z}_{ij}'\widehat{\boldsymbol{\beta}_3} + \overline{\overline{u}} \tag{7}$$

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:

$$RPF_i = \frac{Q2_i}{\widetilde{Q2_i}} \tag{8}$$

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:[51]

$$\begin{aligned} Q2a_i &= RPF_i \cdot \overline{Q2} & (9)\\ \Delta Qa_i &= Q2a_i - \overline{Q1} & (10) \end{aligned}$$

A1.2 Criticisms of the casemix adjustment methodology

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), as all PROMs can only take a restricted range of values. Recent work using limited

[49] 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.

[50] 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 $\overline{\overline{u}} = \sum_{\forall j} \frac{n_j}{N} \hat{u}_j$

[51] 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.

dependent variable models has sought to address this concern (Hernández et al. 2012; Basu and Manca 2012; Gutacker et al. 2013). However, 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 hospital characteristics and individual characteristics (including pre-operative score) made such an outcome very unlikely, the patient *should*, arguably, receive a score greater than 1, even though this implies a conceptually problematic 'more than perfect health'.

Table 15 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.

A1.3 Variables used in NHS casemix adjustment methodology

Variables taken from PROMs survey responses

The $Q1$ score of the outcome measure being adjusted.

The square of the $Q1$ score of the outcome measure being adjusted.

Female dummy.

$Q1$ assisted dummy.

$Q2$ assisted dummy.

$Q1$ living alone dummy.

$Q2$ living alone dummy.

Dummy – previous surgery on same area.

Dummy – patient reported condition – heart disease.

Dummy – patient reported condition – high blood pressure.

Dummy – patient reported condition – stroke.

Dummy – patient reported condition – poor circulation.

Dummy – patient reported condition – lung disease.

Dummy – patient reported condition – diabetes.

Dummy – patient reported condition – kidney disease.

Dummy – patient reported condition – disease of central nervous system.

Dummy – patient reported condition – liver disease.

Dummy – patient reported condition – cancer.

Dummy – patient reported condition – depression.

Dummy – patient reported condition – arthritis.

Dummy – number of patient reported conditions = 2.

Dummy – number of patient reported conditions = 3.

Dummy – number of patient reported conditions = 4 or more.

Dummy – symptoms experienced for less than one year.

Dummy – symptoms experienced for 1-5 years.

Dummy – symptoms experienced for 6-10 years.

Dummy – symptoms experienced for more than 10 years.[52]

Variables taken from Hospital Episode Statistics

Age.

Age squared.

Dummy – mixed ethnicity.

Dummy – Asian.

Dummy – black.

Dummy – other ethnicity.

Dummy – unknown ethnicity.

Dummy – revision of previous hip replacement.

Dummy – HRG code 41.

Dummy – HRG code 72.

Dummy – HRG code 80.

Dummy – HRG code 81.

Charlson score.

Dummy – day case (no overnight stay).

Dummy – one comorbidity.

Dummy – two comorbidities.

Dummy – three or more comorbidities.

Dummy – self-discharge (patient discharged by self, friend or relative).

Index of Multiple Deprivations – score.

[52] For hernia repair, the last three dummies were replaced by a single 'Dummy – symptoms experienced for more than a year'.

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).[53]

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 π_{ij} denote the probability that patient i chooses hospital j: $\pi_{ij} = \Pr(H_i = j)$. Let η_{ij} denote the log of the odds that i will choose hospital j against the reference hospital (namely the Jth hospital). Finally let $\boldsymbol{x}_i$ (with 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. Then the conditional logit model is based on the premise that η_{ij} is a linear function of the explanatory variables, $\boldsymbol{x}_i$ and $\boldsymbol{z}_{ij}$:

$$\begin{aligned} \eta_{ij} &= \log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \boldsymbol{x}_i'\boldsymbol{\theta}_j + \boldsymbol{z}_{ij}'\boldsymbol{\gamma} \\ \pi_{ij} &= \pi_{iJ}\exp(\eta_{ij}) = \pi_{iJ}\exp(\boldsymbol{x}_i'\boldsymbol{\theta}_j + \boldsymbol{z}_{ij}'\boldsymbol{\gamma}) \end{aligned} \tag{11}$$

Summing Equation (11) over all J hospitals and noting that $\sum_{j=1}^{J}\pi_{ij} = 1$ yields that:

$$\pi_{iJ} = \frac{1}{\sum_{j=1}^{J}\exp\left(\eta_{ij}\right)} \tag{12}$$

Plugging the expression for π_{iJ} from Equation (12) into Equation (11), and replacing the js in the former with λs, yields:

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

A2.2 Utility Reformulation of the Conditional Logit Model

Section A2.1 related a patient's probability of choosing a given hospital to a set of individual-specific and hospital-specific explanatory variables. The same idea can also be stated by assuming that patients choose the hospital that maximises their utility, and specifying a patient utility function with a random component. Let patient i's utility from alternative j be an additive function of a systematic component $\boldsymbol{x}_i'\boldsymbol{\theta}_j + \boldsymbol{z}_{ij}'\boldsymbol{\gamma}$ and a random component ε_{ij}:

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

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:

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

If it is assumed that ε_{ij} is distributed standard Type 1 extreme value with cumulative distribution

[53] This Appendix draws from Maddala (1983), Rodríguez (2007) and the Web Appendix accompanying Gaynor et al. (2013).

function $F(\varepsilon) = e^{-e^{-\varepsilon}}$,[54] then it can be shown (Maddala 1983) that:

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

Thus the parameters of the model can be estimated in the same manner as in the formulation presented in Section A2.1.

