Journal of Economic Behavior and Organization 190 (2021) 704–729

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Scale economies in the health sector: The effect of hospital volume on health gains from hip replacement surgery

Laurie Rachet-Jacquet[a,b,*], Nils Gutacker[a], Luigi Siciliani[a,b]

[a] Centre for Health Economics, University of York, Heslington, York YO10 5DD, United Kingdom
[b] Department of Economics and Related Studies, University of York, Heslington, York YO10 5DD, United Kingdom

ARTICLE INFO

Article history:
Received 11 May 2020
Revised 1 August 2021
Accepted 5 August 2021

JEL-codes:
I11
I18
L25

Keywords:
Quality
Hospital
Volumes
Learning-by-doing
Patient-reported outcomes

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 between 2011 and 2015. 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 severity through an accurate measure of pre-surgery health, other patient medical and socioeconomic indicators and a rich set of hospital characteristics. 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, in line with approaches adopted in the hospital competition literature. Results from a pooled OLS model show that the observed effect of volume on health outcomes in hip replacement surgery is positive and clinically small, but no longer statistically significant once we account for the endogeneity of volume. Results from an alternative specification with hospital fixed effects further confirm that hospital volume does not have a causal effect on health outcomes.

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 or learning-by-doing effects often cited in the economics literature (Argote and Epple, 1990; Benkard, 2000; Mukoyama, 2006; Thompson, 2010; 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 (Com-Ruelle et al., 2008; Bauer and Honselmann, 2017; Mesman et al., 2017).

Despite a large literature investigating the relation between volume and health outcomes across a range of procedures and countries (Ferguson et al., 1997; Halm et al., 2002; Gutacker et al., 2017), evidence of a causal effect of volume on quality remains limited due to the potential endogeneity of volume (Luft et al., 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

* Corresponding author.
E-mail addresses: laurie.rachetjacquet@york.ac.uk (L. Rachet-Jacquet), nils.gutacker@york.ac.uk (N. Gutacker), luigi.siciliani@york.ac.uk (L. Siciliani).

https://doi.org/10.1016/j.jebo.2021.08.014
0167-2681/

(Brekke et al., 2014; Gutacker et al., 2016). Yet, understanding the 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 patients' access to care (Blanco et al., 2017). Alternatively, a higher volume of operations can lead to better outcomes through personnel effects, e.g. by increasing surgeons' technical proficiency via repetition (learning-by-doing), through fostering coordination within clinical teams (Bartel et al., 2014; Chan, 2016, 2018; Chan and Gruber, 2020; Chan, 2021), or by making it economically viable for hospitals to invest time and resources in more streamlined production processes that follow patient pathway and to invest in better infrastructure. In contrast to the selective-referral hypothesis, these mechanisms capture different forms of economies of scale.[1] In these instances, more concentrated hospital markets may lead to improved patient outcomes (Gaynor and Town, 2011; Brekke et al., 2017).

This study investigates the 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 between 2011/12 and 2015/16. Hip replacement surgery is a common planned procedure, which involves replacing the damaged part of a hip joint by an artificial one. Hip replacement is well suited to studying economies of scale given the importance of peri–operative, rehabilitation and follow-up care[2] (Reagans et al., 2005). Hospitals with higher volumes of hip replacement patients may exert effort to design better pre-surgery and discharge protocols or build up relationships with healthcare or other providers during patient care pathway (Kizer, 2003; Ho, 2014), which we measure here through hospital volume due to unavailability of more disaggregate data on pre-surgery or rehabilitation care activity. Further, unexplained variations in patient-reported outcomes after hip replacement have been reported at the hospital (Street et al., 2014) and surgeon level (Varagunam et al., 2015a), while fixation methods and implant types are associated with differences in revision rates (Healthcare Quality Improvement Partnership, 2018). Therefore, we allow for surgeons' activity to influence health outcomes, separately from hospital volume. Even modest improvements in the health gains for individual patients would sum to important gains at the health system level given the high incidence of hip surgeries in an ageing population.

To test for the presence of economies of scale in health outcomes, we use two alternative strategies. First, we run a pooled OLS model, exploiting the variation in volume across all public hospitals in England and over time (2011–2015). Variation in volumes can be driven by geographical differences in population density or the organisation of hospital services. To address possible reverse-causality bias due to higher quality increasing hospital demand, we employ 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 (Gaynor and Town, 2011; Gaynor et al., 2013; Cooper et al., 2018). We also control for a rich set of hospital variables and characteristics of the catchment area around the hospital to minimise the risk of omitted variable due to hospital related variables.

Second, we implement a hospital fixed effects model to account for remaining unobserved time-invariant hospital factors beyond our set of hospital control variables. This specification tests for the effect of increasing volumes within hospitals over time on patient health outcomes. Notice that even with hospital fixed effects, a possible concern about reverse causality remains, as hospitals that experience an increase in quality over time may experience an increase in demand.

The results from a pooled OLS model show that the effect of actual volume on health outcomes in hip replacement surgery is positive and clinically small, but no longer statistically significant once we account for the endogeneity of volume by using predicted volumes. Results from an alternative specification with hospital fixed effects further confirm that hospital volume does not have a causal effect on health outcomes, neither with actual or predicted volumes. We therefore conclude that we do not find evidence that economies of scale affect quality to support the argument for concentrating the provision of care in this setting.

Our contribution to the literature (reviewed briefly in Section 1.1) is threefold. First, we use patient-reported outcome measures (PROMs) to estimate the effect of volume in terms of improvements in patients' health status. The English NHS is one of the first healthcare systems to routinely collect these novel data, which permit an examination of the benefit of treatment as perceived by the patient. In contrast, post-operative mortality is low for planned hip replacement patients (0.06% in our data), thus rendering commonly used health outcomes (mortality or complication rates) insensitive to finer variations in quality (Shojania and Forster, 2008; Varagunam et al., 2015a). Second, the availability of rich patient-reported data on functional status collected just before the surgery ensures a more thorough control for patient severity and minimise the risk of omitted variable bias through unobserved severity (Tsai et al., 2006; Kahn et al., 2009). Third, we address the reverse causality bias by employing a measure of *predicted* volumes, rather than *actual* volumes, as outlined above, following the

[1] Rapid increases in volume may also lead to lower quality of care if less clinical time is spent with each patient (i.e. congestion effect).

[2] Patients who undergo a planned hip replacement procedure are typically referred to the hospital by their family physician (called general practitioner in the UK) or after being assessed by a musculoskeletal clinic. They have a pre-surgery assessment with nurse practitioners before being seen by an anaesthetist and operated by an orthopaedic surgeon or one of the team members. After care can be supervised by an occupational therapist or a physiotherapist (Healthcare Quality Improvement Partnership, 2018b). Because of data availability however, we are only able to observe surgeon volumes. The role of pre-surgery, rehabilitation or follow-up care personnel cannot be separately investigated from hospital volumes.

literature on hospital competition (Kessler and McClellan (2000); Gaynor and Town, 2011; Gaynor et al., 2013; Cooper et al., 2018). We also control for a rich set of hospital variables and characteristics of the catchment area around the hospital to minimise the risk of omitted variable due to hospital related variables.

By addressing these two sources of endogeneity, i.e. insufficient risk-adjustment and reverse causality, and controlling for a rich set of hospital characteristics including hospital fixed effects, 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 relation 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 (Kizer, 2003; Ho, 2014). A recent literature has stressed the importance of teamwork and peer effects in surgical settings in increasing productivity and efficiency of care (Chan, 2018; Chan and Gruber, 2020). These studies suggest that high volumes of patients can contribute to better quality of care via more effective work routines and through better coordination between nursing staff and surgeons, a better allocation of patients across surgeons, or through more frequent learning opportunities from senior colleagues in the surgical team (Reagans et al., 2005; Chan, 2016, 2021). At the surgeon level, the volume-outcome effect is more readily understood as a learning-by-doing 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 et al., 2007).

Previous causal studies have used hospital fixed effects to control for time-invariant unobserved hospital quality, using within-hospital variation in volumes over time to estimate learning effects (Hamilton and Hamilton, 1997; Ho, 2002; Sfekas, 2009). Using a quasi-natural experiment in bariatric surgery in the US, several studies exploit the increase in patient referrals to high-volume hospitals after a policy restricted coverage to hospitals with certain minimum volume for Medicare patients (Livingston, 2009; Nguyen et al., 2010; Dimick et al., 2013). Alternatively, the previous literature has exploited the geographical distribution of patients in an IV setting (Gaynor et al., 2005 for cardiac care in the US), or used the variation in volume caused by the closure/opening of surrounding clinics (Avdic et al., 2019 for cancer care in Sweden).

The limited causal literature on orthopaedic surgery suggests mixed results. Luft et al. (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. Hamilton and Hamilton (1997) find no effect of volume on in-hospital mortality for hip fracture patients in Canada, after controlling for unobserved time-invariant hospital quality with hospital fixed effects. Hentschker and Mennicken (2018) use the distribution of patients and hospital competitors around the hospital as an instrument for hospital volumes in Germany. They find that hospital volume reduces in-hospital mortality for emergency hip replacement after hip fracture.

Medical 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, in the US and in the Netherlands (Judge et al., 2006; De Vries et al., 2011; Singh et al., 2011). Previous studies find a negative relation between surgeon volumes and rate of revisions or complications after primary planned hip replacement, using data from the US (Losina et al., 2004) or Canada (Paterson et al., 2010; Ravi et al., 2014). 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 (Losina et al., 2004; Ravi et al., 2014). In a setting close to ours, Varagunam et al. (2015b) use administrative data 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.

Our study relates also to an older literature which estimates scale economies of volume on costs while controlling for quality using hospital-level data (e.g. Dranove (1998), Preyra and Pink (2006), Gaynor et al. (2015), Kittelsen et al. (2018); see Giancotti et al. (2017) for a recent review). It also relates to the literature on estimating hospital efficiency indicators through stochastic frontier techniques or Data Envelopment Analysis, which again needs to control for differences in quality to isolate differences in costs that are due to higher efficiency as opposed to lower quality (Jacobs, 2006. Other studies have instead focused on identifying a potential trade-off between providing high quality and high volumes of care with a given set of factor inputs. For example, Grieco and Mcdevitt (2017) estimate a Cobb Douglas cost function for dialysis care

[1] Our results here differ from theirs, potentially because we exclude private hospitals as their reported volumes do not include all treated patients (see Section 2.1).

in the US and identify a negative elasticity of quality to quantity of approximately −0.2. However, these studies do not explore the feedback effects from volume to quality via e.g. learning or re-organisation of inputs that we seek to uncover in our study.

