UNIVERSITY of York

RESEARCH

CHE

Centre For Health Economics

IMPROVING QUALITY OF CARE IN EUROPE

The Causal Effect of Hospital Volume on Health Gains from Hip Replacement Surgery

Laurie Rachet-Jacquet, Nils Gutacker, Luigi Siciliani

CHE Research Paper 168

The causal effect of hospital volume on health gains from hip replacement surgery

[a]Laurie Rachet-Jacquet
[a]Nils Gutacker
[ab]Luigi Siciliani

[a]Centre for Health Economics, University of York, UK
[b]Department of Economics and Related Studies, University of York, UK

August 2019

Background to series

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement number 721402. Hospital Episode Statistics are Copyright 2015–2016, re-used with the permission of NHS Digital. All rights reserved. Data on consultant characteristics were obtained from the General Medical Council. No ethical approval was needed.

We are grateful to Martin Chalkley, Alistair McGuire, Toni Mora, Nigel Rice, Jonas Schreyögg, Erin Strumpf and participants of the European Training Network research in progress workshop (Rotterdam, May 2018), the European Health Economics Association (EuHEA) conference (Maastricht, July 2018), the EuHEA PhD student-supervisor conference (Catania, September 2018), the Health Econometrics Data Group seminar at the University of York, the Health Economists' Study Group meeting (York, January 2019) and IRDES-Dauphine workshop (Paris, June 2019) for valuable comments and suggestions.

Further copies

Centre for Health Economics
Alcuin College
University of York
York,
YO10 5DD, UK

www.york.ac.uk/che

Abstract

This study investigates the causal effect of hospital volume on health gains from hip replacement surgery in the English National Health Service. We exploit a unique dataset, which links routine hospital records and patient-reported outcome measures (PROMs) for all public hospitals in England. PROMs assess patients' health along key dimensions of pain and mobility shortly before and six months after the surgery. We investigate whether higher hospital volume increases patient health six months post-surgery, conditioning on pre-surgery health and other patient medical and socioeconomic indicators. We address possible reverse-causality bias due to hospital demand being responsive to quality by constructing a measure of predicted hospital volumes based on a patient choice model. The results suggest that the observed volume-outcome effect in hip replacement surgery is clinically small and no longer statistically significant once we account for the endogeneity of volume.

Keywords: quality; hospital; volumes; learning-by-doing; patient reported outcomes.
JEL-codes: I11, I18, L25

1. Introduction

Improving quality of care is a key policy objective in health systems across high income countries (OECD, 2017). Policy initiatives commonly rely on the premise that high-volume healthcare providers are able to deliver better care, by exploiting economies of scale and learning-by-doing effects (Ho, 2014). For instance, the Leapfrog group, a coalition of healthcare purchasers in the United States, has set minimum volume standards for hospital referrals since the early 2000s (Birkmeyer and Dimick, 2004). Similarly, France, Germany and the Netherlands have introduced minimum volume regulations in certain surgeries (Bauer and Honselmann, 2017; Mesman et al., 2017; Com-Ruelle, Or and Renaud, 2008).

Despite a large literature investigating the relationship between volume and health outcomes across a range of procedures and countries (Gutacker et al., 2017; Halm, Lee and Chassin, 2002; Ferguson, Posnett and Sheldon, 1997), evidence of a causal effect of volume on quality remains limited due to the potential endogeneity of volume (Luft, Hunt and Maerki, 1987). Specifically, volume-outcome studies are prone to a reverse causality bias if patients' choice of hospital responds to quality via reputation or public reporting (Gutacker et al., 2016; Brekke et al., 2014). Yet, understanding the causal mechanisms behind the volume-outcome association is essential in the context of policies seeking to improve quality of care by concentrating the provision of care. If the volume-outcome association is driven by demand's responsiveness to quality (sometimes referred to as 'selective-referral'), concentrating surgical activity will not improve quality and may have adverse effects on patient accessibility to care (Blanco et al., 2017). Alternatively, if higher volumes lead to better outcomes due to economies of scale, more concentrated hospital markets may lead to improved patient outcomes (Brekke, Siciliani and Straume, 2017; Gaynor and Town, 2011).

This study investigates the causal effect of hospital volume on the health gains of patients receiving a primary (i.e. non-revision) planned hip replacement procedure in the English NHS in 2015/16. Hip replacement surgery is a common planned procedure, which involves replacing the damaged part of a hip joint by an artificial one. Unexplained variations in patient-reported outcomes after hip replacement have been reported at the hospital (Street et al., 2014) and surgeon levels (Varagunam, Hutchings and Black, 2015a), while fixation methods and implant types are associated with differences in revision rates (Healthcare Quality Improvement Partnership, 2018). Hospital volumes for hip replacement vary little over time. We therefore exploit the cross-sectional variation of volume across all public hospitals in England to test for the presence of economies of scale in health outcomes. We find that hospital volume does not have a causal impact on patient-reported health outcomes for planned hip replacement.

Our contribution to the previous literature (reviewed briefly in section 1.1) is threefold. First, we use patient-reported outcome measures (PROMs) to capture the effect of volume in terms of improvements in patients' health status. Post-operative mortality is very low for planned hip replacement patients, thus rendering commonly used measures (mortality or complication rates) insensitive to finer variations in quality (Varagunam, Hutchings and Black, 2015a; Shojania and Forster, 2008). Second, the availability of rich patient-reported data on functional status collected just before the surgery ensures that we thoroughly control for patient severity and minimise the risk of omitted variable bias through unobserved severity.

Third, we address the reverse causality bias by employing a measure of *predicted* volumes, rather than *actual* volumes. Predicted volumes are derived from a conditional logit choice model where patients' choice of hospital is a function of exogenous determinants, including the distance between patient residence and each hospital. In doing so, we apply a method commonly used in the literature on hospital competition following the seminal study by Kessler and McClellan (2000), which uses predicted volumes (patient flows) to construct Herfindahl-Hirschman indices based on hospital market shares (Cooper, Gibbons and Skellern, 2018; Gaynor, Moreno Serra and Propper, 2013; Gaynor and Town, 2011).

By addressing these two sources of endogeneity, i.e. insufficient risk adjustment and reverse causality, we obtain causal estimates of the effect of hospital volume on patients' health benefits following planned hip replacement surgery. Our findings suggest that failing to account for hospital volume endogeneity generates a spurious positive relationship whereby hospitals of higher quality also face a higher demand and thus higher volumes. We also show that controlling for surgeon volumes does not change our results at the hospital level, suggesting that the relationship between hospital volumes and outcomes does not reflect surgeon effects.

In the remainder of this section we give a brief account of the literature. Section 2 introduces the data. Section 3 lays out the methods and section 4 presents the results. Section 5 concludes.

1.1. Related literature

Quality improvements driven by volume of practice may occur through different channels. At the hospital level, economies of scale may take place through better collaboration between surgeons and nursing staff, familiarity with the operating theatre, the presence of specialists and technology-based services, or more standardised processes of care (Ho, 2014; Kizer, 2003). At the surgeon level, the volume-outcome effect is more readily understood as a practice-makes-perfect effect in surgical skills or a better choice of treatment, with higher volumes leading to better outcomes through improved surgical technique or ability to detect and prevent complications (Chowdhury, Dagash and Pierro, 2007).

Several studies in orthopaedic surgery find a positive association between hospital volume and health outcomes after primary hip replacement surgery. Hospitals with high volume of planned and emergency hip replacement patients are associated with lower mortality rates or complication rates in England, the US and the Netherlands (De Vries et al., 2011; Singh et al., 2011; Judge et al., 2006). Previous studies find a negative relationship between surgeon volumes and rate of revisions or complications after primary planned hip replacement, using data from the US (Losina et al., 2004) and Canada (Ravi et al., 2014; Paterson et al., 2010). These studies further show that patients treated by low-volume surgeons are associated with higher rates of revision, at both low and high-volume hospitals (Ravi et al., 2014; Losina et al., 2004). In a setting close to ours, Varagunam, Hutchings and Black (2015b) use Hospital Episode Statistics from England for 2011/2012 and patient-reported outcome measures for planned hip replacement. They report no association between hospital volume and outcome,[1] but find a significant and positive association between surgeon volume and PROMs scores. The authors, however, implicitly assume volume to be exogenously determined.

[1] Our results here differ from theirs, potentially because we exclude private hospitals (see section 2.1 for a justification).