A2.3 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 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.[55] Specifically, patient i's utility from attending hospital j is defined as:

$$U_{ij} = \sum_{h=1}^{3} \left\{ \begin{array}{c} \beta_1^h \left(d_{ij} - d_{ij+}^h\right) z_j^h + \beta_2^h \left(d_{ij} - d_{ij+}^h\right)\left(1 - z_j^h\right) \\ +\beta_3^h \left(d_{ij} - d_{ij-}^h\right) z_j^h + \beta_4^h \left(d_{ij} - d_{ij-}^h\right)\left(1 - z_j^h\right) \\ +\beta_5^h \left(female_i \cdot mid_i \cdot z_j^h\right) + \beta_6^h \left(female_i \cdot old_i \cdot z_j^h\right) \\ +\beta_7^h \left(male_i \cdot young_i \cdot z_j^h\right) + \beta_8^j \left(male_i \cdot mid_i \cdot z_j^h\right) + \beta_9^h \left(male \cdot old_i \cdot z_j^h\right) \\ +\beta_{10}^h \left(lowseverity_i \cdot z_j^h\right) + \beta_{11}^h \left(highseverity_i \cdot z_j^h\right) + \beta_{12}^h \left(charlson_i \cdot z_j^h\right) \\ +\beta_{13}^h \left(urban_i \cdot z_j^h\right) + \beta_{14}^h \left(poor_i \cdot z_j^h\right) + \boldsymbol{\beta}_{15}^h \left(\boldsymbol{region}_i \cdot z_j^h\right) + \varepsilon_{ij} \end{array} \right\} \quad (13)$$

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 β_1^3 and β_3^3 terms will be turned off; the β_2^3 term will then capture utility from differential distance between j and the nearest private hospital, while the β_4^3 term will capture utility from differential distance between j and the nearest NHS hospital.

[54] 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 $F(x) = \exp(-\exp(-x))$.

[55] 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.

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.

The parameters of the model are estimated separately for each surgical procedure and financial year in the dataset, by maximum likelihood. 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.

A2.4 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

The OPCS4 procedure codes used to define who should be given a PROMs survey are outlined in HSCIC (2013a). 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 were used to identify PROMs procedures in HES. HES provides allows up to 24 surgical procedures to be listed in an episode; matching was conducted on all 24 fields.

Hip replacement:

- Any patient with a procedure field beginning with W37, W38, W39, W46, W47, W48, W93, W94 or W95.
- Any patient with a procedure field beginning with W52, W53, W54, or W58, as well as any other procedure field beginning with Z761, Z756, or Z843.

Knee replacement:

- Any patient with a procedure field beginning with O18, W40, W41, or W42.
- Any patient with a procedure field beginning with W52, W53, or W54, as well as any other procedure field beginning with Z765, Z771, Z774, Z844, Z845, or Z846.

Varicose veins:

- Any patient with a procedure field beginning with L84, L85, L86, L87, L88, or L93.

Hernia:

- Any patient with a procedure field beginning with T19 through to T27.

A knee arthroscopy case is defined as any patient with a procedure field beginning with W82 through to W89.

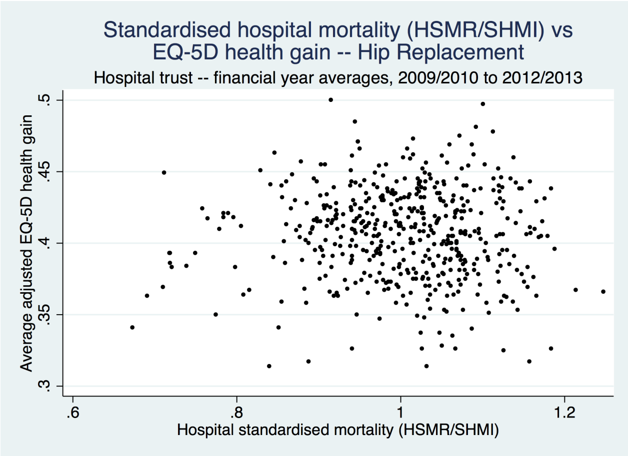
Cataract cases are identified using ICD10 diagnosis codes as well as procedure codes, and are defined as any patient with a procedure code beginning with C71 through to C77, as well as a diagnosis code beginning with H25, H26, H28, or Q120.

Any patients with ICD10 diagnosis codes beginning with I21 and I22 are identified as having experienced an acute myocardial infarction. In addition, to ameliorate concerns about possible upcoding of diagnoses, patients with these diagnosis codes that were discharged alive with a total length of stay of less than three days are excluded. Following Cooper et al. (2011), in the analysis of AMI mortality, attention is restricted to patients that are aged 39 to 100, and admitted on an emergency basis from their permanent or temporary abode (as opposed to, say, from another NHS hospital).