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 2011 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 360,000 surgeries. To ensure sample homogeneity, we exclude revision surgeries, which are rare and more complex procedures.[2] Patients who are younger than 50 years are also excluded from the sample as they represent infrequent (i.e. approximately five percent of planned hip replacements) and atypical medical cases who require replacement of a damaged hip joint much earlier than expected given usual wear. 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 ensure that our results are estimated on a sufficient number of cases and to attenuate the risk of coding errors. Relaxing these sample restrictions based on patient age or hospital size does not affect our results (Table A1 in Appendix).

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. All eligible orthopaedic patients are asked to report through a paper-based survey their health status, functioning and health-related quality of life immediately before and six months after surgery. 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 (Dawson et al., 1996; Ostendorf et al., 2004) (see Section 2.2 below for more detail).

The PROMs and HES data are linked based on a number of identifying characteristics, including their unique NHS number (NHS Digital, 2017). 66% of all hip patient admissions are successfully matched for at least one PROM record, which corresponds to about 235,000 admission records. We discuss potential risks of attrition bias in Section 4.5.

We further exclude private hospitals from our sample. Our dataset includes all patient admissions (privately and NHS-funded) in NHS hospitals, but only admissions for NHS-funded patients in private hospitals. The observed volumes for private hospitals would therefore underestimate their actual volume of activity. The degree of measurement error depends on each hospital's unobserved volume of private patients, such that the relative distribution of the observed volumes for private hospitals will also differ from the distribution of their actual volumes. Results from a volume-outcome analysis based on the observed volumes will therefore be biased. After sample cleaning, restriction to complete PROMs records and exclusion of private hospitals, our final sample includes 105,229 patients.

Table A2 in Appendix indicates that patients in the final sample are slightly older, have more comorbidities and lower pre- and post-surgery Oxford Hip Score than in the sample after sample cleaning but before exclusion of private providers. Though the difference is quantitively small, this suggests that private providers in England treat a healthier population overall (Moscelli et al., 2018a). This would impact the external validity of our results only if the potential for scale economies varies across levels of pre-operative health. However, this also means that we focus on the upper end of the severity distribution of hip replacement patients for which health gains are more likely to occur.

2.2. Dependant 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. Patients are asked to rate the degree or frequency of pain felt ("During the past four weeks, have you had any sudden severe pain (shooting, stabbing, or spasms) from your affected hip?"), their ability to walk ("During the past four weeks, have you been limping when walking because of your hip?"), use public transportation ("Have you had any trouble getting in and out of a car or using public transportation because of your hip?"), climb stairs or do household shopping autonomously, amongst other items.[3] The OHS is the sum of the scores obtained for each item and goes from zero (worst) to 48 (best health status). The same OHS questions are distributed to the patients shortly before and six-month after surgery. We use patients' post-surgery OHS as our dependant variable but control for the pre-surgery OHS, thereby assessing patient's *health gain* from the surgery.

[2] Revision surgeries represented around 10% of total planned hip patient admissions in our data.

[3] The full questionnaire can be found online at: http://www.orthopaedicscore.com/scorepages/oxford_hip_score.html (accessed 02.04.2020).

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 each financial year from 2011/12 to 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 or 2015 index of multiple deprivation (IMD)[4] 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 in filling the questionnaire (Department of Health, 2012). We count the Elixhauser comorbidities reported in a patient's hospital stays up to one year prior to the admission for hip replacement (Elixhauser et al., 1998; Gutacker et al., 2016). We also control for whether the patient has previously undergone hip surgery on the other hip in the past year, 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 control for a large set of hospital characteristics that may be associated with higher quality (e.g. via medical expertise or better resources) independently of volumes. We include controls for hospitals' teaching status, whether the hospital is a specialist (orthopaedic) hospital[5] or a NHS foundation trust (FT) as the latter have greater financial autonomy (Gravelle et al., 2014).

Hospitals located in more affluent areas may enjoy better facilities or find it easier to recruit healthcare staff. We proxy for these exogenous geographical differences by using the market forces factor (MFF) which reflects unavoidable differences in hospital costs of labour or capital and is used to adjust hospital reimbursement tariffs.

To further ensure that we isolate the effect of hip replacement volumes on health outcomes from potential confounders, we also control for the average socio-economic and demographic characteristics of the population who lives in the hospitals' catchment area. Hospitals that serve a more frail or deprived population may have poorer outcomes, independently of volume, and hospitals may face higher demand pressure. We define the hospital catchment area as the area within 30 km of the hospital (in line with the competition literature mentioned above) and measure the proportion of over 65-year-olds and the average deprivation score of the population in the catchment area. Poorer access and availability of primary care for the population in the hospital catchment area may result in lower coordination of care in the community and put more strain on the hospital services as a whole. We therefore include the mean distance to the closest family physician practice, the General Practitioner (GP), for the population who lives in the hospital catchment area.[8]

Robustness check analyses also include controls for the degree of competition in the hospital catchment area, proxied by the number of equivalent public hospitals whose headquarters lie within 30 km of the hospital, for the overall size of the hospital and for hospital staff composition. Hospital size is measured by the total number of beds for general acute care (including overnight and day-only beds) at the trust level. We construct dummies corresponding to seven categories of hospital size: less than 400 beds, 400–549 beds, 550–699 beds, 700–849 beds, 850–999 beds, 1000–1149 beds and over 1150 beds. Data are published quarterly by NHS England and averaged across quarters to obtain hospitals' yearly average number of total beds. Data on hospital staff are reported monthly through the Electronic Staff Records and published quarterly by NHS digital. We construct the proportion of consultants (i.e. senior NHS doctors), the ratio of nursing staff to doctors in full time equivalent (FTE) and the ratio of nursing staff to beds as the yearly mean across quarters.

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:

$$y_{iht} = \alpha + vol'_{ht}\beta_1 + x'_{iht}\beta_2 + x'_{iht}\beta_3 + \delta_t + \varepsilon_{iht}, \tag{1}$$

where y_{iht} is the post-surgical OHS of patient i in hospital h at time of admission t, $\boldsymbol{x_{iht}}$ 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 $\boldsymbol{k_{ht}}$ is a vector of time-varying controls for hospital characteristics (i.e. NHS foundation trust,

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

[5] We extract information on teaching and specialist status from the Estates Returns Information Collection, available at: https://digital.nhs.uk/data-and-information/publications/statistical/estates-returns-information-collection.

[8] We construct these variables based on population statistics from the Office for National Statistics for small homogenous geographic areas (over 32,000 in England) called Lower Super Output Areas (LSOAs). The hospital catchment area comprises all the LSOAs whose centroid falls within 30 km of the hospital's headquarters.

specialist orthopaedic, teaching hospitals and market forces factor in given year t)[6] and control for characteristics of hospitals' catchment area (proportion of population over 65 year-old, mean deprivation and distance to closest GP). $\boldsymbol{\delta_t}$ is a vector of year dummies which account for aggregate change in quality over time. ε_{iht} is a random error term.

Our main interest is in the effect of hospital volume on patients' post-surgery health status. Hospital volume $\boldsymbol{vol_{ht}}$ is entered as a vector of four dummy variables corresponding to volume categories: $vol_{ht}\langle 150,\ 150 \geq vol_{ht}\rangle 200,\ 200 \geq vol_{ht} > 300$ and $vol_{ht} \geq 300$. 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. Using volume quartiles would not permit comparability of the results across specifications given that observed and predicted volumes follow different distributions. We therefore define category thresholds that allow for more weight to be placed on the lower volume categories given expected diseconomies of scale, whilst ensuring that we have enough hospitals in each category and both volume distributions for consistent estimation.[7] We also present results using the log of hospital volume.

We estimate Eq. (1) with pooled OLS. We also estimate a second model where we add hospital fixed effects, denoted with $\boldsymbol{\gamma_h}$, to control for unobserved time-invariant hospital factors. We adjust standard errors for clustering at the hospital level.[8]

3.2. Endogeneity concerns

Regression models, such as that defined in Eq. (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 Eq. (1). To address this, we draw 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 (Gowrisankaran and Town, 2003; Gaynor et al., 2013). 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).

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.[8] Our identification strategy is therefore based on assumptions commonly made in the literature that i) patients' residential choices are not based on the quality of the surgical interventions provided by the surrounding hospitals, and ii) that patients derive higher disutility and costs from travelling further (Kessler and McClellan, 2000; Gaynor et al., 2005; Gutacker et al., 2016; Hentschker and Mennicken, 2018; Moscelli et al., 2018a). Whilst residential sorting is plausible in the context of education, as families may choose to live close to a charter school for instance (Horowitz et al., 2009; Chung, 2015), this is unlikely in the context of acute hospital care in high-income countries. Residential sorting in our case would imply that patients anticipate their future need for a specific healthcare procedure, here orthopaedic surgery, when choosing where to live, and have a good knowledge of this specific aspect of hospital quality. This is even less plausible in the context of a hip replacement, which is a one-off acute treatment, for relatively healthy individuals (as opposed to patients with chronic conditions who may require repeated treatment for the rest of their lives). Further, hospital quality has been shown to be only weakly correlated across conditions (low-risk vs high-risk conditions) and types of care (emergency vs planned care) (Gravelle et al., 2014; Skellern, 2017) and is likely to vary over time. In particular, Gravelle et al. (2014) test for correlation between different measures of quality for a sample of English hospitals and find that hospitals' overall mortality rates are not correlated with any measures of quality related to planned orthopaedic activity (i.e., readmission or revisions after hip or knee replacements). 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 (Geweke et al., 2003; Hentschker and Mennicken, 2018). 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 are limited.