Causal studies in the literature exploit the geographical distribution of patients, or use variation in volume caused by the closure/opening of surrounding clinics (Avdic, Lundborg and Vikström, 2019 for cancer care in Sweden) or surgeons' retirement from practice (Ramanarayanan, 2008 for cardiac care in the US). In orthopaedic surgery, Luft, Hunt and Maerki (1987) find evidence of both demand's response to quality ('selective-referral') and practice-makes-perfect effect in hip replacement using simultaneous equation methods with US data. Hentschker and Mennicken (2018) use the distribution of patients and hospital competitors around the hospital as an instrument for hospital volumes with 2007 data for Germany. They find that hospital volume reduces in-hospital mortality for emergency hip replacement after hip fracture.

2. Data

2.1. Sample

We extract data from the Hospital Episodes Statistics (HES) on all planned (i.e. non-emergency) hip replacement surgeries in England performed between April 2015 and March 2016. HES is an administrative dataset on hospital admissions in England, which includes detailed patient demographic and medical information. The original sample consists of about 70,000 patients. To ensure sample homogeneity, we exclude patients below 50 years old and those having a revision surgery, which is a rare and more complex procedure.[2] We further exclude uncommon types of hip replacement (e.g. total prosthetic replacement of the head of the femur or resurfacing arthroplasty of the joint, which count for less than 0.01% of the sample). Hospitals reporting an unusually low number of cases (below 20 annual hip replacements) are also excluded to attenuate the risk of coding errors.

Since April 2009, a national programme requires hospitals in England to collect patient-reported outcome measures (PROMs) from patients who undergo certain planned surgeries (hip or knee replacement, varicose vein and groin hernia repair). Participation in this programme is voluntary for patients but mandatory for all hospitals that treat NHS-funded patients. Using a paper-based survey, all eligible orthopaedic patients are asked to report their health status, functioning and health-related quality of life immediately before and six months after surgery.[3] We use data collected via the Oxford Hip Score (OHS) questionnaire, which is a hip-specific instrument that has been clinically validated as an accurate measure of health status for patients with problems of the hip joint (Ostendorf et al., 2004; Dawson et al., 1996).

The PROMs and HES data are linked based on a number of identifying characteristics, including their unique NHS number (NHS Digital, 2017). Sixty three per cent of all hip patient admissions are successfully matched, which corresponds to about 30,000 admission records.[4]

We further exclude all private hospitals from our sample. HES includes all patient admissions (privately and NHS-funded) in NHS hospitals, but only admissions for NHS-funded patients in private hospitals. The volumes for private hospitals provided by HES would therefore underestimate their total volume of activity. This measurement error may vary across hospitals, depending on each hospital's share of activity spent on private patients. This is problematic as the distribution of the measured volumes for these hospitals may differ from the distribution of the actual total volumes, thus confounding results from a volume-outcome relationship based on the observed volumes. As a result of this exclusion, our final sample includes 18,979 patients.

2.2. Dependent variable

Our measure of patient health, the OHS, contains 12 items relating to functional status (mobility) and pain, each of which is evaluated on a scale from 0 to 4. The OHS is the sum of the scores obtained for each item and goes from 0 (worst) to 48 (best health status). The same OHS questions are distributed to the patients shortly before and six months after surgery. We use patients' post-

[2] Revision surgeries represented around 11% of total hip patient admissions in HES in 2015/16.
[3] The questionnaire can be found at:
https://nhs-digital.citizenspace.com/rocr/213-056/supporting_documents/213056%20HIP_questionnaire.pdf.
[4] Patient participation in PROMs is voluntary and attrition may happen for different reasons. If attrition is correlated with hospital volumes and health outcomes, our estimates will be biased. We find no evidence of a correlation between hospitals' response rates and volumes (see Figure A1 in the Appendix). Furthermore, our rich set of risk-adjusted variables limits the risk of attrition bias linked to unobserved patient severity. Overall, we conclude that attrition bias is unlikely to be severe.

surgery OHS as our dependent variable but control for the pre-surgery OHS, thereby assessing patient's *health gain* from the surgery.

2.3. Independent variables

Our key independent variables are the annual hospital volumes, measured as the number of patients who have undergone a planned primary hip replacement at a given hospital during the financial year 2015/16. In the English NHS, hospitals are organised into legal entities, formally called NHS trusts. We measure volume at the more disaggregated hospital (site) level rather than at the Trust level to obtain the physical concentration of activity in a facility, which we assume to be more relevant to economies of scale.

We control for patients' demographic characteristics (age, gender and ethnic group) and socio-economic deprivation, where the latter is based on the quintiles of the 2010 index of multiple deprivation (IMD)[5] measured at the small residence area level (lower-level super output area, LSOA) of the patients. Our model includes pre-surgery OHS grouped in narrow bands to capture potential non-linear effects. Our model also controls for the patient's self-assessed disability status prior to the surgery, symptom duration and living arrangements, as well as self-reported depression and assistance with completing the questionnaire (Department of Health, 2012, p. 3). We count the Elixhauser comorbidities reported in a patient's hospital stays up to one year prior to the admission for hip replacement (Gutacker et al., 2016; Elixhauser et al., 1998). We also control for whether the patient has previously undergone hip surgery on the other hip in the past four years, the primary diagnosis (e.g. osteoarthritis) (Losina et al., 2004) and the type of surgery (i.e. total hip replacement vs hybrid prosthetic replacement).

We further include controls for hospital characteristics that may be associated with higher quality (e.g. via medical expertise or better resources) independently of volumes. In particular, we control for hospitals' teaching status, whether the hospital is a specialist (orthopaedic) hospital[6] or an NHS Foundation Trust (FT) as the latter have greater financial autonomy (Gravelle, Santos and Siciliani, 2014).

[5] The index of multiple deprivation measures deprivation across seven domains, including income, employment and education.

[6] We extract information on teaching and specialist status from the Estates Returns Information Collection https://digital.nhs.uk/data-and-information/publications/statistical/estates-returns-information-collection/estates-returns-information-collection-eric-england-2015-16.

3. Methods

3.1. Baseline model with observed volumes

We study the effect of hospital volume on health gains after hip replacement surgery. Our econometric model is specified as follows:

$$y_{ih} = \alpha + \boldsymbol{vol}'_{\boldsymbol{h}}\boldsymbol{\beta_1} + \boldsymbol{x}'_{\boldsymbol{ih}}\boldsymbol{\beta_2} + \boldsymbol{k}'_{\boldsymbol{h}}\boldsymbol{\beta_3} + \varepsilon_{ih} \ , \tag{1}$$

where y_{ih} is the post-surgical OHS of patient i in hospital h, $\boldsymbol{x_{ih}}$ is a vector of patient characteristics (age in 10-year bands, gender, comorbidities, the pre-surgery OHS, socio-economic status) to adjust for differences in case-mix across hospitals, and $\boldsymbol{k_h}$ is a vector of dummy variables for hospital characteristics (i.e. NHS Foundation trust, specialist orthopaedic or teaching hospitals).[7] ε_{ih} is a random error term.

Our main interest is the effect of hospital volume on patients' post-surgery health status. Hospital volume $\boldsymbol{vol_h}$ is entered as a vector of four dummy variables corresponding to volume categories: $vol_h < 150$, $150 \geq vol_h > 200$, $200 \geq vol_h > 300$ and $vol_h \geq 300$. This allows for a non-linear relationship due to decreasing marginal returns to scale; especially at the lower end of the volume distribution where scale economies are likely to occur. To our knowledge, there is no evidence on the safety threshold for planned hip replacement using PROMs. We therefore define category thresholds that are both policy relevant and fit the distribution of the observed and predicted volumes.[8] Using defined categories rather than using quartiles ensures comparability of the results across specifications given that observed and predicted volumes follow different distributions. Equation (1) is estimated with OLS. We adjust standard errors for clustering at the hospital level.[9]

3.2. Endogeneity concerns

Cross-sectional regression models, such as that defined in (1), may provide a biased estimate of the volume-outcome relationship in the presence of reverse causality from quality to volume, or omitted variables linked to unobserved patient severity or hospital characteristics.

First, low (high) quality hospitals will face a lower (higher) demand, thus inducing a positive correlation in our estimates of (1). To address this, we borrow from the literature on the effect of competition on hospital quality. A similar challenge in this literature arises since hospitals' market share, measuring the hospital market structure via the Herfindahl-Hirschman Index (HHI), may be potentially (endogenously) determined by the quality of the hospital and of its competitors (Gaynor, Moreno Serra and Propper, 2013; Gowrisankaran and Town, 2003). These studies use discrete patient choice models, based on patients' distance to the hospital and hospital characteristics, to obtain predicted patient volumes and thus predicted market shares to derive an exogenous measure of hospital market structure (i.e. exogenous HHI).