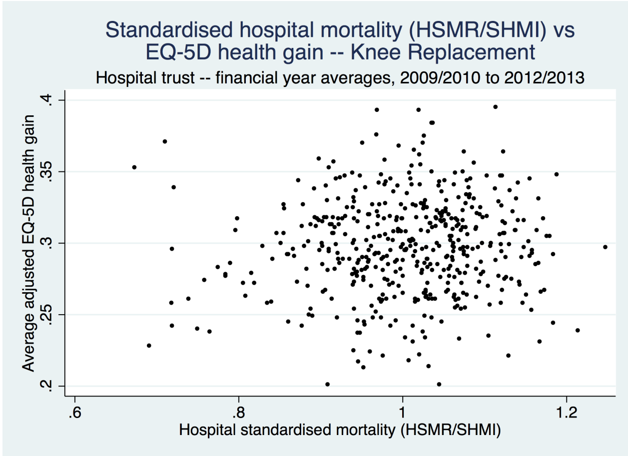
Figures

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

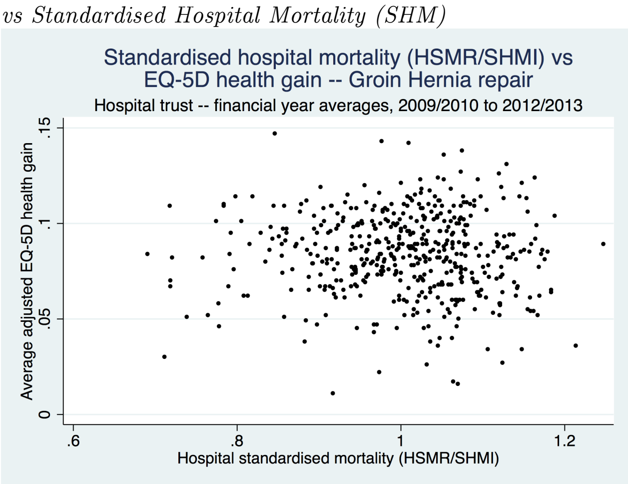
(A) Adjusted EQ-5D health gain from hip replacement vs Standardised Hospital Mortality (SHM)



(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)

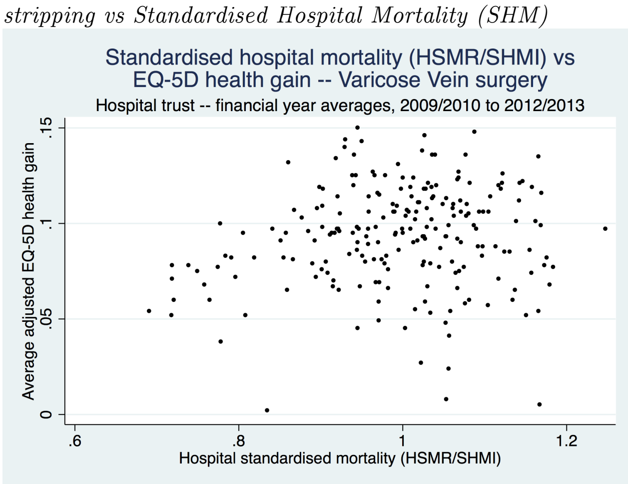
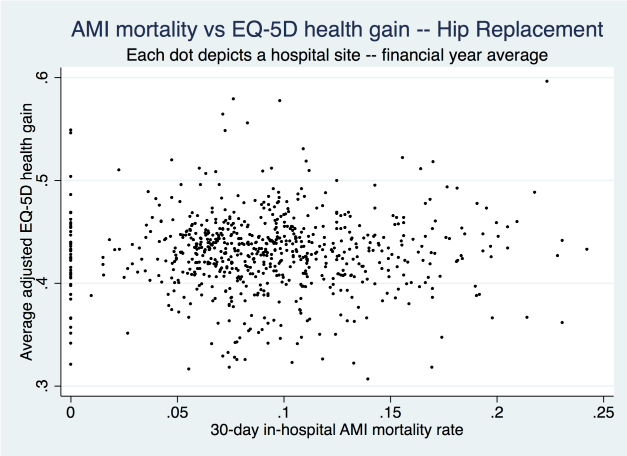


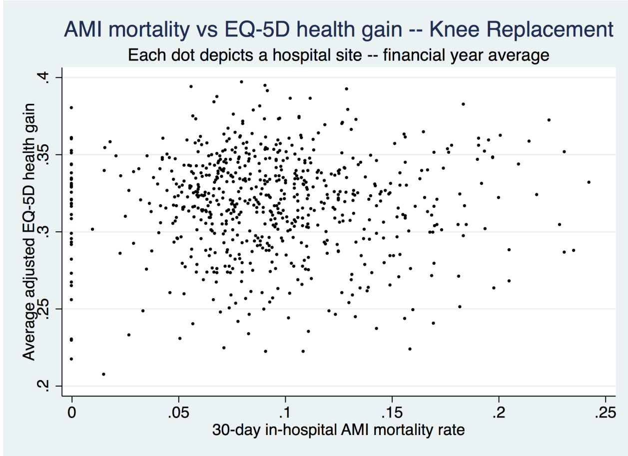
Figure graphs the relationship between hospital trusts' standardised (risk-adjusted) mortality rates (Standardised Hospital Mortality or SHM) and average casemix-adjusted health gain from elective surgery as captured by the EQ-5D PROM. There is one observation for each hospital trust and financial year. For 2009/2010, SHM is captured by Dr Foster's Hospital Standardised Mortality Ratios (HSMR) (Dr Foster 2011). For 2010/2011 onwards, SHM is captured by the NHS Standardised Hospital Mortality Indicator (SHMI) (HSCIC 2013b).