Finally, hospitals may be able to provide higher quality through unobserved determinants of quality that also correlate with volume. By failing to control for these, parameter estimates in Eq. (1) will suffer from omitted variable bias. We address these concerns by running specifications with hospital fixed effects. Results from a hospital fixed effects model estimate the

[6] These characteristics are defined at the trust level. For simplicity, we use the same subscript h for hospitals and trusts. Hospital status may change over time.

[7] The smallest volume category accounts for 10% of the volume distribution and a minimum of 21 hospitals.

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

[12] There is an analogy between our method and previous instrumental variable strategies (Gaynor et al., 2005; Hentschker and Mennicken, 2018) 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.

effect of change in volume within hospitals over time on patient health outcomes. While using hospital fixed effects may also curtail relevant variation in volume, e.g. across hospitals, it allows for a thorough control of potential unobserved time-invariant hospital factors, and thus mitigates further the risk of omitted variable bias. Notice that even after controlling for hospital fixed effects, reverse-causality bias may still remain if hospitals that experience improvements in quality over time attract more patients leading to an increase in volume. We therefore use both observed and predicted volume in the fixed effects specification. In addition, we test the sensitivity of our results to the inclusion of additional time-varying hospital control variables that may be potentially correlated with volume (degree of competition) or endogenous to volume (overall size of the hospital, or hospital staff composition) in the robustness checks reported in Section 4.5.

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 they participated in the PROM survey. The utility of patient i choosing hospital h at time t can be written as:

$$u_{iht} = V_{iht} + \nu_{iht}, \tag{2}$$

where V_{iht} is the utility of patient i derived from observed characteristics of hospital h and ν_{iht} is the unobserved utility.

We specify V_{iht} as:

$$V_{iht} = \gamma_1 d_{iht} + \gamma_2 d_{iht}^2 + \gamma_3 d_{iht}^3 + \gamma_4 close_{iht} + z_{ht}^{'}\gamma_z + \sum_{k=1}^{K} x_{ikt}\left(\gamma_{1k} d_{iht} + \gamma_{2k} d_{iht}^2 + \gamma_{3k} d_{iht}^3\right) \tag{3}$$

where d_{iht} represents the distance between patient i and hospital h at time t, measured as the straight-line distance between hospital's postcode and the centroid of 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, $close_{iht}$, to capture the utility of avoiding any excess travel past the closest hospital. The vector $\boldsymbol{z_{ht}}$ consists of dummy variables for hospital characteristics at time t (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 $\boldsymbol{x} = (x_{ikt},\ k = 1, .., \ K)$, a vector of K patient characteristics (age, sex, socio-economic status, Elixhauser comorbidities, and whether the patient lives in a rural area[9]) as the impact of distance on hospital choice also depends on patients' socioeconomic and clinical factors.[10] 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 ν_{iht} are iid extreme-value (Train, 2009), the probability that patient i chooses hospital h at time t can be estimated by maximum likelihood and is given by:

$$\hat{p}_{iht} = \frac{\exp\left(\hat{V}_{iht}\right)}{\sum_{h' \epsilon M_{it}} \exp\left(\hat{V}_{ih't}\right)} \tag{4}$$

where M_{it} 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 $\hat{p}_{iht}$ across all patients of choosing hospital h:

$$\widehat{vol}_{ht} = \sum_{i=1}^{N} \hat{p}_{iht} = \sum_{i=1}^{N} \frac{\exp\left(\hat{V}_{iht}\right)}{\sum_{h' \epsilon M_{it}} \exp\left(\hat{V}_{ih't}\right)}\ . \tag{5}$$

We estimate Eq. (4) for the whole sample of planned primary hip replacement patients in England for all years between 2011/12 and 2015/16, after exclusion of patients under 50-years old who are atypical medical cases for planned hip replacement and hospitals with less than 20 hip replacement cases per year. In the robustness check presented in Table A1 in Appendix, we also relax these sample restrictions when estimating the choice model. Preferences in hospital choice may vary across years. We therefore estimate the choice model separately for each year before merging all years to obtain the final sample.

Appendix Table A3, Table A4 and Fig. A1 present summary statistics for the choice model sample, consisting of 261,743 patients.[11] Predicted hospital volumes are less dispersed than observed volumes (Fig. A2, 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

[9] The geographical information for lower super output areas (LSOAs) comes from the Office for National Statistics.

[10] In a sensitivity analysis, we included an indicator variable for patients who had a hip replacement surgery in the previous year (slightly under four percent of the sample), to account for the fact they will likely return to the same hospital. Predicted volumes under this alternative specification were highly correlated with our baseline predicted volumes (Pearson correlation coefficient = 0.99).

[11] Estimated coefficients from the choice model for the last year of our sample are available in Table S1 in the supplementary material.

Table 1
Hospital characteristics, for all hospital-years (2011/12–2015/16).

	Mean	SD			Min.	Max.
		total	between	within		
Observed volume	222.54	156.45	147.69	40.31	20.00	1238.00
Foundation Trust	0.57	0.49	0.49	0.07	0.00	1.00
Teaching hospital	0.21	0.41	0.41	0.06	0.00	1.00
Specialist hospital	0.02	0.13	0.12	0.00	0.00	1.00
Market forces factor	1.08	0.07	0.07	0.00	1.00	1.30
Hospital catchment area						
% of pop. over 65 years	16.93	3.10	3.05	0.49	11.23	24.55
Mean deprivation rank	15,717.74	3143.93	3224.20	393.15	11,383.87	24,665.11
Mean distance to GP (km)	1.48	0.54	0.52	0.04	0.78	3.44
Hospital-years	892					

Notes: Volume is the number of annual planned primary hip replacements per hospital. The market forces factor index adjusts hospital resource allocation for unavoidable geographical differences in the costs of labour and capital. A hospital catchment area comprises all the small homogenous geographic areas (Lower Super Output Areas, LSOAs) whose centroid falls within 30 km of the hospital's headquarters. The index of multiple deprivation ranks each LSOA according to their level of deprivation, from 0 (the most deprived) to 32,844 (the least deprived).

patient population.[12] The correlation coefficient between both observed and predicted volumes in our estimation sample is 0.61 ($p < 0.001$).

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 *omit* hospital quality[13] from our model specification, thus creating 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). Our model is therefore an approximation of patients' demand for hospitals, if they were to ignore hospital quality considerations. We re-estimate our model with varying sets of alternatives (comprising the 30 or 10 closest hospitals in patients' choice set). Hospitals' predicted volumes under these alternative specifications are highly correlated (minimum 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 222 hip replacements patients annually, ranging from 20 to 1238 surgeries. Table 1 also reports the total, between and within-hospital standard deviation (SD) for all hospital characteristics. The within-hospital standard deviation is much smaller (40.31), about 27%, than the between-hospital standard variation (147.69) as hospitals are less likely to experience dramatic changes in volumes over time. 57% of hospitals are NHS Foundation Trusts (Table 1). Teaching hospitals and specialist orthopaedic hospitals account respectively for 21% and two percent of hospitals. The average hospital has a market forces factor of 1.08 and a (total) standard deviation of 0.07. On average, 16.93% of the population in the hospital catchment area is over 65 years old. Population in the hospital catchment area has a mean deprivation rank of 15,971 (relative to the catchment area with highest deprivation with a rank of 24,208) and lives on average 1.48 km away from the closest GP practice.

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

Table 2 presents descriptive statistics of patient characteristics. The average pre-surgery OHS is 17.52 points, and patients gain on average 22 points (from 17.52 to 38.69) 6 months after surgery. Patients are on average 70 years old and 40% of our sample are male. On average, patients report slightly over 1 (1.37) Elixhauser comorbidity. The large majority of patients (70%) report having hip-related symptoms for between 1 and 5e years.

Table 3 presents summary statistics of the surgeon characteristics used in Section 4.3. On average, a surgeon treats around 57 planned hip replacement patients per year. 99% of orthopaedic surgeons are male. 63% of the surgeons are trained in the UK, and, on average, had their primary medical qualification 24 years ago. We use surgeons' years since primary medical qualification as a proxy for seniority. We exclude surgeons who report less than 10 cases a year ($N = 1130$ surgeons, more than half of which only treat 1 hip replacement patient per year), and 4 surgeon outliers who report more than 300 annual cases.

[12] The observed volumes with and without these sample restrictions have a correlation coefficient of 0.98.

[13] We also estimate our choice model *with* hospital quality using hospitals' average lagged risk-adjusted Oxford Hip Score gain and standardized overall mortality rates, before removing these effects for the computation of predicted volumes. Predicted volumes under this alternative specification of the choice model are highly correlated with our baseline choice model (Pearson correlation coefficient =0.98) and results (available upon request) are unchanged.

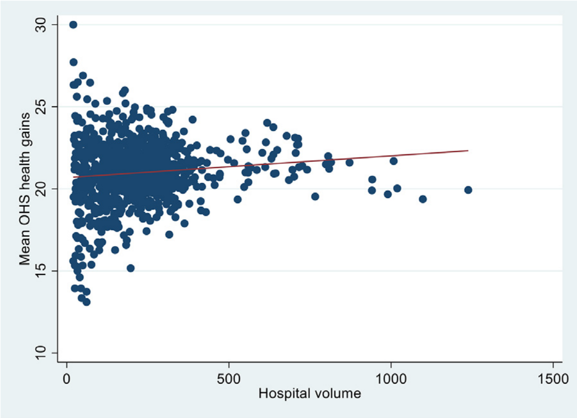

Fig. 1. Association between hospital volume and Oxford Hip Score (OHS) health gain
Notes: Plot of the average OHS health gains (post-surgery minus pre-surgery OHS, unadjusted for other patient characteristics) per hospital against hospital volumes, for all hospital-years between 2011/12 and 2015/16.

Table 2
Patient characteristics, for all sample (2011/12–2015/16).