---

[7] These characteristics are defined at the trust level. For simplicity, we use the same subscript h for hospitals and trusts.

[8] The median hospital volume is 200 cases for observed and predicted volumes, which is the cut-off value between the second and third volume category.

[9] Technically, we cluster at the trust (i.e. legal entity) level, given possible correlation across hospitals within a trust.

We follow a similar approach but focus on predicted hospital volumes rather than market shares. This amounts to constructing the volumes that would be observed if patients were choosing hospitals based on proximity.[10] Our identification strategy is based on the assumptions that i) patients' residential choices are not based on the quality of surgical interventions provided by the surrounding hospitals, and ii) that patients derive higher disutility and costs from travelling further. This seems reasonable in the case of planned hip surgery, where patients face an uncertain future demand for care and where timely access to care does not affect health outcomes (Brealey et al., 2012; Tuominen et al., 2010).[11] A formal presentation of the choice model is given in section 3.3.

Second, family physicians may refer their most severely ill patients to hospitals with better quality and higher volumes (Hentschker and Mennicken, 2018; Geweke, Gowrisankaran and Town, 2003). We control for differences in hospitals' case-mix with patients' self-reported pre-surgery health and a comprehensive set of comorbid conditions. Pre-surgery measures of functional status and pain allow us to adjust more thoroughly for differences in patients' ability to benefit from surgery than has been possible in previous studies. Any remaining differences between hospitals in terms of unobserved patient severity should be limited.

Finally, hospitals may be able to provide higher quality through unobserved determinants of quality that also correlate with volume, such as the presence of more experienced surgeons or nursing staff. By failing to control for these, parameter estimates in equation (1) will suffer from omitted variable bias. We address these concerns in a series of robustness checks in section 4.

3.3. A model of patient choice of hospital

To implement the empirical strategy outlined above, we estimate a conditional logit model of patient choice of hospital (McFadden, 1974). We include in the choice set all public and private hospitals that treat NHS patients. The sample includes the whole population of planned hip replacement patients who had surgery, regardless of whether or not they participated in the PROM survey. The utility of patient i choosing hospital h can be written as:

$$u_{ih} = V_{ih} + v_{ih} , \tag{2}$$

where V_{ih} is the utility of patient *i* derived from observed characteristics of hospital h and v_{ih} is the unobserved utility. We specify V_{ih} as:

$$V_{ih} = \gamma_1 d_{ih} + \gamma_2 d_{ih}^2 + \gamma_3 d_{ih}^3 + \gamma_4 close_{ih} + \boldsymbol{z}_{\boldsymbol{h}}'\boldsymbol{\gamma}_{\boldsymbol{z}} + \sum_{k=1}^{K} x_{ik}\left(\gamma_{1k} d_{ih} + \gamma_{2k} d_{ih}^2 + \gamma_{3k} d_{ih}^3\right), \tag{3}$$

where d_{ih} represents the distance between patient i and hospital h, measured as the straight-line distance between a hospital's postcode and the centroid of the patient's LSOA of residence, and γ_1 is the associated (dis)utility of travel. We include quadratic and cubic terms of distance to allow for a non-linear effect on patient's choice utility. We add a dummy variable, $close_{ih}$, to capture the utility of avoiding any excess travel past the closest hospital. The vector $\boldsymbol{z}_{\boldsymbol{h}}$ consists of dummy variables for

[10] There is an analogy between our method and previous instrumental variable (IV) strategies (Hentschker and Mennicken, 2018; Gaynor, Seider and Vogt, 2005) because both rely on the exogeneity of patient's distance to the hospitals. However, the conditional logit model allows for non-linear effects whereas the first stage in an IV strategy is estimated by OLS and thus assumes linearity in the parameters.

[11] This is in contrast to certain acute conditions that require immediate treatment, e.g. heart attacks or strokes, where delays in access to care are likely to result in negative health outcomes.

hospital characteristics (i.e. NHS Foundation trusts, specialist (orthopaedic) hospital, teaching hospital, private hospital) as well as the number of hospitals (sites) within a trust and whether the hospital is a treatment centre. Hospital groups (trusts) may direct their patients to a specific hospital (site). Treatment centres typically do not admit complex patients. We therefore control for these two admission restrictions. We add interaction terms between all the distance terms and $\boldsymbol{x} = (x_{ik}, k = 1,..,K)$, a vector of K patient characteristics (age, sex, socio-economic status, Elixhauser comorbidities, and whether the patient lives in a rural area[12]) as the impact of distance on hospital choice also depends on patients' socioeconomic and clinical factors. Standard errors are clustered at the family physician practice level to account for correlation in hospital choice across patients of the same practice.

Assuming that the unobserved utility terms v_{ih} are iid extreme-value (Train, 2009), the probability that patient i chooses hospital h can be estimated by maximum likelihood and is given by:

$$\hat{p}_{ih} = \frac{exp(\hat{V}_{ih})}{\sum_{h' \epsilon M_i} exp(\hat{V}_{ih'})}, \quad (4)$$

where M_i is the patient choice set containing patient i's fifty closest hospitals. The predicted volume of hospital h is equal to the sum of the estimated probabilities $\hat{p}_{ih}$ across all patients of choosing hospital h:

$$\widehat{vol}_h = \sum_{i=1}^{N} \hat{p}_{ih} = \sum_{i=1}^{N} \frac{exp(\hat{V}_{ih})}{\sum_{h' \epsilon M_i} exp(\hat{V}_{ih'})}. \quad (5)$$

We estimate equation (4) for the whole sample of planned primary hip replacement patients in England in the financial year 2015/16. This results in 59,833 patients, after exclusion of patients given age, uncommon surgeries or missing variables.

Appendix Tables A1, A2 and Figure A2 present summary statistics for the choice model sample.[13] Predicted hospital volumes are less dispersed than observed volumes (Figure A3, Appendix). We use the same sample restrictions to compute observed volumes to ensure that both predicted and observed volumes sum up to the same total patient population.[14] The correlation coefficient between both measures of volumes in our estimation sample is 0.59 (p<0.01). In specifications with predicted volumes, we bootstrap standard errors (100 replications) to account for the fact that the predicted volumes are *generated* in a first stage choice model.[15]

Conditional logit models have the advantage that they are tractable and computationally simple. These properties however rely on the assumption of independent error terms. If this holds, estimated coefficients are invariant to which alternatives/choices are available (independence of irrelevant alternatives (IIA)). We deliberately *omit* hospital quality from our model specification, thus creating a potential correlation in the error terms. The IIA property of logit models is problematic in forecasting exercises (i.e. when forecasting the demand for a new alternative) as it imposes strong restrictions on substitution behaviours. However, it is considered less crucial when estimating average aggregate preferences (Train, 2009, p. 36). Our model is therefore an approximation of

[12] The geographical information for lower super output areas (LSOAs) comes from the Office of National Statistics.
[13] Estimated coefficients from the choice model are available upon request.
[14] The observed volumes with and without these sample restrictions have a correlation coefficient of 0.98.
[15] Model-based standard errors do not account for sampling variation in the predicted volumes, which may lead to downward-biased standard errors (Murphy and Topel, 1985).

patients' demand for hospitals, if they were to ignore hospital quality considerations. We re-estimate our model with varying sets of alternatives (comprising the 50, and 10 closest hospitals in patients' choice set). Hospitals' predicted volumes under both specifications are highly correlated (Pearson correlation coefficient= 0.97), suggesting that any potential violation of the IIA assumption does not affect our results.

4. Results

4.1. Descriptive statistics

In our sample, hospitals treat on average 237 hip replacements patients annually, ranging from 21 to 1,238 surgeries. Sixty per cent of hospitals are NHS Foundation trusts (Table 1). Teaching hospitals and specialist orthopaedic hospitals account respectively for 23% and 2% of hospitals.

Table 1. Hospital characteristics, for NHS hospitals in 2015/16 in England

	Mean	Std. Dev.	Min.	Max.
Observed volume	236.72	170.67	21	1238
Foundation trust	0.60	0.49	0	1
Teaching hospital	0.23	0.42	0	1
Specialist hospital	0.02	0.13	0	1
N	178			

Notes: Volume is the number of annual planned primary hip replacements per hospital.

Figure 1 shows the unadjusted relationship between OHS health gains and hospital volumes, suggesting a small positive association between hospital volumes and surgery health gain.