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

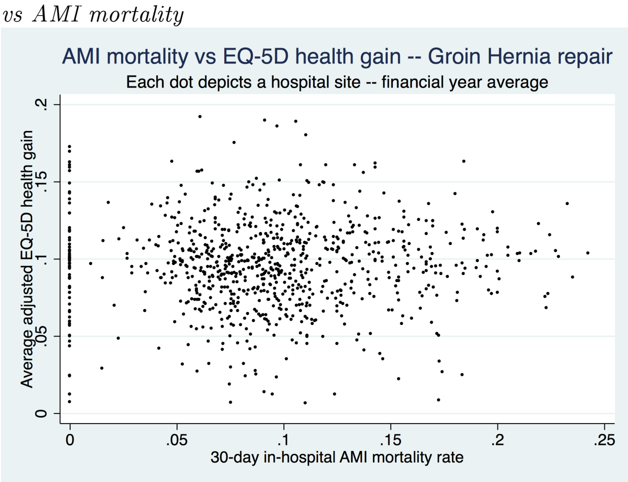
(A) Adjusted EQ-5D health gain from hip replacement vs AMI mortality



(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

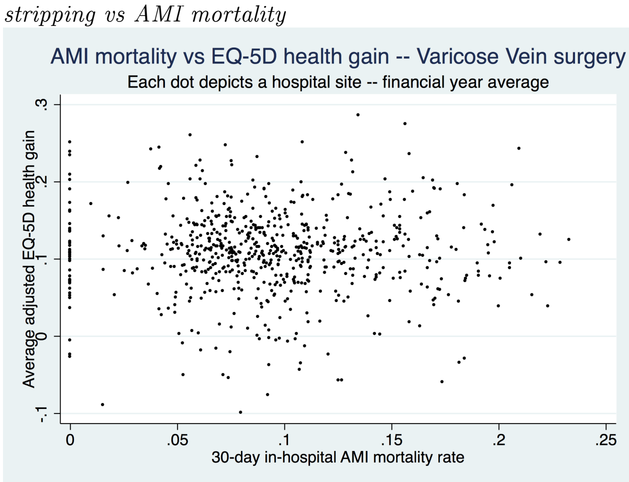
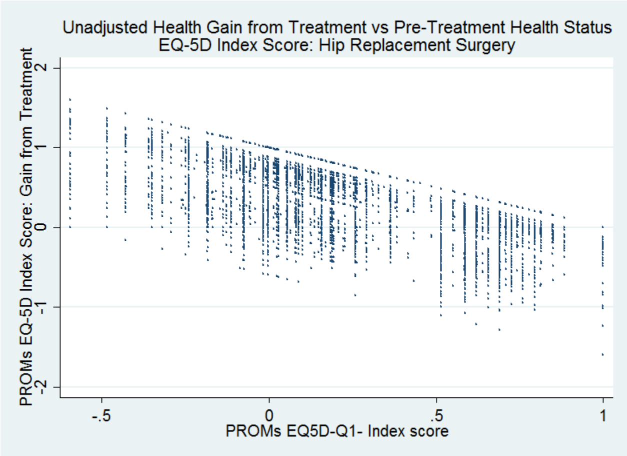


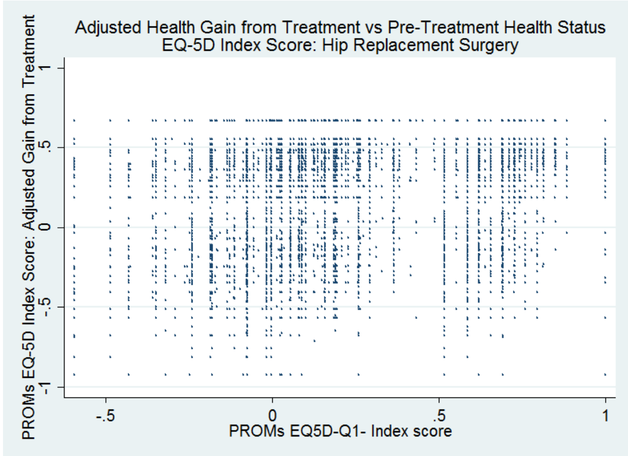
Figure graphs the relationship between individual hospital sites' mortality rates from acute myocardial infarction (AMI) and average casemix-adjusted health gain from elective surgery as captured by the EQ-5D PROM. There is one observation for each hospital site and financial year. The AMI mortality 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.

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

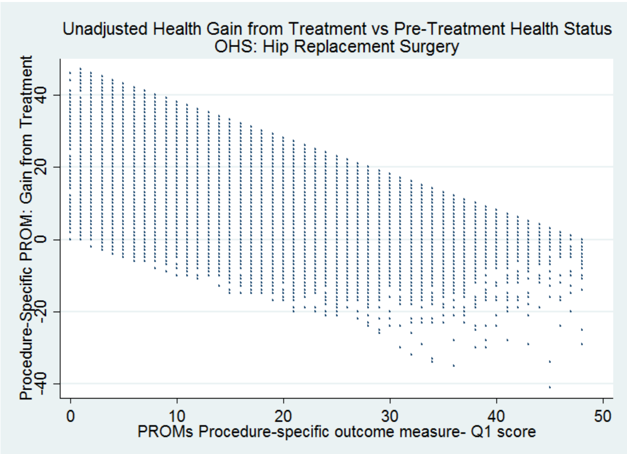
(A) EQ-5D Index Score for Hip Replacement: Unadjusted Health Gains vs Pre-Treatment Health Status