	Mean	SD	Min.	Max.
Post-surgery OHS	38.69	9.03	0.00	48.00
Pre-surgery OHS	17.52	8.07	0.00	48.00
Age	69.87	8.90	50.00	99.00
Male	0.40	0.49	0.00	1.00
Elix. comorbidity count	0.25	0.73	0.00	8.00
Index of multiple deprivation (IMD):				
1st quintile	0.24	0.43	0.00	1.00
2nd quintile	0.25	0.43	0.00	1.00
3rd quintile	0.22	0.42	0.00	1.00
4th quintile	0.17	0.38	0.00	1.00
5th quintile – most deprived	0.12	0.32	0.00	1.00
Ethnicity: white	0.91	0.28	0.00	1.00
Surgery on the other hip	0.18	0.39	0.00	1.00
Diagnosed with osteoarthritis	0.97	0.18	0.00	1.00
Hybrid prosthetic replacement	0.19	0.39	0.00	1.00
PROMs questions:				
Self-reported disability	0.61	0.49	0.00	1.00
Self-reported depression	0.08	0.27	0.00	1.00
Received assistance in filling questionnaire	0.07	0.26	0.00	1.00
Symptoms duration:				
<1 year	0.13	0.34	0.00	1.00
1–5years	0.70	0.46	0.00	1.00
6–10years	0.11	0.31	0.00	1.00
>10years	0.06	0.23	0.00	1.00
Living arrangements:				
Lives alone	0.28	0.45	0.00	1.00
Lives with family	0.71	0.45	0.00	1.00
Other	0.01	0.08	0.00	1.00
N	105,229			

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 6 months after the operation.

Table 4 presents summary statistics for the additional hospital controls used in robustness checks in Section 4.5. On average, a hospital has slightly under 6 equivalent rival hospitals in its catchment area. Hospital bed categories are relatively evenly distributed. Hospitals with less than 400 beds, the smallest category, stands for 8% of the hospital sample, and hospitals with over 1150 beds, the largest category, representing 24% of the hospital sample. On average, doctors account for 12.68% of hospital staff in full-time equivalent. There are on average 2.28 nursing staff for one doctor, and 1.53 nurses per bed in our sample of hospitals.

4.2. Main results

Table 5 provides the results for the pooled 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

Table 3
Surgeon characteristics, for all surgeon-years (2011/12–2015/16).

	Mean	SD			Min.	Max.
		total	between	within		
Surgeon yearly volume	56.67	40.69	37.31	14.59	10.00	270.00
Male surgeon	0.99	0.11	0.12	0.00	0.00	1.00
Years since qualification	24.73	7.52	7.79	1.29	9.00	45.00
Qualified in the UK	0.63	0.48	0.49	0.00	0.00	1.00
Surgeon-years	4619					

Notes: Surgeon volume comprises all patients treated in all hospitals (if the surgeon holds multiple appointments). We exclude surgeons with less than 10 annual cases or above 300 annual cases who are probable volume outliers. Years since qualification is the time since primary medical qualification, after which surgeons follow additional speciality training.

Table 4
Additional hospital statistics, for all hospitals in pooled sample (2011/12–2015/16).

	Mean	SD			Min.	Max.
		total	between	within		
Number of rivals	5.89	6.79	6.96	0.67	0.00	25.00
Overall number of beds						
<400 beds	0.08	0.27	0.25	0.11	0.00	1.00
400–549 beds	0.13	0.34	0.29	0.17	0.00	1.00
550–699 beds	0.15	0.36	0.30	0.18	0.00	1.00
700–849 beds	0.16	0.37	0.32	0.21	0.00	1.00
850–999 beds	0.15	0.36	0.29	0.21	0.00	1.00
1000–1149 beds	0.09	0.29	0.25	0.17	0.00	1.00
>1150 beds	0.24	0.43	0.41	0.16	0.00	1.00
% of doctors in hospital staff	12.68	2.37	2.24	0.76	5.39	25.87
Nurses-doctors ratio	2.28	0.54	0.47	0.23	0.82	7.45
Nurses-beds ratio	1.53	0.34	0.32	0.16	0.90	3.46
Hospital-years	892					

Notes: The number of equivalent hospital rivals in the hospital's catchment area comprises all public hospitals (trusts) whose headquarters lie within 30 km of the hospital. The number of beds (total of overnight and day only beds, published quarterly by NHS England) and staff data (the proportion of hospital staff who are doctors, the nurses-to-doctors ratio and nurses-to-beds ratio) are yearly average for hospital trusts and are lagged by one year. Data on hospital staff are reported monthly through the Electronic Staff Records and published quarterly by NHS digital.

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 0.72 fewer OHS points. The estimate is statistically significant at the 0.1% level. Patients treated in hospitals performing between 150 and 200 cases are estimated to gain 0.48 fewer OHS points, and the coefficient is statistically significant at the one percent 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 4 points (Varagunam et al., 2015b). Therefore, the estimated association is quantitatively small, as it accounts for slightly under 20% (0.7 points) of a clinically meaningful change.[14]

The coefficients on patient characteristics are all statistically significant, though are not substantial in clinical terms. Patients with higher pre-surgery health (OHS score) also have higher health after surgery. Coefficients on pre-surgery score are close to or above 4 OHS points, suggesting that the difference is clinically important. 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 outcome than less deprived ones. Having one more Elixhauser comorbidity leads to a reduced post-surgery OHS score by about 0.47 points (12% of a minimally clinically important difference). Self-reported depression, disability, or help in filling questionnaires are negatively associated with post-surgery outcomes by two OHS points or above. Only specialist orthopaedic hospitals are associated with slightly better health outcomes, though the difference is not clinically important.

In panel A (top panel) of Table 6, we report the regression results for observed hospital volumes and predicted hospital volumes using pooled OLS. The covariate coefficients are similar for the specifications with observed and predicted volumes. We therefore only present the coefficients for hospital volume, under different functional forms: using volume categories or continuous volume in log form. Both functional forms allow for a nonlinear effect of volume on health outcomes. Table 6, panel A, shows observed hospital volume is associated with higher patient post-surgery OHS scores, irrespectively of the functional form of volume chosen. However, when using the predicted (exogenous) hospital volumes, the volume

[14] Note however that the reasoning holds at the individual level; the total OHS gains at the aggregate level can amount to more substantial health gains.

Table 5
Results from pooled OLS regression with observed hospital volumes.

	Post-surgery OHS	
	Coefficient	SE
Volume *[Ref. >=300]*		
<150 cases	−0.720***	(0.191)
150–200 cases	−0.484**	(0.177)
200–300	−0.242	(0.130)
Pre-surgery OHS [Ref. 0-6 pts]		
6–12 pre-surgery OHS	2.996***	(0.155)
12–18 pre-surgery OHS	5.007***	(0.163)
18–24 pre-surgery OHS	6.314***	(0.168)
24–30 pre-surgery OHS	7.174***	(0.175)
30–36 pre-surgery OHS	7.987***	(0.180)
36–42 pre-surgery OHS	8.730***	(0.226)
42–48 pre-surgery OHS	9.250***	(0.402)
Age [Ref. 50-59 years]		
60–69 years	0.380***	(0.097)
70–79 years	−0.653***	(0.109)
80–89 years	−1.463***	(0.126)
90–105 years	−1.196***	(0.294)
Male patient	0.969***	(0.056)
Ethnic group [Ref. white]		
Other ethnic group	−0.934**	(0.332)
Ethnicity not coded	1.083***	(0.134)
Deprivation Index IMD [Ref. 1st quintile]		
2nd quintile	−0.307***	(0.075)
3rd quintile	−0.579***	(0.093)
4th quintile	−1.499***	(0.095)
5th quintile - Most deprived	−2.732***	(0.127)
Surgery on the other hip	−0.677***	(0.075)
Elix. comorbidity count	−0.470***	(0.055)
Diagnosed with osteoarthritis	1.468***	(0.173)
Hybrid prosthetic replacement	0.209*	(0.084)
Self-reported disability	−2.414***	(0.066)
Self-reported depression	−2.995***	(0.132)
Received assistance in filing questionnaire	−2.749***	(0.137)
Symptoms duration [Ref. <1 year]		
1 to 5 years	−0.767***	(0.082)
6 to 10 years	−1.504***	(0.106)
More than 10 years	−1.815***	(0.154)
Living arrangements [Ref. alone]		
Lives with family	0.393***	(0.059)
Other	−0.835*	(0.374)
Teaching hospital	−0.069	(0.137)
Specialist hospital	0.632*	(0.281)
Foundation Trust	−0.007	(0.126)
Market forces factor	0.271**	(0.101)
% of pop. over 65 years old (catchment area)	0.058	(0.043)
Mean deprivation rank (catchment area)	0.000	(0.000)
Mean distance to GP (catchment area)	0.307	(0.189)
Year dummies [Ref. 2011]		
2012	0.638***	(0.102)
2013	0.613***	(0.119)
2014	0.758***	(0.111)
2015	1.626***	(0.132)
R^2	0.175	
Observations	105,229	

Notes: In parentheses, robust standard errors clustered on hospitals.
* $p < 0.05$,.
** $p < 0.01$,.
*** $p < 0.001$.

coefficients reported in columns 3–4 are no longer statistically significant. This suggests the presence of a spurious positive relation between health outcomes and volumes. After accounting for reverse causality, due to hospitals with higher quality attracting more patients, hospital volumes are no longer associated with improved health outcomes (Table 6, panel A, columns 3–4).[15]

[15] Note that with predicted volumes, bootstrapping standard errors would be the appropriate technique to account for the fact that the predicted volumes are generated in a first stage choice model. Model-based standard errors do not account for sampling variation in the predicted volumes, which may lead

Table 6
Effect of observed and predicted hospital volumes on patient post-surgery OHS.