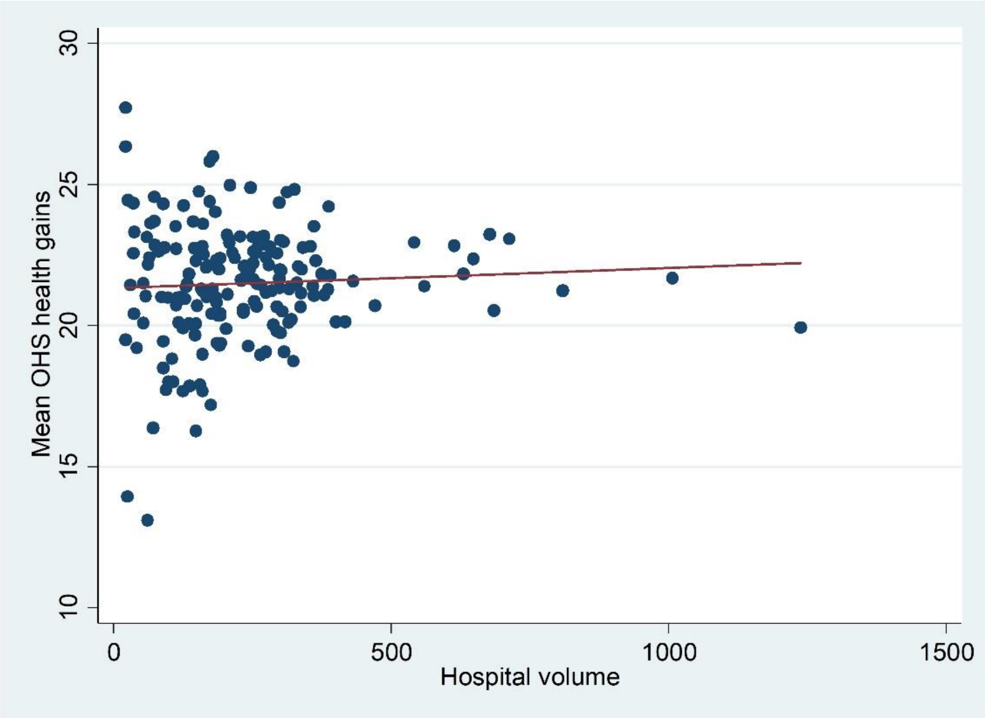

Figure 1. Association between hospital volume and Oxford Hip Score (OHS) health gain

Notes: Plot of the average OHS health gains (post-surgery minus pre-surgery OHS, unadjusted for other patient characteristics) per hospital against hospital volumes.

Table 2 presents descriptive statistics of patient characteristics. The average pre-surgery OHS is 17 points, and patients gain on average 22 points (from 17 to 39) six months after surgery. Patients are on average 70 years old and 40 percent of our sample are male. On average, patients report slightly over one (1.37) Elixhauser comorbidity. The large majority of patients (70%) report having hip-related symptoms for between one and five years.

Table 2. Patient characteristics for hip replacement: NHS patients in England in 2015/16

	Mean	Std. Dev.	Min.	Max.
Post-surgery OHS	39.11	8.74	0	48
Pre-surgery OHS	17.41	8.01	0	48
Age	70.13	8.97	50	98
Male patient	0.40	0.49	0	1
Elix. comorbidity count	1.37	1.21	0	8
Index of multiple deprivation (IMD):				
1st quintile	0.25	0.43	0	1
2nd quintile	0.25	0.43	0	1
3rd quintile	0.21	0.41	0	1
4th quintile	0.17	0.38	0	1
Most deprived quintile	0.12	0.32	0	1
Ethnicity: white	0.90	0.31	0	1
Surgery on the other hip	0.15	0.36	0	1
Diagnosed with osteoarthritis	0.97	0.17	0	1
Hybrid prosthetic replacement	0.24	0.42	0	1
Self-reported disability	0.59	0.49	0	1
Self-reported depression	0.09	0.28	0	1
Received assistance completing questionnaire	0.06	0.24	0	1
Symptoms duration:				
<1year	0.13	0.33	0	1
1-5years	0.70	0.46	0	1
6-10years	0.12	0.32	0	1
>10years	0.06	0.23	0	1
Living arrangements:				
Lives alone	0.28	0.45	0	1
Lives with family	0.72	0.45	0	1
Other	0.01	0.08	0	1
N	18979			

Notes: The IMD is calculated for small residence areas (LSOAs) in England. The Oxford Hip Scores (OHS) range from the worst reported health state (=0) to the best (=48) and are collected for each patient shortly before and six months after the operation. Variables may not add up exactly to 100% due to rounding of numbers.

Table 3 presents summary statistics of the surgeon characteristics used in Section 4.3. On average, a surgeon treats 60 planned hip replacement patients per year. Ninety nine per cent of orthopaedic surgeons are male. Sixty four per cent of the surgeons are trained in the UK and have on average 25 years of experience since their primary medical qualification. We exclude two surgeon outliers who report around 400 annual cases or more.

Table 3. Summary statistics for surgeon characteristics

	Mean	Std. Dev.	Min.	Max.
Surgeon volume	60.17	41.68	10	262
Male	0.99	0.12	0	1
Years since qualification	25.23	7.48	10	45
Qualified in the UK	0.64	0.48	0	1
N	938			

Notes: Surgeon volume comprises all patients treated in all hospitals (if the surgeon holds multiple appointments). We exclude volume outliers (400 annual cases or above). Years since qualification is the time since primary medical qualification, after which surgeons follow additional specialty training.

4.2. Regression results

Table 4 provides the results for the OLS regression with observed hospital volumes, indicating a positive and statistically significant association between hospital volume and health. Relative to patients treated in hospitals with more than 300 hip replacement cases per year (the reference group with highest volume), patients treated in hospitals with less than 150 cases (with lowest volume) are estimated to gain 0.9 (0.923) fewer OHS points. Patients treated in hospitals performing between 150 and 200 cases are estimated to gain 0.7 (0.686) fewer OHS points. Both estimates are statistically significant at the 5% level. There is no statistically significant effect of hospital volumes of 200-300 patients relative to the base category. The volume-outcome association is therefore weakly monotonic. A change in OHS is considered clinically meaningful if above four points (Varagunam, Hutchings and Black, 2015b). Therefore, the estimated association is quantitatively small, as it accounts for between around 18 percent (0.7 points) and 23 percent (0.9 points) of such clinical change.

The coefficients on patient characteristics are all statistically significant. Patients with higher pre-surgery health (OHS score) also have higher health after surgery. Older patients tend to have worse outcomes and male patients tend to report slightly better outcomes. Our results suggest the existence of a socioeconomic gradient. More deprived patients have worse outcomes than less deprived ones. Having one more Elixhauser comorbidity leads to a reduced post-surgery OHS score by about 0.5 (0.546) points. Self-reported depression, disability, or help completing questionnaires are negatively associated with post-surgery outcomes by two OHS points or above. Among the hospital-level covariates, only specialist orthopaedic trusts are associated with better health outcomes, even though the difference is small (around 0.6 OHS points).

Table 4. Results from OLS regression with observed volumes

	Post-surgery OHS	
	Coefficient	**SE**
Volume [Ref. ≥300]		
<150 cases	-0.923**	(0.293)
150-200 cases	-0.686*	(0.331)
200-300	0.150	(0.200)
Pre-surg. OHS [Ref. 0-6 pts]		
6-12 pre-surg. OHS	3.015***	(0.374)
12-18 pre-surg. OHS	4.789***	(0.371)
18-24 pre-surg. OHS	5.857***	(0.378)
24-30 pre-surg. OHS	6.469***	(0.409)
30-36 pre-surg. OHS	7.339***	(0.422)
36-42 pre-surg. OHS	8.154***	(0.464)
42-48 pre-surg. OHS	7.958***	(1.254)
Age [Ref. 50-59 years]		
60-69 years	0.329	(0.174)
70-79 years	-0.622**	(0.211)
80-89 years	-1.241***	(0.259)
90-105 years	-1.530*	(0.673)
Male	0.808***	(0.133)
Ethnic group [Ref. white]		
Other ethnic group	-2.695**	(0.872)
Ethnicity not coded	1.319***	(0.193)
Deprivation index IMD [Ref. 1st quintile]		
2nd quintile	-0.320	(0.163)
3rd quintile	-0.696***	(0.175)
4th quintile	-1.625***	(0.206)
Most deprived 5th quintile	-2.827***	(0.242)
Surgery on the other hip	-0.629***	(0.177)
Elixhauser comorbidity count	-0.546***	(0.059)
Diagnosed with osteoarthritis	1.678***	(0.436)
Hybrid prosthetic replacement	0.556***	(0.151)
Self-reported disability	-2.174***	(0.130)
Self-reported depression	-2.572***	(0.267)
Received assistance in filling questionnaire	-2.744***	(0.309)
Symptoms duration [Ref. <1 year]		
1 to 5 years	-0.793***	(0.168)
6 to 10 years	-1.377***	(0.223)
More than 10 years	-2.063***	(0.322)
Living arrangements [Ref. alone]		
Lives with family/friends	0.510***	(0.151)
Living arrangements: other	0.095	(0.882)
Teaching hospital	-0.092	(0.234)
Specialist hospital	0.597*	(0.286)
Foundation trust	0.035	(0.172)
R^2	0.175	
Observations	18979	