(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

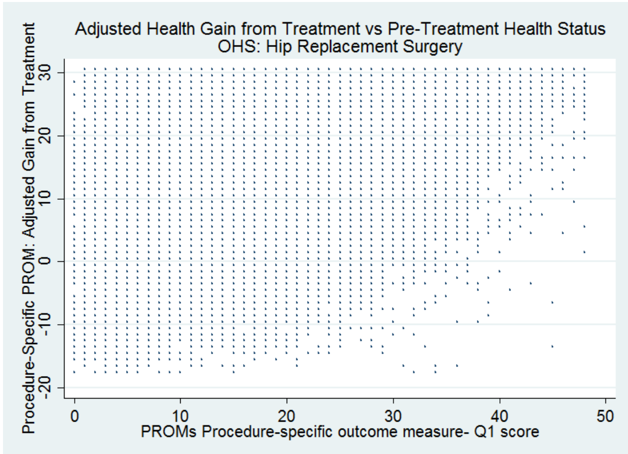


Figure graphs the relationship between hip replacement patients' pre-treatment health status and their adjusted and unadjusted health gains from treatment. Panels (A) and (B) show this relationship for the EQ-5D Index Score, with unadjusted and adjusted health gains on the y-axis respectively. Panels (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)
Varicose veins	0.0000
Knee replacement	0.0794
Hernia repair	0.0367
Hip replacement	0.0973
Acute myocardial infarction (AMI)	7.8411

Table reports 30-day in-hospital mortality rates in the four years from 2009/2010 to 2012/2013. 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	Groin Hernia	72409	0.095	0.168	-1.372	1.060
EQ-5D	Hip Replacement	110547	0.432	0.223	-0.926	1.516
EQ-5D	Knee Replacement	115402	0.323	0.237	-0.975	1.610
EQ-5D	Varicose Veins	24870	0.105	0.186	-1.341	1.335
OHS	Hip Replacement	122511	20.429	8.488	-17.592	67.002
OKS	Knee Replacement	126280	15.552	9.133	-18.181	83.047
AVVQ	Varicose Veins	26300	8.182	8.699	-65.870	73.386
EQ-VAS	Groin Hernia	68464	0.037	13.990	-79.099	243.520
EQ-VAS	Hip Replacement	106854	10.762	16.623	-64.149	93.224
EQ-VAS	Knee Replacement	111215	4.838	17.066	-66.627	103.204
EQ-VAS	Varicose Veins	24018	0.196	14.295	-78.504	96.747

Table reports the average casemix-adjusted health gain, by procedure, for each PROM 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
EQ-5D	1.25	0.93	0.38	0.39
EQ-VAS	0.43	0.18	-0.03	-0.01
OHS/OKS/AVVQ	2.32	1.87	N/A	0.71

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

Table 4: Convergent validity of PROMs outcome measures

(A) Correlation between hip replacement PROMs

	OHS	EQ-5D	EQ-VAS
OHS	1		
EQ-5D	0.6158	1	
EQ-VAS	0.4107	0.4685	1

(B) Correlation between knee replacement PROMs

	OKS	EQ-5D	EQ-VAS
OKS	1		
EQ-5D	0.6304	1	
EQ-VAS	0.3933	0.4329	1

(C) Correlation between groin hernia repair PROMs

	EQ-5D	EQ-VAS
EQ-5D	1	
EQ-VAS	0.3954	1

(D) Correlation between varicose vein stripping PROMs

	AVVQ	EQ-5D	EQ-VAS
AVVQ	1		
EQ-5D	0.3597	1	
EQ-VAS	0.1893	0.3447	1

Table reports the within-observation correlation between different PROMs for each of the surgical procedures studied in this paper. See Table 2 notes for explanation of 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	0.1345	1			
Knee replacement: average EQ-5D health gain	0.1816	0.2745	1		
Groin hernia: average EQ-5D health gain	0.068	0.0806	0.0519	1	
Varicose veins: average EQ-5D health gain	0.0559	-0.0158	-0.0145	0.0458	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	-0.0603	1			
Knee replacement: average adjusted EQ-5D health gain	0.0532	0.2575	1		
Groin hernia: average adjusted EQ-5D health gain	-0.0807	0.217	0.2229	1	
Varicose veins: average adjusted EQ-5D health gain	0.0877	0.2197	0.186	0.0851	1

Table reports the correlation between hospital trusts' standardised (risk-adjusted) mortality rates (Standardised Hospital Mortality or SHM) and average health gains from elective surgery as captured by the EQ-5D PROM. 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.

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	0.0037	1			
Knee replacement: average EQ-5D health gain	0.0259	0.2387	1		
Groin hernia: average EQ-5D health gain	0.0384	0.0254	0.0354	1	
Varicose veins: average EQ-5D health gain	0.0026	-0.0705	0.0156	0.0575	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	-0.0118	1			
Knee replacement: average adjusted EQ-5D health gain	-0.0085	0.1465	1		
Groin hernia: average adjusted EQ-5D health gain	0.0238	0.0354	-0.0088	1	
Varicose veins: average adjusted EQ-5D health gain	-0.0265	-0.0553	0.0302	0.0273	1

Table reports the correlation between hospital sites' mortality rate from acute myocardial infarction (AMI) with average health gains from elective surgery as captured by the EQ-5D PROM. 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.

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	0.0177	1			
Knee replacement: change in average adjusted EQ-5D health gain	0.0351	0.0099	1		
Groin hernia: change in average adjusted EQ-5D health gain	-0.0112	0.0122	0.0222	1	
Varicose veins: change in average adjusted EQ-5D health gain	-0.1988	0.1697	0.168	0.012	1

Table reports the correlation between first differenced Standardised Hospital Mortality (SHM) and first differenced trust-level average casemix-adjusted health gain from elective surgery as captured by the EQ-5D PROM. See Figure 1 notes for sources of SHM data.