	Observed hospital volumes		Predicted hospital volumes	
	(1)	(2)	(3)	(4)
Panel A: Pooled OLS				
Volume *[Ref. >=300]*				
<150 cases	−0.720***		0.012	
	(0.191)		(0.250)	
150–200 cases	−0.484**		0.011	
	(0.177)		(0.152)	
200–300	−0.242		−0.199	
	(0.130)		(0.140)	
Log(volume)		0.404***		−0.028
		(0.089)		(0.178)
R^2	0.175	0.175	0.174	0.174
Panel B: Fixed Effects				
Volume *[Ref. >=300]*				
<150 cases	−0.316		−0.106	
	(0.239)		(0.274)	
150–200 cases	−0.092		−0.151	
	(0.174)		(0.204)	
200–300	−0.146		−0.101	
	(0.098)		(0.118)	
Log(volume)		0.202		0.157
		(0.195)		(0.263)
R^2	0.182	0.182	0.182	0.182
Observations	105,229	105,229	105,229	105,229

Notes: The same covariates (patient controls, hospital time-varying controls and year dummies) as in Table 5 are included. In parentheses, robust standard errors clustered on hospitals (columns 1–4). $^{*}p < 0.05$,.
** $p < 0.01$,.
*** $p < 0.001$.

In panel B (bottom panel) of Table 6, we report the same results but estimate a hospital fixed-effects model to account for potential unobserved time-invariant heterogeneity in hospital quality. Our results are essentially similar: predicted hospital volumes are not statistically significant, indicating no causal effect of volume on patient health (panel B, columns 3–4). However, observed hospital volumes are no longer significantly associated with better patient health outcomes in this context (panel B, columns 1–2). This may be because either the fixed effect specification controls for the correlation between (time-invariant) quality and volume across hospitals, and/or because it relies solely on the within-hospital variation in volume over time, which is limited in this context as shown in summary statistics in Table 1.[16] The comparison of the results with observed or predicted volume also suggests that reverse-causality bias seems limited in a fixed effects specification. Results are also robust to the exclusion of (observed) volume outliers (i.e., hospitals treating over 1000 annual hip replacement cases). In summary, the fixed effects results confirm our key finding that volume is not associated with quality.

4.3. Testing for the effect of surgeon volume

Volume-outcome effect may be due to hospital factors and/or personnel effects. Healthcare personnel, and chiefly, surgeons, may experience positive learning effects as their volumes of activity increase. They may gain technical proficiency and become more apt at detecting complications. A higher volume of patients, if it entails a more regular practice, may also ensure that operating skills are maintained over time (Ramanarayanan, 2008; Hockenberry and Helmchen, 2014). To test that hospital volume is not simply a proxy for individual surgeon effects, we run the same models but additionally control for individual surgeons' yearly volume and characteristics, such as their gender, years since graduation as a proxy for seniority and being 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

to downward-biased standard errors (Murphy and Topel, 1985). However, because the procedure is computationally intensive and because larger standard errors would not affect our results as we find a null effect for predicted hospital volumes, we do not bootstrap the standard errors throughout the study. A comparative table for the main results in Table 5 with and without bootstrapping is available upon request.

16 More substantial changes in hospital volumes over time may arise after the closure of nearby hospitals or hospital mergers. Such quasi-exogenous shocks in volume have been used elsewhere in the literature (e.g. Avdic et al. (2019)).

Table 7
Regression results with surgeon volumes, pooled OLS and surgeon fixed effects model.

	Observed hospital volumes		Predicted hospital volumes	
	(1)	(2)	(3)	(4)
Panel A: Pooled OLS				
Hospital volume *[Ref. >=300]*				
<150 cases	−0.412*	−0.534**	0.103	0.065
	(0.172)	(0.174)	(0.226)	(0.237)
150–200 cases	−0.243	−0.347*	0.153	0.080
	(0.165)	(0.168)	(0.129)	(0.142)
200–300	−0.122	−0.194	−0.099	−0.140
	(0.117)	(0.126)	(0.122)	(0.136)
Log(surgeon volume)	0.487***		0.537***	
	(0.071)		(0.071)	
Surgeon volume >= 35		0.570***		0.637***
		(0.108)		(0.108)
Male surgeon	0.608	0.683*	0.592	0.689*
	(0.310)	(0.309)	(0.313)	(0.311)
Qualified in the UK	0.281*	0.359**	0.303*	0.397**
	(0.118)	(0.120)	(0.120)	(0.124)
Years since qualification	−0.030***	−0.027***	−0.030***	−0.027***
	(0.006)	(0.006)	(0.006)	(0.006)
R^2	0.177	0.177	0.177	0.177
Panel B: Surgeon Fixed Effects				
Hospital volume *[Ref. >=300]*				
<150 cases	−0.507**	−0.542**	0.348	0.309
	(0.183)	(0.184)	(0.243)	(0.248)
150–200 cases	−0.142	−0.173	0.111	0.073
	(0.171)	(0.168)	(0.174)	(0.176)
200–300	−0.088	−0.106	0.033	0.005
	(0.098)	(0.097)	(0.108)	(0.110)
Log(surgeon volume)	0.282*		0.324**	
	(0.126)		(0.124)	
Surgeon volume >= 35		0.203		0.216
		(0.155)		(0.153)
R^2	0.200	0.200	0.200	0.200
Observations	101,304	101,304	101,304	101,304

Notes: The same covariates (patient controls, hospital time-varying controls and year dummies) as in Table 5 are included. In parentheses, robust standard errors are clustered on hospitals. In panel B, surgeon characteristics (being male, qualification in the UK or years since qualification) are not included as they would be collinear with the surgeon fixed effects or the year dummies. Sample size is slightly smaller due to missing surgeon characteristics or exclusion of surgeon outliers.
* $p < 0.05$,.
** $p < 0.01$,.
*** $p < 0.001$.

NHS Commissioning Board (2012), and reported that *all* surgeons were in line with expectations[17] (Varagunam et al., 2015a). 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 7 shows the results of a specification where we allow for surgeon effects. As in Table 6, we report results from two specifications: first, we run a pooled OLS model (panel A) and second, we estimate a surgeon fixed effects model to account for unobserved surgeon effects. Hospital fixed effects will be highly collinear with surgeon fixed effects, therefore we only include surgeon fixed effects. We alternatively use continuous surgeon volumes in logs to allow for a non-linear effect of surgeon volumes on health outcomes (columns 1 and 3) and a categorical variable (columns 2 and 4) for surgeon volume above the safety threshold (35 annual cases) identified by Ravi et al. (2014) for surgeons performing total hip arthroplasty in the US.

Results from the pooled OLS model indicate that there is a significant effect of surgeon volume on patient health. The quantitative effect however is small. A 10% increase in surgeon volume is associated with around 0.05 additional OHS points, equivalent to 1.25% of a clinically minimal important difference (four OHS points). The effect is stronger (coefficient around 0.6, i.e. approximately 15% of a minimally important difference) when we compare surgeons who perform less than 35 annual cases with surgeons above that threshold. The coefficient for surgeon volumes is stable across specifications.

Hospitals with small volumes of hip replacement 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 surgeon characteristics. Surgeons

[17] This is also stated on the National Joint Registry (NJR) website to inform patient choice for current data, available at: http://www.njrsurgeonhospitalprofile.org.uk/FAQ#10.

Table 8
Regression results with more inclusive measures of hospital volumes.

	Observed hospital volumes			Predicted hospital volumes		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pooled OLS						
Log(volume)	0.404***			−0.028		
	(0.089)			(0.178)		
Log(volume) – with emergency hip replacements		0.325**			−0.173	
		(0.105)			(0.177)	
Log(volume) – with emergency hip and planned knee replacements			0.295**			−0.205
			(0.106)			(0.188)
R^2	0.175	0.175	0.174	0.174	0.174	0.174
Panel B: Fixed effects						
Log(volume)	0.202			0.157		
	(0.195)			(0.263)		
Log(volume) – with emergency hip replacements		0.070			−0.225	
		(0.236)			(0.266)	
Log(volume) – with emergency hip and planned knee replacements			0.061			−0.151
			(0.234)			(0.296)
R^2	0.183	0.183	0.183	0.183	0.183	0.183
Observations	105,229	105,229	105,229	105,229	105,229	105,229

Notes: The same covariates (patient controls, hospital time-varying controls and year dummies) as in Table 5 are included. In parentheses, robust standard errors are clustered on hospitals.
* $p < 0.05$,.
** $p < 0.01$,.
*** $p < 0.001$.

who qualified in the UK are associated with slightly better health outcomes, while 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.

Results from the specification with surgeon fixed effects (panel B) however show that surgeon volume is associated with differences in patient health but the magnitude of the association is smaller once we control for time-invariant surgeon effects with surgeon fixed effects.

Overall, the results in Table 7 confirm that observed hospital volume is significantly associated with patient health outcomes, even after controlling for surgeon effects. Predicted hospital volume has no effect on health gains even after controlling for surgeon effects (Table 7, columns 3 and 4), even when we include surgeon fixed effects (Table 7, panel B).

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 whom we observe to work only in NHS hospitals (around 55%, $N = 689$).[18] Our results (Table A5 in Appendix) are essentially unchanged.

The results show that observed hospital volume is associated with health outcomes, even after controlling for individual surgeon volume and surgeon fixed effects. In a robustness check in Section 4.5, we control for other measures of personnel effects, such as the proportion of staff who are doctors, the nurses-to-doctors and the nurses-to-beds ratio measured at the hospital level.

4.4. More inclusive measures of volume

Scale economies may also arise from performing treatments that are different but related to planned hip replacements, and therefore contribute to improvement in health outcomes for these patients through ameliorated processes or learning-by-doing effects (Schilling et al., 2003). To address this, in this section, we employ more comprehensive measures of orthopaedic volumes. We first include emergency hip replacements in our measure of hospital and surgeon volumes, as this relates to the same procedure but in an emergency rather than an elective setting. Second, we further add knee replacements, which are also performed in an elective setting and involve similar surgeon skills. Table A6 in Appendix presents summary statistics for these different definitions of volumes.

Table 8 compares the results when using different measures of hospital volumes. Given that different definitions of volume have a different support and distribution, we use the log of volume, rather than the previously defined volume categories, to compare the results. Columns (1) and (4) correspond to our baseline measure of volume comprising all planned hip replacements. The coefficient on observed volume is significant in most specifications in columns 1–3 but diminishes

[18] While private hospitals represent an important share of all 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.

Table 9
Regression results with broader measures of surgeon volumes (pooled OLS).