Notes: In parentheses, robust standard errors clustered on hospitals. IMD is measured at the patient's residence level.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5 compares the regression results for observed hospital volumes versus predicted hospital volumes. The covariate coefficients are similar in both specifications. We therefore only present the coefficients for hospital characteristics and pre-surgery health. The volume coefficients for the predicted (exogenous) hospital volumes are smaller and no longer statistically significant even though the precision of the estimates (i.e. the standard errors) is essentially unchanged. This suggests that hospitals with higher quality attract more patients, thus creating a spurious positive relation between health outcomes and volumes. After accounting for reverse causality, hospital volumes are no longer associated with improved health outcomes. The coefficient for specialist hospitals is similar in magnitude but less precisely estimated, and no longer statistically significant (Table 5, column 2).

Table 5. Results with observed and predicted hospital volumes

	Observed volumes		Predicted volumes	
	Post-surgery OHS (1)		Post-surgery OHS (2)	
	Coefficient	**SE**	**Coefficient**	**SE**
Volume [Ref. ≥300]				
<150 cases	-0.923^{**}	(0.293)	-0.168	(0.252)
150-200 cases	-0.686^{*}	(0.331)	-0.352	(0.318)
200-300 cases	0.150	(0.200)	-0.302	(0.178)
Pre-surg. OHS [0-6 pts]				
6-12 pre-surg. OHS	3.015^{***}	(0.374)	3.025^{***}	(0.395)
12-18 pre-surg. OHS	4.789^{***}	(0.371)	4.795^{***}	(0.382)
18-24 pre-surg. OHS	5.857^{***}	(0.378)	5.867^{***}	(0.379)
24-30 pre-surg. OHS	6.469^{***}	(0.409)	6.483^{***}	(0.414)
30-36 pre-surg. OHS	7.340^{***}	(0.422)	7.377^{***}	(0.429)
36-42 pre-surg. OHS	8.154^{***}	(0.464)	8.167^{***}	(0.516)
42-48 pre-surg. OHS	7.959^{***}	(1.254)	7.958^{***}	(1.204)
Teaching hospital	-0.092	(0.234)	-0.077	(0.261)
Specialist hospital	0.591^{*}	(0.288)	0.568	(0.449)
Foundation trust	0.035	(0.172)	0.078	(0.179)
Patient controls	*Yes*		*Yes*	
R^2	0.175		0.174	
Observations	18979		18979	

Notes: The same covariates as in Table 4 are included. In parentheses, robust standard errors clustered on hospitals (column 1) or bootstrapped (column 2) to account for the use of generated regressors.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We test the sensitivity of our results to different functional forms of the volume variable (linear-quadratic, square root and log terms). Our results (in the Appendix, Table A3) are unchanged, but the coefficients for predicted volumes are less precisely estimated. Results are also robust to the exclusion of (observed) volume outliers (i.e. hospitals treating over 1,000 annual cases).

4.3. Testing for the effect of surgeon volume

Volume-outcome effect may be due to hospital factors and/or surgeon effects (e.g. practice-makes-perfect in surgical skills). To test whether our results may be driven by surgeon effects, we run the same models with surgeons' volume and characteristics (i.e. gender, years since graduation as a proxy for seniority and trained in the UK).

Our strategy of predicting hospital volumes based on distance cannot be extended to surgeon volumes because patients would travel the same distance to surgeons within the same hospital. We argue that selective referral to high-quality surgeons should be limited, given that little information was available on surgeon performance during our study period. In 2015/16, online statistics were limited to a surgeon's 90-day mortality rate after primary hip replacement (NHS Commissioning Board, 2012), and reported that *all* surgeons were in line with expectations[16] (Varagunam, Hutchings and Black, 2015a). In addition, we compare our results to estimates from a specification with hospital fixed effects (Table 6, column 5). The latter control for unobserved hospital quality and thus account for patients' sorting into high-quality hospitals. However, we cannot completely exclude that surgeon volumes are endogenous, e.g. if allocation of patients to surgeons within a hospital is not random. We, therefore, do not claim causality of the surgeon volume effect.

Table 6 shows the results of a regression with surgeon volumes. The quantitative effect is small. An increase of 10 cases in the annual volume of the surgeon is associated with around 0.06 additional OHS points, equivalent to 1.5% of a clinically minimal important difference (four OHS points). The coefficient for surgeon volumes is stable across specifications (Table 6, columns 1-4), including for the specification with hospital fixed effects (Table 6, column 5). Hospitals with small volumes of hip patients, i.e. under 150 cases a year, are associated with worse health outcomes, even though the effect is smaller once we control for surgeon volumes and not significant for larger hospitals (Table 6, column 1). Adding surgeon characteristics does not change the overall results (Table 6, column 2). The number of years since graduation is associated with slightly worse outcomes, possibly because older surgeons are less familiar with the medical state-of-the-art knowledge. In an alternative specification, we included surgeon volume with a linear and a quadratic term but the latter was not significant.

Overall, the results in Table 6 confirm that predicted hospital volume has no effect on health gains even after controlling for surgeon volume (Table 6, columns 3 and 4).

[16] This is also stated on the NJR website to inform patient choice for current data: http://www.njrsurgeonhospitalprofile.org.uk/FAQ#10.

Table 6. Regression results with surgeon volumes

	Observed hospital volumes		Predicted hospital volumes		Hospital FEs
	(1)	(2)	(3)	(4)	(5)
Volume [Ref. ≥300]					
<150 cases	-0.679* (0.291)	-0.594* (0.291)	-0.058 (0.279)	-0.004 (0.270)	
150-200 cases	-0.466 (0.317)	-0.391 (0.308)	-0.101 (0.294)	-0.061 (0.285)	
200-300	0.286 (0.204)	0.331 (0.207)	-0.155 (0.201)	-0.128 (0.204)	
Teaching hospital	-0.074 (0.239)	-0.102 (0.229)	-0.099 (0.227)	-0.141 (0.220)	
Specialist hospital	0.548 (0.315)	0.489 (0.287)	0.502 (0.489)	0.433 (0.463)	
Foundation trust	-0.013 (0.172)	-0.038 (0.170)	0.004 (0.190)	-0.028 (0.184)	
Surgeon volume (10s)	0.058*** (0.016)	0.058*** (0.016)	0.060*** (0.016)	0.059*** (0.016)	0.051*** (0.016)
Male surgeon		0.635 (0.681)		0.739 (0.752)	0.032 (0.611)
Qualified in the UK		0.227 (0.159)		0.220 (0.162)	0.046 (0.156)
Years since qualification		-0.035** (0.010)		-0.036*** (0.011)	-0.035*** (0.009)
Patient controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Hospital fixed effects	*No*	*No*	*No*	*No*	*Yes*
R^2	0.177	0.178	0.176	0.177	0.164
Observations	18470	18470	18470	18470	18470

Notes: In parentheses, robust standard errors clustered on hospitals (columns 1, 2 and 5) or bootstrapped (columns 3, 4) to account for the use of generated regressors. Sample size is smaller due to missing surgeon characteristics or exclusion of surgeon outliers. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Some surgeons in public hospitals may also treat privately-funded patients in private (independent sector) hospitals. For these surgeons, our measure of volume is smaller than the total carried out across public and private hospitals. We run the same analysis on the subsample of surgeons (around 40%, N=363) whom we observe to work only in NHS hospitals.[17] Our results (Table A4 in the Appendix) are virtually unchanged. The magnitude of the surgeon effect is slightly higher (between 0.082 and 0.096), which is consistent with the hypothesis that surgeon volumes are under-reported in the whole sample.