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	0.0495	1			
Knee replacement: change in average adjusted EQ-5D health gain	0.0098	0.2816	1		
Groin hernia: change in average adjusted EQ-5D health gain	0.0198	0.076	-0.0196	1	
Varicose veins: change in average adjusted EQ-5D health gain	0.045	0.0011	0.1017	0.0882	1

Table reports the correlation between hospital sites' first differenced mortality rate from acute myocardial infarction (AMI) and first differenced average casemix-adjusted health gain from elective surgery as captured by the EQ-5D PROM. See Figure 2 notes for AMI sample restrictions.

Table 9: Correlation between different indicators of competition intensity

	GP-level, GP-centred 95% radius HHI	Site-level, GP-centred 95% radius HHI	Site-level, MSOA-centred HHI (actual choices)	Site-level, MSOA-centred HHI (predicted choices)	Number of sites within 30km	Lives in urban area
GP-level, GP-centred 95% radius HHI	1.000					
Site-level, GP-centred 95% radius HHI	0.448	1.000				
Site-level, MSOA-centred HHI (actual choices)	0.361	0.814	1.000			
Site-level, MSOA-centred HHI (predicted choices)	0.237	0.439	0.360	1.000		
Number of hospital sites within 30km	0.350	0.398	0.228	0.740	1.000	
Dummy: lives in urban area	-0.097	0.035	0.005	0.240	0.278	1.000

Table reports the correlation between measures of competition intensity used in the paper. HHIs are reported as 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: Number of PROMs procedures and linkage rates

	Total number in dataset	Number linked to a PROMs survey	Linkage rate (%)
Hernia Repair	442,101	102,671	23.22
Hip Replacement	274,854	158,200	57.56
Knee Replacement	294,989	165,973	56.26
Varicose Veins	119,927	41,734	34.80
Total	1,131,871	468,578	41.40

Table reports the number of PROMs procedures in the dataset used to calculate this paper's competition indices, and the number of observations successfully linked to a PROMs survey observation. The linkage rate for Hernia Repair is low because, for the purpose of calculating competition indices, this paper includes all Hernia Repair patients, whereas the PROMs surveys are targeted only at Groin Hernia Repair patients. See Appendix 3 for the procedure and diagnosis codes used to define each condition.

Table 11: Control variables – averages

	Groin hernia	Hip Replacement	Knee Replacement	Varicose Veins
Specialist hospital (%)	0.052	4.775	3.820	0.108
Teaching hospital (%)	13.840	12.935	12.721	21.450
Non-teaching university hospital (%)	14.339	16.123	16.391	15.098
Standard acute hospital (%)	55.815	50.116	51.331	56.647
Private hospital (%)	15.093	15.706	15.327	6.553
Age	58.705	67.753	69.163	50.602
Urban dweller (%)	75.649	71.958	75.554	79.635
Catchment area population density	27.207	25.590	26.562	33.779
Charlson score	0.817	1.330	1.515	0.446
Dummy- 1 diagnosis (%)	40.414	15.009	13.285	60.025
Dummy- 3+ diagnoses (%)	34.523	64.588	67.923	18.858
IMD04 income deprivation score (%)	13.247	12.417	13.464	15.061
Female (%)	7.282	59.267	57.033	62.159
Procedure-specific site FCEs/year	448.8	422.0	417.3	275.0
Total trust admissions/year	92431.3	88377.7	90283.5	112150.3

Table reports average values of key control variables used in this paper, separated by PROMs procedure.

Table 12: Main estimates using EQ-5D and procedure-specific PROMs

	Cross-sectional estimates		First stage estimates: pre-reform HHI instrument		Second stage estimates: pre-reform HHI instrument	
	(1)	(2)	(3)	(4)	(5)	(6)
	MSOA-Actual	MSOA-Predicted	MSOA-Actual	MSOA-Predicted	MSOA-Actual	MSOA-Predicted
Orthopaedic surgery:	-0.229	-0.322**	0.761***	0.956***	-1.492***	-0.322**
Oxford Score health gain	(0.253)	(0.134)	(0.0726)	(0.0263)	(0.501)	(0.153)
Orthopaedic surgery:	-0.0015	-0.00446	0.759***	0.955***	-0.0239**	-0.00454
EQ-5D health gain	(0.00485)	(0.00289)	(0.0727)	(0.0265)	(0.0113)	(0.00341)
Groin hernia repair:	0.00157	-0.00046	0.682***	0.990***	-0.0111	-0.000944
EQ-5D health gain	(0.00403)	(0.00188)	(0.0831)	(0.0205)	(0.00929)	(0.00212)
Varicose vein stripping: Aberdeen	-1.009*	-0.767***	0.720***	1.009***	0.178	-0.813***
Questionnaire health gain	(0.524)	(0.238)	(0.0744)	(0.0304)	(0.97)	(0.245)
Varicose vein stripping:	-0.0044	-0.00770*	0.720***	1.008***	0.0215	-0.00809*
EQ-5D health gain	(0.00834)	(0.00417)	(0.0746)	(0.0304)	(0.0144)	(0.00467)
All PROMs procedures:	-0.00749*	-0.00454	0.705***	0.947***	-0.0148*	-0.00573*
EQ-5D health gain	(0.00388)	(0.00293)	(0.0677)	(0.0238)	(0.00854)	(0.00314)