	Observed hospital volumes			Predicted hospital volumes		
	(1)	(2)	(3)	(4)	(5)	(6)
Hospital volume *[Ref. >=300]*						
<150 cases	−0.412*	−0.439*	−0.538**	0.103	0.103	0.044
	(0.172)	(0.172)	(0.179)	(0.226)	(0.228)	(0.235)
150–200 cases	−0.243	−0.271	−0.346*	0.153	0.152	0.092
	(0.165)	(0.164)	(0.170)	(0.129)	(0.129)	(0.142)
200–300	−0.122	−0.142	−0.186	−0.099	−0.107	−0.137
	(0.117)	(0.117)	(0.128)	(0.122)	(0.123)	(0.136)
Log(surgeon volume)	0.487***			0.537***		
	(0.071)			(0.071)		
Log(surgeon volume) – with emergency hip replacements		0.510***			0.566***	
		(0.078)			(0.078)	
Log(surgeon volume) – with emergency hip and planned knee replacements			0.295***			0.358***
			(0.079)			(0.076)
Male surgeon	0.608	0.626*	0.682*	0.592	0.613*	0.675*
	(0.310)	(0.305)	(0.309)	(0.313)	(0.309)	(0.312)
Qualified in the UK	0.281*	0.281*	0.360**	0.303*	0.304*	0.393**
	(0.118)	(0.118)	(0.123)	(0.120)	(0.121)	(0.127)
Years since qualification	−0.030***	−0.027***	−0.027***	−0.030***	−0.027***	−0.027***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
R^2	0.177	0.177	0.177	0.177	0.177	0.176
Observations	101,304	101,304	101,304	101,304	101,304	101,304

Notes: The same covariates (patient controls, hospital time-varying controls and year dummies) as in Table 5 are included. In parentheses, robust standard errors are clustered on hospitals. Sample size is slightly smaller due to missing surgeon characteristics or exclusion of surgeon outliers.
* $p < 0.05$,.
** $p < 0.01$,.
*** $p < 0.001$.

in size as we include additional activity, first by adding emergency hip replacements (column 2), and then, planned knee replacements (column 3). These results suggest that these additional surgeries are less relevant to returns of scale and including them potentially introduces some measurement error. Coefficients for predicted hospital volumes are not statistically significant for either definitions of volumes (columns 4–6). When we introduce hospital fixed effects (panel B), as with our baseline results, no measure of volume is statistically significant.

Similarly, Table 9 shows the results when adding surgeon volume using these different definitions. We keep our baseline measure of hospital volume (planned hip replacements) as results from Table 8 show these seem the most relevant to measure returns to scale. Surgeon volume is associated with better health post-surgery OHS, though the magnitude of the association is similar or lower when we include related orthopaedic activity, as with hospital volumes in Table 8. Measuring the effect of all planned activity (hip and knee replacements) first gives the same results, both for hospital and surgeon volume (see results in Tables A7 and A8 in Appendix). The overall results are unchanged: observed hospital volumes are associated with patient post-surgery health, even after controlling separately for individual surgeon volume and characteristics.

4.5. Robustness checks

Our estimation of predicted volumes based on the patient choice model relies on patients' distance to the hospital being exogenous to hospital quality, conditional on our set of controls. This assumes that there are no unobserved patient confounders that relate both to patient distance to the hospital and to health outcomes. Though untestable, this is plausible considering our large set of patient controls. Unlike for certain acute conditions that require immediate treatment, e.g., heart attacks or strokes, where delays in access to care may impact health outcomes, timely access to care does not affect health outcomes in the case of planned hip surgery (Tuominen et al., 2010; Brealey et al., 2012). Further, we show in Appendix, Table A9, that patient distance to the hospital is not correlated with patient pre-surgery OHS or with the number of comorbidities after controlling for patients' main socio-economic characteristics,[19] suggesting that remaining unobserved confounders are unlikely.

Our dependant variable, post-surgery Oxford Hip Score, is a subjective measure of health. Patients who care about hospital quality, and therefore select actively into better quality hospitals, may also answer the outcome questionnaire differently.

[19] Results from Table A9 show that patient distance to the closest hospital is weakly correlated with the number of comorbidities (i.e. increasing patient distance to the closest hospital by 100 km is associated with around 0.6 fewer comorbidities) but the effect is driven by patients with the highest distance to the closest hospital. The correlation disappears when we remove the top percentile of patients with the largest distance to their closest hospital (Table A9, column 5).

Table 10
Results with additional hospital control variables (pooled OLS).

	Observed volumes	Predicted volumes
Hip replacement volume *[Ref. >=300]*		
<150 cases	−0.690***	−0.023
	(0.188)	(0.246)
150–200 cases	−0.432*	0.053
	(0.178)	(0.180)
200–300	−0.276*	−0.200
	(0.140)	(0.154)
Measure of hospital competition		
Number of rivals	−0.012	−0.016
	(0.030)	(0.031)
Total number of beds (Ref. >1500 beds)		
<400 beds	−0.267	−0.275
	(0.336)	(0.338)
400–549 beds	−0.368	−0.452
	(0.303)	(0.308)
550–699 beds	−0.304	−0.339
	(0.286)	(0.287)
700–849 beds	−0.071	−0.061
	(0.281)	(0.276)
850–999 beds	0.043	0.042
	(0.278)	(0.275)
1000–1149 beds	−0.126	−0.078
	(0.253)	(0.249)
Staff composition		
% of doctors in hospital staff	−0.004	0.026
	(0.071)	(0.074)
Nurses-doctors ratio	0.178	0.196
	(0.139)	(0.149)
Nurses-beds ratio	−0.131	−0.140
	(0.089)	(0.093)
R^2	0.175	0.175
Observations	85,918	85,918

Notes: In parentheses, robust standard errors are clustered on hospitals. A hospital catchment area comprises all the small homogenous geographic areas (Lower Super Output Areas, LSOAs) whose centroids fall within 30 km of the hospital's headquarters. The 2015 index of multiple deprivation ranks each LSOA according to their level of deprivation, from 0 (the most deprived) to 32,844 (the least deprived). The number of equivalent hospital rivals in the hospital's catchment area comprises all public hospitals (trusts) whose headquarters lie within 30 km of the hospital. The number of beds and staff data are lagged yearly average statistics at the trust level.

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

As a robustness check, we use the probability of having a revision surgery within three years of the index surgery as an alternative and objective measure of patient outcomes.[20] Results, in Appendix Table A10, confirm our baseline results with PROMs: there is no statistically significant effect of predicted hospital volume on mortality rates. Like with PROMs, patients treated in hospitals with smaller observed volumes are associated with a higher risk of having a revision surgery, though the effect is only statistically significant at the 10% level.

Furthermore, patient participation in the PROM survey is voluntary and attrition may happen for different reasons. If attrition is systematically correlated with health outcomes and hospital volumes, our estimates will be biased. If non-response to the PROMs questionnaires is driven by poorer underlying patient health, our rich set of risk-adjustment variables (including the pre-surgery OHS) ensures that differences in hospitals' case mix are accounted for. In addition, we regress the rate of PROMs participation per hospital, corresponding to the number of patients who answered the questionnaires out of the total number of eligible patients, on hospital volumes, patient case-mix and hospital status. Results in Table A11 in Appendix indicate no systematic correlation between hospitals' rate of participation to PROMs and hospital volumes. Overall, this suggests that bias linked to attrition is unlikely.

Hospitals may provide better quality through unobserved determinants that correlate with hospital volume. Our large 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 the following additional robustness checks. First, previous studies

[20] Post-surgical mortality is another objective measure of health outcomes but it is very low for planned hip replacement surgery (around 0.06% in our initial sample), as opposed to around 1.3% for three-year hip revision surgery.

have found that hospitals in more competitive areas respond to the competition by increasing quality (Cooper et al., 2012; Gaynor et al., 2013; Bloom et al., 2015), though the effect varies across countries and procedures.[21] We therefore control for the degree of competition in the hospital market, proxied by the number of equivalent rival hospitals in the hospital catchment area (Bloom et al., 2015; Moscelli et al., 2018b). This is because competition could be correlated with hospital volume, if higher aggregate supply causes lower volume for each provider, for a given demand, or if more competitive areas face proportionally larger demand than less competitive areas even accounting for higher supply.

Second, we check that our results are due to the effect of the volume of hip replacement patients on health outcomes, rather than the overall size of the hospital. For instance, larger hospitals may benefit from economies of scope across clinical departments, by pooling resources or skills, which may benefit the quality of care. We measure hospital size by the total number of hospital beds for general acute care and include controls for seven categories: less than 400 beds, 400–549 beds, 550–699 beds, 700–849 beds, 850–999 beds, 1000–1149 beds and over 1150 beds. We do not control for hospitals' size in our baseline model because, similarly to hip replacement volume, it is prone to a reverse causality effect, whereby hospitals with high-quality reputation will attract more patients, thus driving hospitals' bed capacity upward.

Third, hospital staff composition may impact the quality of care, independently of volume. The presence of experienced colleagues may have a positive effect on the team (Ayoubi et al., 2017). We additionally control for the proportion of hospital staff who are doctors, the nurses-to-doctors ratio and the nurses-to-beds ratio across the hospital.[22] Again, we do not include this variable in our baseline regressions because staff composition may also reflect hospital quality, insofar as high-quality hospitals will be more successful in recruiting more and better qualified personnel.

The regression results for the different specifications, shown in Table 10, do not differ from our main results. The positive relationship between the observed hospital volumes (lowest volume category) and the patient outcomes is unchanged and statistically significant. The predicted volumes show no causal effect on patient health gains in any of the specifications.

The coefficients on the pre-surgery Oxford Hip Scores are stable across all specifications. None of the additional hospital characteristics shows a statistically significant relationship with patient outcomes. The sample size in Table 10 is smaller because of some missing staff characteristics for certain years for certain hospitals.

Other unobserved mechanisms might be at work within hospitals. For instance, the presence of physical therapists, whom we do not have data on, could improve patient rehabilitation. Overall, however, our set of additional controls suggest that our results successfully isolate the potential for economies of scale in hip replacement from a range of potential confounders.