4.4. Placebo test

Hospitals may provide better care through unobserved determinants of quality that correlate with volume. Our set of hospital-level controls together with the exclusion of private hospitals mitigates the risk of systematic quality differences (e.g. linked to ownership type). In addition, we run a placebo test where we regress patient outcomes (as measured by the EQ-5D instrument) after groin

[17]To increase our confidence that these surgeons do not work for private hospitals, we restrict the sample to surgeons who have had no HES patient records in private hospitals over the past five years. While private hospitals represent a substantial share of NHS funded care (i.e. in 2015/16, close to one third of NHS-funded planned hip replacement patients were treated in private hospitals in our data), privately-funded hip replacements across all Hospitals accounted for less than 13% of the total hip replacement volume in 2010/11 according to Kelly and Stoye (2016). The likelihood that surgeons working in private hospitals only treat private patients is therefore low, and any remaining unobserved volumes would be small in magnitude.

hernia repair on hip replacement volumes. Hernia repair is performed by surgeons working in the clinical specialty of general surgery, whereas hip replacements are performed by orthopaedic surgeons. The quality of hernia repair surgeries should therefore not be affected by hip replacement volumes, except if the latter also reflect unobserved aspects of hospital quality (e.g. the quality of infrastructure or general nursing input). Results (in Appendix, Table A5) show that hip volumes do not affect hernia health outcomes (the volume coefficient for the lowest volume category is -0.002 (p-value= 0.008)).

5. Conclusions

This study has investigated the causal effect of hospital volumes on health gains, as measured by patient-reported outcomes, for planned hip replacement surgery in public hospitals in England. Our key finding is that there is a clinically small and positive association between observed hospital volume and health outcomes but this disappears once we adjust for volume endogeneity due to reverse causality (i.e. hospitals with higher quality attract more patients).

In the analysis, we cannot exclude the possibility that most hospitals may already be operating at the flat end of the volume-outcome curve, despite being able to observe hospitals with low volumes (the lowest volume category starts at 20 cases per year). However, the fact that we do find a positive association between outcomes and observed volumes suggests that there remains room for improvement.

Our findings have policy implications. They suggest that concentrating the provision of planned hip replacements in the English healthcare market would not result in improvements in health outcomes. Given that the hospital market in England is already concentrated, further concentration may have adverse effects on patient access to care. This would be especially concerning in the context of hip replacement, given that older patients place a larger value on proximity to home when selecting providers (Gutacker et al., 2016).

Our results also suggest that the hospital-level association can be driven by a hospital demand's responsiveness to quality rather than economies of scale. This shows the importance of accounting for volume endogeneity in volume-outcome studies for planned procedures whose results may otherwise be biased. Our methodology to deal with volume endogeneity can be applied to other hospital markets and healthcare procedures.

Appendix

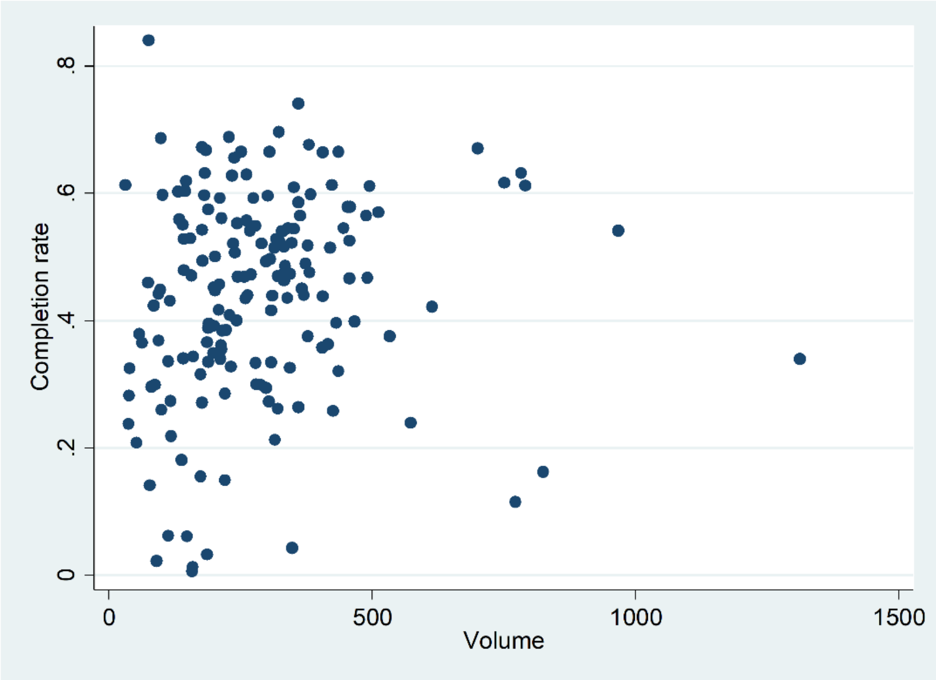

Figure A1. PROMs completion rate against hospital volumes

Notes: Completion rate (both pre- and post-surgery PROM questionnaires are filled in) per hospital against hospital volume. The Pearson correlation coefficient is 0.13 and is not statistically significant (p-value= 0.08).

Table A1. Summary statistics of sample used for hospital patient choice model (patients)

	Mean	Std. Dev.	Min.	Max.
Patient age	69.99	9.19	50	101
Male patient	0.39	0.49	0	1
# Elixhauser conditions	1.33	1.21	0	8
Index of multiple deprivation:				
1st quintile	0.25	0.43	0	1
2nd quintile	0.25	0.43	0	1
3rd quintile	0.22	0.41	0	1
4th quintile	0.17	0.37	0	1
5th quintile	0.12	0.32	0	1
Patient living in rural area	0.27	0.45	0	1
Distance to chosen hospital (km)	13.14	13.07	0	154
Patient choosing closest hospital	0.42	0.49	0	1
N	59833			

Table A2. Summary statistics of sample used for hospital patient choice model (providers)

	Mean	Std. Dev.	Min.	Max.
NHS Treatment Centre (TC) site	0.03	0.18	0	1
Teaching trust	0.14	0.35	0	1
Specialist trust	0.01	0.10	0	1
# hospitals within trust	1.60	1.05	1	5
Provider type:				
Independent sector TC	0.08	0.28	0	1
Independent sector non-TC	0.32	0.47	0	1
NHS Foundation trust (FT)	0.35	0.48	0	1
NHS non-FT	0.25	0.43	0	1
N	312			

Notes: Statistics are calculated for the total sample of planned hip replacements in England in 2015/16, after exclusion of hip admissions for revision surgeries, patients below 50 and providers with less than 20 cases a year.

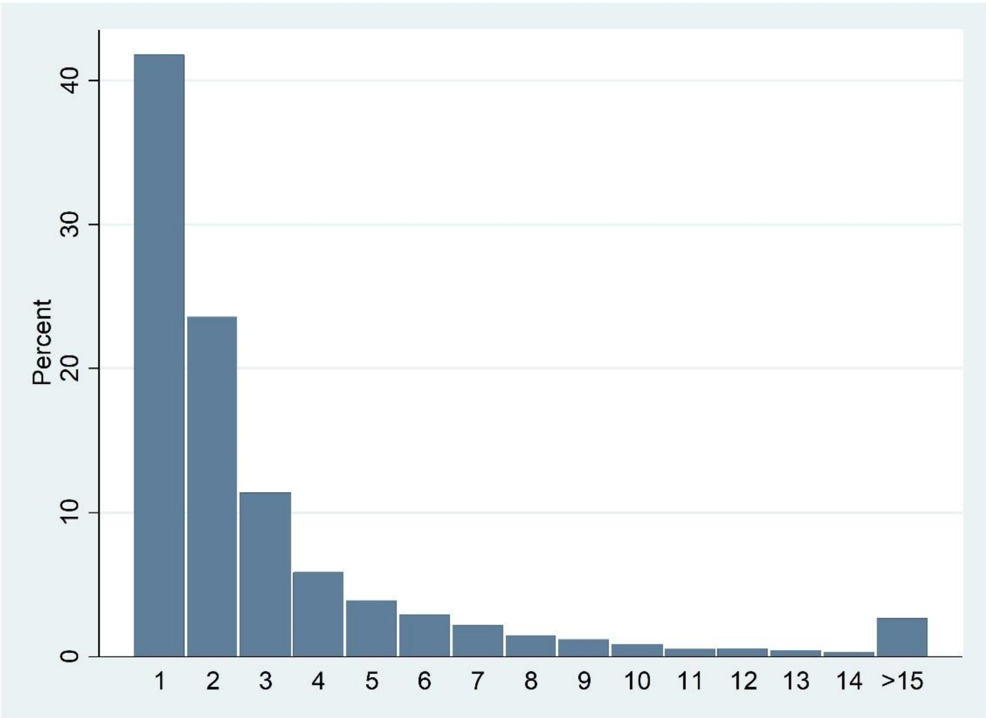

Figure A2. Percentage of planned hip replacement patients who went to their Nth nearest provider
Notes: This corresponds to the total sample of 59,833 planned primary hip replacements in England in 2015/16 over which the multinomial logit model of hospital choice is run.