Table reports the paper's main estimates of the effect of hospital competition on casemix-adjusted health gain from elective surgery as captured by the EQ-5D PROM and the procedure-specific PROMs (Oxford Hip Score, Oxford Knee Score and Aberdeen Varicose Vein Questionnaire). Estimates for orthopaedic surgery pool hip and knee replacement surgery observations in a single regression, with a dummy variable for knee replacement to capture any differences in the level of health gains from the two procedures. Columns (1), (3) and (5) use a neighbourhood (MSOA) centred, hospital-level HHI based on actual patient choices. Columns (2), (4) and (6) use a neighbourhood (MSOA) centred, hospital-level HHI based on predicted patient choices. Columns (1) and (2) report cross-sectional estimates, using use current-period HHI as the treatment intensity variable, and current-period values of total and procedure-specific admissions (and their respective quadratics) as controls. Columns (3) to (6) report estimates when current period competition intensity is instrumented by average pre-reform (2002/3-2004/5) competition intensity, and lagged values of total and procedure-specific admissions (and their respective quadratics) as used as controls in place of their current-period values. Columns (3) and (4) report first stage estimates – i.e. the coefficient on the excluded variable (pre-reform average HHI) when current-period HHI is regressed on the excluded variable plus all the control variables included in the second stage. Columns (5) and (6) report the second stage estimates from regression of hospital quality (as captured by PROMs health gains) on the instrumented variable plus controls. 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. Statistical significance is reported as follows: *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the hospital level are reported in parentheses.

Table 13: Estimates using EQ-VAS PROM

	Cross-sectional estimates		First stage estimates: pre-reform HHI instrument		Second stage estimates: pre-reform HHI instrument	
	(1)	(2)	(3)	(4)	(5)	(6)
	MSOA-Actual	MSOA-Predicted	MSOA-Actual	MSOA-Predicted	MSOA-Actual	MSOA-Predicted
Orthopaedic surgery:	-0.751**	-0.172	0.761***	0.956***	-1.348*	-0.127
EQ-VAS health gain	(0.362)	(0.179)	(0.0728)	(0.0263)	(0.719)	(0.201)
Groin hernia repair:	0.188	0.129	0.680***	0.988***	0.0297	0.0453
EQ-VAS health gain	(0.370)	(0.164)	(0.0828)	(0.0206)	(0.756)	(0.178)
Varicose vein stripping:	-0.522	-0.539*	0.716***	1.008***	1.532	-0.585*
EQ-VAS health gain	(0.638)	(0.284)	(0.0746)	(0.0305)	(0.996)	(0.315)
All PROMs procedures:	-0.650**	-0.104	0.705***	0.948***	-1.004*	-0.136
EQ-VAS health gain	(0.291)	(0.191)	(0.0677)	(0.0237)	(0.538)	(0.194)

Table reports the paper's estimates of the effect of hospital competition on casemix-adjusted health gain from elective surgery as captured by the EQ-VAS PROM. See Table 12 for further information.

Table 14: Alternative specifications

	(1) No Year-Region Interaction	(2) No Urban or Pop Density Controls	(3) All London Obs Omitted	(4) Instrument All Scale Variables	(5) Omit All Scale Variables	(6) All Variables in Logs	(7) Procedure-Level Comp Measures
(A) Competition measure: Site-level, MSOA-centred HHI (actual choices)							
Orthopaedic surgery: Oxford Score health gain	-1.417***	-1.359***	-1.439***	-1.551***	-0.708*	-0.0195***	-1.977***
	(0.511)	(0.525)	(0.512)	(0.552)	(0.389)	(0.00726)	(0.561)
Orthopaedic surgery: EQ-5D health gain	-0.0220*	-0.0224*	-0.0236**	-0.0239*	-0.01	-0.0105*	-0.0355***
	(0.0116)	(0.0115)	(0.0117)	(0.0129)	(0.00938)	(0.00579)	(0.012)
Groin hernia repair: EQ-5D health gain	-0.0108	-0.012	-0.0134	-0.0123	-0.0140**	-0.00695	-0.0133
	(0.00932)	(0.00944)	(0.0104)	(0.00897)	(0.00681)	(0.00525)	(0.0155)
Varicose vein stripping: Aberdeen Questionnaire health gain	0.281	0.296	0.775	0.222	-0.318	-0.000322	3.512
	(0.988)	(0.978)	(1.136)	(0.963)	(0.842)	(0.00833)	(4.025)
Varicose vein stripping: EQ-5D health gain	0.0202	0.0234*	0.0269	0.0245*	0.0142	0.0161*	0.0917
	(0.0148)	(0.014)	(0.019)	(0.0141)	(0.0111)	(0.00908)	(0.0656)
(B) Competition measure: Site-level, MSOA-centred HHI (predicted choices)							
Orthopaedic surgery: Oxford Score health gain	-0.322**	-0.322**	-0.350**	-0.514***	-0.322**	-0.00573**	-0.352**
	(0.153)	(0.153)	(0.156)	(0.181)	(0.153)	(0.00256)	(0.153)
Orthopaedic surgery: EQ-5D health gain	-0.00454	-0.00454	-0.00524	-0.00748*	-0.00454	-0.00291	-0.00464
	(0.00341)	(0.00341)	(0.00347)	(0.00396)	(0.00341)	(0.00198)	(0.00323)
Groin hernia repair: EQ-5D health gain	-0.000948	-0.000948	-0.000695	-0.00103	-0.000948	-0.00134	-0.00191
	(0.00212)	(0.00212)	(0.00216)	(0.00225)	(0.00212)	(0.00144)	(0.00232)
Varicose vein stripping: Aberdeen Questionnaire health gain	-0.822***	-0.822***	-0.732***	-0.838***	-0.822***	-0.00787***	-0.575*
	(0.246)	(0.246)	(0.257)	(0.294)	(0.246)	(0.00236)	(0.325)
Varicose vein stripping: EQ-5D health gain	-0.00837*	-0.00837*	-0.00612	-0.00734	-0.00837*	-0.00616*	-0.00639
	(0.00467)	(0.00467)	(0.00486)	(0.0052)	(0.00467)	(0.00324)	(0.00562)