5. Conclusions

This study investigates the effect of hospital volume 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). Our results differ from the study by Hentschker and Mennicken (2018) who find a positive causal effect of hospital volume on patient outcomes after emergency hip fracture in Germany. Results may differ because they investigate a more complex procedure typically involving frail patients, with a high post-surgery mortality rate at around six percent.

Our pooled OLS results overall suggest that the hospital-level association can be driven by 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. In the absence of a causal effect of volume on patient outcomes, increasing the provision of planned hip replacements at any hospital would not result in improvements in health outcomes in that hospital. We conclude that we do not find evidence that economies of scale affect quality to support the argument for concentrating the provision in this setting.

Nevertheless, we find a small positive correlation between volume and outcomes. Transferring patients from low-volume hospitals with lower-performance to better quality hospitals could result in health improvements for these patients. However, the potential gains are unlikely to offset the potential adverse effects of concentrating hip replacement activity. The hospital market in England is already concentrated, and further concentration may have adverse effects on patient access to care, as proximity remains a key determinant of patient choice to the provider, especially for relatively older patients in need of a hip replacement (Gutacker et al., 2016). Policies that concentrate provision of health care may also risk shifting some of the NHS cost onto patients and carers by increasing travel times or transportation costs, which may be particularly problematic for patients from disadvantaged socio-economic backgrounds (Ferguson et al., 1997). Further concentration of care may also affect hospital competition, which could have knock-on effects on quality efforts (Cooper et al., 2012; Gaynor et al., 2013; Moscelli et al., 2018b).

The study has some limitations. First, 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

[21] Feng et al. (2015) find no association between market competition of hospitals and patient-reported health outcomes for planned hip replacements in England, while Skellern (2017) finds that a pro-competition reform in the English NHS had a negative effect on PROMs for hip and knee surgeries.

[22] Variables that are directly under hospital control, such as staff composition and the hospital's total number of beds, are lagged by one year, ensuring that they were measured *before* our dependent variable and thus could not have been affected by the contemporaneous level of quality.

volumes suggests that there remains room for improvement. Second, we cannot perfectly disentangle the various channels behind higher hospital volume. We have focused on a procedure which is likely to benefit most from economies of scale and for which concentration of care would be a possible policy option given the planned nature of the surgery. We use hospital volumes to measure hospital economies of scale (e.g. improvement in processes of care), but hospital volume may also be a proxy for increased team experience. However, we include analyses where we control for a range of personnel factors (such as individual surgeon characteristics and volume, and hospital staff composition). The results are robust to the inclusion of these personnel effects, which indicates that attenuation bias linked to measurement error should not be a concern in this case. Future research could explore empirical strategies that disentangle the causal contributions of personnel and hospital factors. Volume effects that are driven primarily by individual surgical learning-by-doing would have different policy implications, and may be best addressed during surgeons' medical residency, or via regular trainings or work schedules that allow for regular practice (Hockenberry and Helmchen, 2014). Provided that data become available, future research could also investigate the effect of volume on other dimensions of quality, such as care responsiveness or facility features .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

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. The views expressed here are those of the authors and not necessarily those of the funder. Hospital Episode Statistics are Copyright 2011–2015, re-used with the permission of NHS Digital (Data Sharing Agreement DARS-NIC-84254-J2G1Q). 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. We also thank the editor and two anonymous referees for numerous suggestions and improvements to the paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.08.014.

Appendix

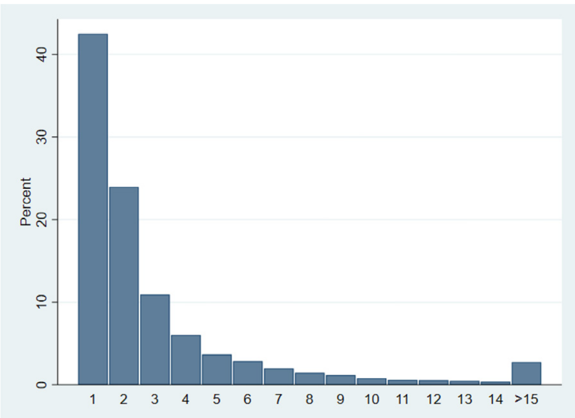

Fig. A1. Percentage of planned hip patients who went to their Nth nearest provider.
Notes: This corresponds to the total sample of 261,743 planned primary hip replacements in England between 2011/12 and 2015/16 over which the multinomial logit model of hospital choice is run.

Table A1
Results after relaxing restriction to hospitals above 20 annual cases (1) or to patients above 50 years old (2), for observed or predicted hospital volumes.

	(1) Observed	Predicted	(2) Observed	Predicted
Volume *[Ref. >=300]*				
<150 cases	−0.743***	−0.042	−0.744***	−0.175
	(0.203)	(0.271)	(0.192)	(0.219)
150–200 cases	−0.449**	−0.018	−0.440*	−0.008
	(0.170)	(0.151)	(0.169)	(0.143)
200–300	−0.231	−0.135	−0.273	−0.193
	(0.128)	(0.132)	(0.139)	(0.143)
R^2	0.178	0.177	0.178	0.177
Observations	110,559	110,559	110,669	110,669

Notes: Results with different sample restrictions. In parentheses, robust standard errors are clustered on hospitals. Patient controls include, besides the pre-surgery Oxford Hip Scores (OHS), the patient age, sex, ethnicity, index of multiple deprivation, the number of Elixhauser comorbidities, diagnosis with osteoarthritis, control for the type of surgery carried out (hybrid prosthetic versus total hip replacement) and for having had a surgery on the other hip, self-reported disability, depression, assistance in filling questionnaires, the symptoms duration and the patient' living arrangements. Hospital controls include hospital status (teaching, specialist orthopaedic and Foundation Trust hospitals), for the market forces factor index, and for socio-demographic characteristics of population in hospitals' catchment area (proportion of over 65-year-old, mean deprivation rank and mean distance to the closest GP).

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

Table A2
Patient characteristics for sample with private providers (1) and final sample (2).

	(1) Mean	SD	(2) Mean	SD
Post-surgery OHS	39.44	8.69	38.69	9.03
Pre-surgery OHS	18.11	8.11	17.52	8.07
Age	69.68	8.73	69.87	8.90
Male patient	0.40	0.49	0.40	0.49
Elix. comorbidity count	0.23	0.68	0.23	0.71
Index of multiple deprivation (IMD):				
1st quintile	0.26	0.44	0.24	0.43
2nd quintile	0.26	0.44	0.25	0.43
3rd quintile	0.22	0.42	0.22	0.42
4th quintile	0.16	0.37	0.17	0.38
5th quintile	0.10	0.30	0.12	0.32
Ethnicity: white	0.88	0.32	0.91	0.28
Surgery on the other hip	0.19	0.40	0.18	0.39
Diagnosed with osteoarthritis	0.96	0.19	0.97	0.18
Hybrid prosthetic replacement	0.18	0.39	0.19	0.39
Self-reported disability	0.57	0.49	0.61	0.49
Self-reported depression	0.08	0.26	0.08	0.27
Received assistance in filling questionnaire	0.06	0.24	0.07	0.26
Symptoms duration:				
<1year	0.14	0.35	0.13	0.34
1–5years	0.69	0.46	0.70	0.46
6–10years	0.11	0.31	0.11	0.31
>10years	0.06	0.23	0.06	0.23
Living arrangements:				
Lives alone	0.27	0.44	0.28	0.45
Lives with family	0.72	0.45	0.71	0.45
Other	0.01	0.07	0.01	0.08
Observations	148,617		105,229	

Notes: The index of multiple deprivation (IMD) measures deprivation across seven domains, including income, employment and education, and 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 surgery.

Table A3
Summary statistics of sample used for hospital patient choice model (patients).

	Mean	SD	Min.	Max.
Patient age	70.01	9.14	50.00	102.00
Male patient	0.39	0.49	0.00	1.00
# Elixhauser conditions	0.27	0.76	0.00	8.00
Index of multiple deprivation (IMD):				
1st quintile	0.24	0.43	0.00	1.00
2nd quintile	0.25	0.43	0.00	1.00
3rd quintile	0.22	0.42	0.00	1.00
4th quintile	0.17	0.38	0.00	1.00
5th quintile	0.12	0.33	0.00	1.00
Patient living in rural area	0.27	0.45	0.00	1.00
Distance to chosen hospital (km)	12.94	12.97	0.00	326.35
Patient choosing closest hospital	0.42	0.49	0.00	1.00
N	261,743			

Table A4
Summary statistics of sample used for hospital patient choice model (hospitals).

	Mean	SD	Min.	Max.
NHS Treatment Centre (TC) site	0.03	0.17	0.00	1.00
Teaching trust	0.16	0.37	0.00	1.00
Specialist trust	0.01	0.09	0.00	1.00
# hospital sites within trust	1.89	1.21	1.00	6.00
Provider type:	ref.			
Independent Sector TC	0.07	0.26	0.00	1.00
Independent Sector non-TC	0.23	0.42	0.00	1.00
NHS Foundation Trust (FT)	0.38	0.49	0.00	1.00
NHS non-FT	0.31	0.46	0.00	1.00
N	465			

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.

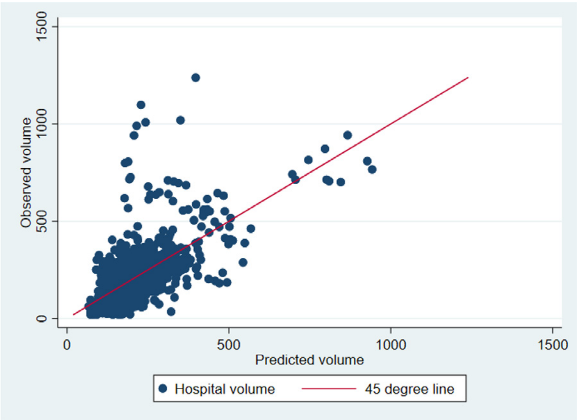

Fig. A2. Scatterplot of the observed against the predicted hospital volumes
Notes: Hospital volumes are the number of planned primary (non-revision) patients treated by a hospital in between 2011/12 and 2015/16 in the English NHS. The predicted volumes are constructed using a conditional logit model of hospital choice. The data points are for each hospital-year in the final sample.