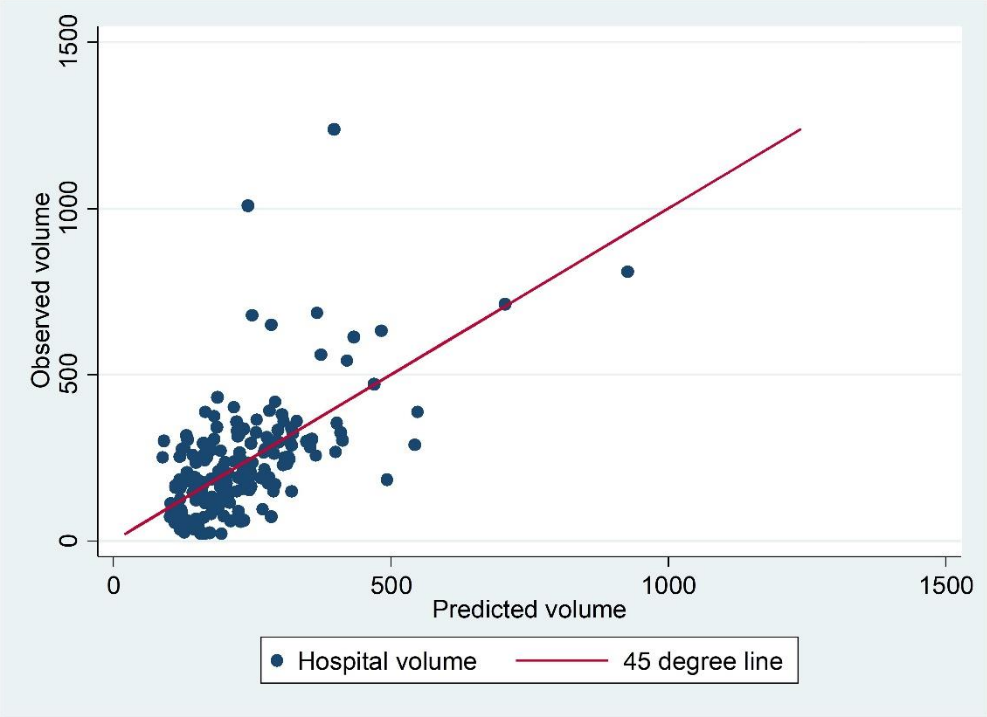

Figure A3. Scatterplot of the observed against the predicted hospital volumes

Notes: Hospital volumes are the number of planned primary (non-revision) patients treated by a hospital in 2015/16 in the English NHS. The predicted volumes are constructed using a conditional logit model of hospital choice. The data points are the hospitals in the estimation sample.

Table A3. Results with different functional forms for volumes

	Observed hospital volumes			Predicted hospital volumes		
	(1)	(2)	(3)	(4)	(5)	(6)
Volume (in 100s)	0.323*			0.272		
	(0.133)			(0.296)		
Volume^2	-0.023*			-0.024		
	(0.010)			(0.042)		
Volume ($\sqrt{vol}$)		0.035*			0.033	
		(0.016)			(0.025)	
Log(volume)			0.375*			0.269
			(0.148)			(0.215)
Patient controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Hospital controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
R^2	0.174	0.174	0.174	0.174	0.174	0.174
Observations	18979	18979	18979	18979	18979	18979

Notes: Volume is in 100s (1 unit corresponds to 100 patients). In parentheses, robust standard errors clustered on hospitals (columns 1-3) or bootstrapped (columns 4-6) to account for the use of generated regressors.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4. Results on a subsample of surgeons who work for public hospitals only

	Observed volumes	Predicted volumes
	(1)	(2)
Hospital volume *[Ref. ≥300]*		
<150 cases	-1.105* (0.478)	-0.792 (0.426)
150-200 cases	-0.443 (0.396)	-0.138 (0.398)
200-300	0.152 (0.288)	-0.352 (0.276)
Teaching hospital	-0.110 (0.329)	-0.040 (0.300)
Specialist hospital	0.325 (0.382)	0.307 (0.569)
Foundation trust	-0.101 (0.261)	-0.130 (0.260)
Surgeon volume (in 10s)	0.082* (0.037)	0.096* (0.040)
Male surgeon	0.030 (0.918)	0.023 (0.986)
Qualified in the UK	0.510 (0.281)	0.454 (0.257)
Years since qualification	-0.009 (0.016)	-0.015 (0.015)
Patient controls	*Yes*	*Yes*
R^2	0.173	0.173
Observations	6783	6783

Notes: Results for subsample of surgeons working for NHS hospitals only. In parentheses, robust standard errors clustered on hospitals (column 1) or bootstrapped (column 2) to account for the use of generated regressors.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5. Results of a placebo test of the effect of hip volumes on health gains for groin hernia patients

	Observed volumes		Predicted volumes	
	Post-surgery EQ5D		Post-surgery EQ5D	
	(1)		(2)	
	Coefficient	SE	Coefficient	SE
Hip replacement volume [Ref.>300]				
<150 cases	-0.002	(0.008)	-0.003	(0.007)
150-200 cases	0.001	(0.006)	0.009	(0.007)
200-300 cases	0.001	(0.006)	0.011	(0.006)
Pre-surg. EQ5D Index [Ref. -0.594 to 0]				
0-0.5 EQ5D	0.324***	(0.044)	0.324***	(0.039)
0.5-0.75 EQ5D	0.371***	(0.044)	0.371***	(0.038)
0.75-1 EQ5D	0.466***	(0.044)	0.466***	(0.038)
Age [Ref. 20-29 years]				
30-39 years	-0.032	(0.022)	-0.032	(0.020)
40-49 years	-0.032	(0.018)	-0.032	(0.017)
50-59 years	-0.018	(0.017)	-0.018	(0.016)
60-69 years	-0.004	(0.017)	-0.004	(0.016)
70-79 years	0.001	(0.016)	0.002	(0.015)
80-89 years	0.003	(0.018)	0.004	(0.016)
90-105 years	-0.018	(0.035)	-0.018	(0.035)
Male	0.038***	(0.008)	0.038***	(0.010)
Ethnic group [Ref. White]				
Other ethnic group	-0.026*	(0.011)	-0.025*	(0.010)
Ethnicity not coded	0.006	(0.006)	0.005	(0.006)
Deprivation Index [Ref. 1st quintile]				
2nd quintile	-0.001	(0.005)	-0.001	(0.005)
3rd quintile	-0.006	(0.006)	-0.006	(0.006)
4th quintile	-0.018**	(0.006)	-0.017**	(0.006)
Most deprived quintile	-0.043***	(0.008)	-0.044***	(0.008)
Laparoscopic hernia	0.012**	(0.004)	0.011*	(0.005)
Elixhauser comorbidity count	-0.019***	(0.003)	-0.019***	(0.003)
Hypertension	0.015*	(0.006)	0.014**	(0.005)
Self-reported disability	-0.156***	(0.009)	-0.156***	(0.009)
Self-reported depression	-0.081***	(0.014)	-0.081***	(0.014)
Assistance in filling questionnaire	-0.041***	(0.012)	-0.041***	(0.011)
Teaching hospital	-0.002	(0.004)	-0.001	(0.004)
Foundation trust	0.004	(0.005)	0.004	(0.005)
R^2	0.308		0.308	
Observations	8250		8250	

Notes: Results from a regression of hip replacement volumes on hernia repair outcomes for groin hernia patients. Outcomes are calculated using EQ5D indexes before and after-surgery collected in PROM data. In parentheses, robust standard errors clustered on hospitals (column 1) or bootstrapped (column 2) to account for the use of generated regressors.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Avdic, D., Lundborg, P. and Vikström, J. (2019). Estimating returns to hospital volume: evidence from advanced cancer surgery. *Journal of Health Economics,* 63, 81-99.

Bauer, H. and Honselmann, K. (2017). Minimum volume standards in surgery - Are we there yet? *Visceral Medicine,* 33(2), 106-116.

Birkmeyer, J. and Dimick, J. (2004). Potential benefits of the new Leapfrog standards: effect of process and outcomes measures. *Surgery,* 135(6), 569-575.