Table reports the coefficient on the treatment intensity variable (impact of competition on our main outcome variables) using a range of alternative specifications. Panel (A) reports estimates when treatment intensity is captured using a hospital-level, neighbourhood-centred HHI based on actual patient choices. Panel (B) reports estimates when treatment intensity is captured by the same measure, but based on predicted patient choices. Unless indicated below, all specifications instrument current-period competition intensity by its pre-reform average value, and use pre-reform average values for all scale variables. Column (1) includes separate year and region of England dummies, as opposed to interacting these dummies, as is the case in the headline specification. Column (2) omits controls for patient's urban status and hospital catchment area population density. Column (3) omits all observations from London hospitals. Column (4) instruments total trust admissions and procedure-specific hospital site admissions (as well as their respective quadratic terms) with their pre-reform average values, as opposed to using pre-reform average values in place of current-period values. Column (5) omits all scale effects controls (total trust admissions, procedure-specific hospital site admissions, and their respective quadratic terms). Column (6) converts all outcome variables and non-dummy control variables to logs. Column (7) uses procedure-specific HHIs (e.g. competition intensity calculated using varicose vein observations when running regressions using a varicose vein outcome measure), as opposed to competition measures averaged across six procedures. See Table 12 for further explanation.

Table 15: Variable radius HHI using GP-centred markets

	Cross-sectional estimates		First stage estimates: pre-reform HHI instrument		Second stage estimates: pre-reform HHI instrument	
	(1) GP-centred, Hosp-level	(2) GP-centred, GP-level	(3) GP-centred, Hosp-level	(4) GP-centred, GP-level	(5) GP-centred, Hosp-level	(6) GP-centred, GP-level
Orthopaedic surgery: Oxford	-0.208	0.0221	0.924***	0.770***	-0.481**	-0.198
Score health gain	(0.135)	(0.0696)	(0.0585)	(0.0361)	(0.202)	(0.122)
Orthopaedic surgery:	-0.00139	-0.000904	0.923***	0.768***	-0.00512	-0.00233
EQ-5D health gain	(0.00286)	(0.00142)	(0.0585)	(0.0357)	(0.00425)	(0.00277)
Groin hernia repair:	0.00331*	0.00110	0.709***	0.785***	-0.00159	-0.00345
EQ-5D health gain	(0.00201)	(0.00145)	(0.0975)	(0.028)	(0.00398)	(0.00244)
Varicose vein stripping: Aberdeen	-0.631**	-0.205	0.777***	0.790***	-0.114	-0.156
Questionnaire health gain	(0.305)	(0.160)	(0.0843)	(0.033)	(0.607)	(0.269)
Varicose vein stripping:	-0.00535	-0.00141	0.777***	0.785***	0.00646	-0.00128
EQ-5D health gain	(0.00498)	(0.00328)	(0.0842)	(0.0329)	(0.00822)	(0.00513)
All PROMs procedures:	-0.00364*	-0.00196*	0.833***	0.772***	-0.00321	-0.00255
EQ-5D health gain	(0.00216)	(0.00115)	(0.0613)	(0.0299)	(0.00355)	(0.00222)

Table reports the same set of estimates as Table 12, but using indicators of competition intensity centred on GP surgeries rather than neighbourhoods (MSOAs). Each GP surgery's market is defined by the 95th percentile of patient distance travelled by the surgery's patients. Columns (2), (4) and (6) report estimates when this indicator of competition intensity is calculated at the GP level, while Columns (1), (3) and (5) report estimates when these GP-centred, GP-level competition indices are aggregated to hospital level. See Table 12 for further explanation.

Table 16: Additional Robustness Tests

	(1) Number of hospitals within 30km	(2) MSOA-centred, hospital-level predicted patient choice HHI with hospital fixed effects
Orthopaedic surgery: Oxford	-0.0291	-0.125
Score health gain	(0.0183)	(0.121)
Orthopaedic surgery:	-0.000446	-0.00558
EQ-5D health gain	(0.000358)	(0.0036)
Groin hernia repair:	0.0000893	0.0117
EQ-5D health gain	(0.000186)	(0.0106)
Varicose vein stripping: Aberdeen	-0.0553**	-0.00447
Questionnaire health gain	(0.0235)	(0.0106)
Varicose vein stripping:	0.000241	0.0524
EQ-5D health gain	(0.000531)	(0.453)

Table reports the coefficient on the treatment intensity variable (impact of competition on our main outcome variables) for additional robustness tests. Column (1) reports estimates using a fixed-radius measure of market structure, in which competition intensity is equal to the number of hospitals within 30km. Column (2) reports estimates using an MSOA-centred, hospital-level HHI based on predicted patient choices, when hospital fixed effects are included in the regression. See Table 12 for further explanation.

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	2.082
EQ-5D Visual Analogue Score	0	0	0	100	100	322.619
Oxford Hip Score	0	0	0	48	48	84.594
Oxford Knee Score	0	0	0	48	48	101.229
Aberdeen Varicose Vein Questionnaire	14.383	0.342	13.362	100	100	153.639

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.