Blanco, B., et al. (2017). “Take the volume pledge” may result in disparity in access to care. *Surgery,* 161(3), 837-845.

Brealey, S., et al. (2012). The effect of waiting times from general practitioner referral to MRI or orthopaedic consultation for the knee on patient-based outcomes. *The British Journal of Radiology,* 85(1019), e1134.

Brekke, K., et al. (2014). Patient Choice, Mobility and Competition Among Health Care Providers. In: Levaggi, R. and Montefiori, M. eds. *Health Care Provision and Patient Mobility: Health Integration in the European Union*. Milano: Springer Milan, pp. 1-26.

Brekke, K., Siciliani, L. and Straume, O. (2017). Hospital mergers with regulated prices. *Scandinavian Journal of Economics,* 119(3), 597-627.

Chowdhury, M., Dagash, H. and Pierro, A. (2007). A systematic review of the impact of volume of surgery and specialization on patient outcome. *British Journal of Surgery Society,* 94(2), 145-161.

Com-Ruelle, L., Or, Z. and Renaud, T. (2008). The volume-outcome relationship in hospitals: lessons from the literature. *Questions d'économie de la santé* [Online], September 2008. Available at: https://www.irdes.fr/EspaceAnglais/Publications/IrdesPublications/QES135.pdf [Accessed 02/05/2018].

Cooper, Z., Gibbons, S. and Skellern, M. (2018). Does competition from private surgical centres improve public hospitals' performance? Evidence from the English National Health Service. *Journal of Public Economics,* 166, 63-80.

Dawson, J., et al. (1996). Questionnaire on the perceptions of patients about total hip replacement. *Journal of Bone and Joint Surgery,* 78(2), 185.

De Vries, L., et al. (2011). Complications after hip arthroplasty and the association with hospital procedure volume: a nationwide retrospective cohort study on 50,080 total hip replacements with a follow-up of 3 months after surgery. *Acta Orthopaedica,* 82(5), 545-552.

Department of Health (2012). *A methodology for applying casemix adjustment - Annex C: coefficients for hip replacement models*. Retrieved from https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms/proms-methodologies#case-mix.

Elixhauser, R., et al. (1998). Comorbidity measures for use with administrative data. *Medical Care,* 36(1), 8-27.

Ferguson, B., Posnett, J. and Sheldon, T. (1997). *Concentration and choice in the provision of hospital services. Summary Report*. University of York, York.

Gaynor, M., Moreno Serra, R. and Propper, C. (2013). Death by market power: reform, competition, and patient outcomes in the National Health Service. *American Economic Journal-Economic Policy,* 5(4), 134-166.

Gaynor, M., Seider, H. and Vogt, W. (2005). The volume-outcome effect, scale economies, and learning-by-doing. *American Economic Review,* 95(2), 243-247.

Gaynor, M. and Town, R. (2011). Competition in Health Care Markets. In: Pauly, M., McGuire, T. and Barros, P. eds. *Handbook of Health Economics*. Vol. 2. Elsevier, pp. 499-637.

Geweke, J., Gowrisankaran, G. and Town, R. (2003). Bayesian inference for hospital quality in a selection model. *Econometrica,* 71(4), 1215-1238.

Gowrisankaran, G. and Town, R. (2003). Competition, payers, and hospital quality. *Health Services Research,* 38(6p1), 1403-1422.

Gravelle, H., Santos, R. and Siciliani, L. (2014). Does a hospital's quality depend on the quality of other hospitals? A spatial econometrics approach. *Regional Science and Urban Economics,* 49, 203-216.

Gutacker, N., et al. (2017). Hospital surgical volumes and mortality after coronary artery bypass grafting: using international comparisons to determine a safe threshold. *Health Services Research,* 52(2), 863-878.

Gutacker, N., et al. (2016). Choice of hospital: which type of quality matters? *Journal of Health Economics,* 50, 230-246.

Halm, E., Lee, C. and Chassin, M. (2002). Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature. *Annals of Internal Medicine,* 137(6), 511-520.

Healthcare Quality Improvement Partnership (2018). *National Joint Registry - 15th Annual Report*. London, England. Retrieved from https://www.hqip.org.uk/resource/national-joint-registry-15th-annual-report-2018/.

Hentschker, C. and Mennicken, R. (2018). The volume–outcome relationship revisited: practice indeed makes perfect. *Health Services Research,* 53(1), 15-34.

Ho, V. (2014). *Learning by Doing*. In: *Encyclopedia of Health Economics*, pp. 141-145.

Judge, A., et al. (2006). The effects of surgical volumes and training centre status on outcomes following total joint replacement: analysis of the Hospital Episode Statistics for England. *Journal of Public Health,* 28(2), 116-124.

Kelly, E. and Stoye, G. (2016). New joints: private providers and rising demand in the English National Health Service. *IFS Working Papers* [Online], (W16/15). Available at: https://www.econstor.eu/handle/10419/173949 [Accessed 22/03/2019].

Kessler, D. and McClellan, M. (2000). Is hospital competition socially wasteful? *Quarterly Journal of Economics,* 115(2), 577-615.

Kizer, K. (2003). The volume–outcome conundrum. *The New England Journal of Medicine,* 349(22), 2159-2161.

Losina, E., et al. (2004). Early failures of total hip replacement: effect of surgeon volume. *Arthritis & Rheumatism,* 50(4), 1338-1343.

Luft, H., Hunt, S. and Maerki, S. (1987). The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health Services Research,* 22(2), 157.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P. ed. *Frontiers in Economics*. Vol. 4. New York: Academic Press, pp. 105-142.

Mesman, R., et al. (2017). Evaluation of minimum volume standards for surgery in the Netherlands (2003–2017): a successful policy? *Health Policy,* 121(12), 1263-1273.

Murphy, K. M. and Topel, R. H. (1985). Estimation and Inference in Two-Step Econometric Models. *Journal of Business & Economic Statistics,* 3(4), 370-379.

NHS Commissioning Board (2012). *Everyone counts: planning for patients 2013-14*. Retrieved from https://www.jcpmh.info/resource/everyone-counts-planning-for-patients-2013-14/.

NHS Digital (2017). *A guide to PROMs methodology*. Patient Reported Outcome Measures (PROMs) in England. Retrieved from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms/proms-methodologies.

OECD (2017). *Caring for quality in health - Lessons learnt from 15 reviews of health care quality*. OECD Reviews of Health Care Quality. Paris, France. Retrieved from https://www.oecd.org/health/caring-for-quality-in-health-9789264267787-en.htm.

Ostendorf, M., et al. (2004). Patient- reported outcome in total hip replacement. A comparison of five instruments of health status. *The Journal of Bone and Joint Surgery. British volume,* 86(6), 801-808.

Paterson, J. M., et al. (2010). Provider volumes and early outcomes of primary total joint replacement in Ontario. *Canadian Journal of Surgery,* 53(3), 175-183.

Ramanarayanan, S. (2008). *Does practice make perfect: an empirical analysis of learning-by-doing in cardiac surgery*. Unpublished manuscript. Available at SSRN: http://dx.doi.org/10.2139/ssrn.1129350.

Ravi, B., et al. (2014). Relation between surgeon volume and risk of complications after total hip arhroplasty: propensity score matched cohort study. *British Medical Journal,* 348(7961), 14.

Shojania, K. and Forster, A. (2008). Hospital mortality: when failure is not a good measure of success. *Canadian Medical Association Journal,* 179(2), 153-157.

Singh, J., et al. (2011). Hospital volume and surgical outcomes after elective hip/knee arthroplasty: a risk-adjusted analysis of a large regional database. *Arthritis & Rheumatism,* 63(8), 2531-2539.

Street, A., et al. (2014). Variations in outcome and costs among NHS providers for common surgical procedures: econometric analyses of routinely collected data. *Health Services and Delivery Research,* 2(1), 1-90.

Train, K. (2009). *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge University Press, Cambridge.

Tuominen, U., et al. (2010). Is Longer Waiting Time for Total Knee Replacement Associated with Health Outcomes and Medication Costs? Randomized Clinical Trial. *Value in Health,* 13(8), 998-1004.

Varagunam, M., Hutchings, A. and Black, N. (2015a). Do patient-reported outcomes offer a more sensitive method for comparing the outcomes of consultants than mortality? A multilevel analysis of routine data. *BMJ Quality & Safety,* 24(3), 195.

Varagunam, M., Hutchings, A. and Black, N. (2015b). Relationship between patient-reported outcomes of elective surgery and hospital and consultant volume. *Medical Care,* 53(4), 310-316.