Table A5
Results on subsample of surgeons who work for public hospitals only.

	Observed hospital volumes		Predicted hospital volumes	
	(1)	(2)	(3)	(4)
Panel A: Pooled OLS				
Hospital volume *[Ref. >=300]*				
<150 cases	−0.441*	−0.650**	0.247	0.129
	(0.201)	(0.202)	(0.334)	(0.355)
150–200 cases	−0.156	−0.310	0.079	−0.054
	(0.206)	(0.209)	(0.148)	(0.169)
200–300	−0.047	−0.162	−0.177	−0.254
	(0.147)	(0.163)	(0.141)	(0.163)
Log(surgeon volume)	0.568***		0.624***	
	(0.091)		(0.087)	
Surgeon volume >= 35		0.544***		0.626***
		(0.130)		(0.131)
Male surgeon	0.645	0.711*	0.625	0.723*
	(0.342)	(0.341)	(0.347)	(0.343)
Qualified in the UK	0.256	0.352*	0.281	0.401**
	(0.148)	(0.154)	(0.147)	(0.154)
Years since qualification	−0.030***	−0.025**	−0.031***	−0.026**
	(0.008)	(0.008)	(0.008)	(0.008)
R^2	0.179	0.178	0.179	0.178
Panel B: Surgeon Fixed Effects				
Hospital volume *[Ref. >=300]*				
<150 cases	−0.799**	−0.854**	0.585	0.530
	(0.266)	(0.262)	(0.357)	(0.365)
150–200 cases	−0.163	−0.205	−0.022	−0.071
	(0.253)	(0.248)	(0.243)	(0.245)
200–300	−0.140	−0.166	−0.102	−0.136
	(0.171)	(0.169)	(0.185)	(0.188)
Log(surgeon volume)	0.288		0.362*	
	(0.161)		(0.152)	
Surgeon volume >= 35		0.193		0.220
		(0.187)		(0.184)
R^2	0.205	0.205	0.205	0.205
Observations	61,668	61,668	61,668	61,668

Notes: Results for subsample of surgeons working for NHS hospitals only. In parentheses, robust standard errors are clustered on hospitals. Patient controls include, besides the pre-surgery Oxford Hip Scores (OHS), the patient age, sex, ethnicity, index of multiple deprivation, the number of Elixhauser comorbidities, diagnosis with osteoarthritis, control for the type of surgery carried out (hybrid prosthetic versus total hip replacement) and for having had a surgery on the other hip, self-reported disability, depression, assistance in filling questionnaires, the symptoms duration and the patient' living arrangements. Hospital controls include hospital status (teaching, specialist orthopaedic and Foundation Trust hospitals), for the market forces factor index, and for socio-demographic characteristics of population in hospitals' catchment area (proportion of over 65-year-old, mean deprivation rank and mean distance to the closest GP).

* $p < 0.05$,.

** $p < 0.01$,.

*** $p < 0.001$.

Table A6
Summary statistics for broader definitions of hospital and surgeon volumes.

	Mean	SD	Min.	Max.
Hospital volume:				
Planned hip replacements (baseline)	222.54	156.45	20.00	1238.00
With emergency hip replacements	307.32	176.46	20.00	1238.00
With emergency hip + planned knee replacements	553.61	317.57	24.00	2422.00
Hospital-years	892			
Surgeon volume:				
Planned hip replacements (baseline)	56.67	40.69	10.00	270.00
With emergency hip replacements	66.63	44.24	10.00	589.00
With emergency hip + planned knee replacements	121.83	68.22	10.00	589.00
Surgeon-years	4619			

Table A7
Results with more inclusive measures of volume, all planned activity first.

	Observed hospital volumes			Predicted hospital volumes		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pooled OLS						
Log(volume)	0.404***			−0.028		
	(0.089)			(0.178)		
Log(volume) – with planned knee replacements		0.313**			−0.173	
		(0.095)			(0.177)	
Log(volume) – with planned knee and emergency hip replacements			0.295**			−0.205
			(0.106)			(0.188)
R^2	0.175	0.175	0.174	0.174	0.174	0.174
Panel B: Fixed effects						
Log(volume)	0.202			0.157		
	(0.195)			(0.263)		
Log(volume) – with emergency hip replacements		0.101			0.033	
		(0.199)			(0.273)	
Log(volume) – with emergency hip and planned knee replacements			0.061			−0.151
			(0.234)			(0.296)
R^2	0.183	0.183	0.183	0.183	0.183	0.183
Observations	105,229	105,229	105,229	105,229	105,229	105,229

Notes: In parentheses, robust standard errors are clustered on hospitals.
* $p < 0.05$,.
** $p < 0.01$,.
*** $p < 0.001$.

Table A8
Results with broader measures of surgeon volumes, all planned activity first (pooled OLS).

	Observed hospital volumes			Predicted hospital volumes		
	(1)	(2)	(3)	(4)	(5)	(6)
Hospital volume *[Ref. >=300]*						
<150 cases	−0.412*	−0.530**	−0.538**	0.103	0.043	0.044
	(0.172)	(0.178)	(0.179)	(0.226)	(0.234)	(0.235)
150–200 cases	−0.243	−0.338*	−0.346*	0.153	0.092	0.092
	(0.165)	(0.171)	(0.170)	(0.129)	(0.142)	(0.142)
200–300	−0.122	−0.180	−0.186	−0.099	−0.135	−0.137
	(0.117)	(0.128)	(0.128)	(0.122)	(0.136)	(0.136)
Log(surgeon volume)	0.487***			0.537***		
	(0.071)			(0.071)		
Log(surgeon volume) – with emergency hip replacements		0.288***			0.348***	
		(0.074)			(0.073)	
Log(surgeon volume) – with emergency hip and planned knee replacements			0.295***			0.358***
			(0.079)			(0.076)
Male surgeon	0.608	0.676*	0.682*	0.592	0.667*	0.675*
	(0.310)	(0.312)	(0.309)	(0.313)	(0.313)	(0.312)
Qualified in the UK	0.281*	0.361**	0.360**	0.303*	0.394**	0.393**
	(0.118)	(0.123)	(0.123)	(0.120)	(0.127)	(0.127)
Years since qualification	−0.030***	−0.028***	−0.027***	−0.030***	−0.028***	−0.027***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
R^2	0.177	0.177	0.177	0.177	0.177	0.176
Observations	101,304	101,304	101,304	101,304	101,304	101,304

Notes: In parentheses, robust standard errors are clustered on hospitals. Sample size is slightly smaller due to missing surgeon characteristics or exclusion of surgeon outliers.
* $p < 0.05$,.
** $p < 0.01$,.
*** $p < 0.001$.

Table A9
Correlation between patient distance to closest hospital and pre-surgery patient severity.

	Pre-surgery OHS		Elixhauser comorbidities		
	(1)	(2)	(3)	(4)	(5)
Distance to hospital	0.011	−0.007	−0.002	−0.007**	−0.003
	(0.010)	(0.018)	(0.002)	(0.002)	(0.003)
Distance to hospital^2		0.001		0.000**	0.000
		(0.000)		(0.000)	(0.000)
Patient age	−0.042***	−0.042***	0.028***	0.028***	0.028***
	(0.004)	(0.003)	(0.001)	(0.001)	(0.001)
Male patient	2.333***	2.333***	0.015*	0.015*	0.015*
	(0.057)	(0.057)	(0.007)	(0.007)	(0.007)
Deprivation index (rank)	0.000***	0.000***	−0.000***	−0.000***	−0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ethnicity: white	−0.595**	−0.595**	0.147***	0.147***	0.148***
	(0.186)	(0.186)	(0.020)	(0.020)	(0.020)
Constant	17.623***	17.684***	−0.742***	−0.728***	−0.738***
	(0.354)	(0.367)	(0.043)	(0.044)	(0.045)
R^2	0.041	0.041	0.060	0.060	0.060
Observations	100,930	100,930	100,930	100,930	100,347

Notes: Distance to the hospital is patient's distance to their closest hospital. Controls also include year dummies. In column (5), the top percentile of distances in our sample were dropped. The deprivation index ranks all small geographical areas in England (LSOAs) and goes from 1 (the most deprived) to 32,844 (the least deprived). In parentheses, robust standard errors clustered on hospitals.

* $p < 0.05$,.

** $p < 0.01$,.

*** $p < 0.001$.

Table A10
Results of the effect of hospital volume on 3-year revision rates.

	Pooled OLS		Fixed Effects	
	(1) Observed	(2) Predicted	(3) Observed	(4) Predicted
Volume [Ref. >300]				
<150 cases	0.241+	0.116	0.572+	−0.160
	(0.143)	(0.154)	(0.332)	(0.318)
150–200 cases	−0.043	0.090	0.242	−0.000
	(0.159)	(0.145)	(0.258)	(0.232)
200–300	0.001	0.003	0.240	0.039
	(0.111)	(0.116)	(0.170)	(0.167)
R^2	0.003	0.003	0.006	0.006
Observations	105,229	105,229	105,229	105,229

Notes: In parentheses, robust standard errors are clustered on hospitals.

\+ $p < 0.10$,.

* $p < 0.05$,.

** $p < 0.01$,.

*** $p < 0.001$.

Table A11
Determinants of patient participation to PROM questionnaires for Oxford Hip Score (hospital-level data).

	PROMs participation rate		
	(1)	(2)	(3)
Hospital volume	0.017*	0.013	0.012
	(0.007)	(0.007)	(0.008)
Average patient age		−0.899**	−1.054*
		(0.341)	(0.458)
% of male patients		0.151	−0.009
		(0.166)	(0.175)
% of patients of white ethnicity		0.099	−0.066
		(0.098)	(0.114)
Average deprivation rank		0.000	0.000
		(0.000)	(0.000)

(*continued on next page*)

Table A11 (*continued*)

	PROMs participation rate		
	(1)	(2)	(3)
% of patients with surgery on other hip		0.406**	0.343**
		(0.131)	(0.116)
Average Elixhauser comorbidity count		1.074	1.232
		(2.620)	(2.625)
% of patients with osteoarthritis		0.367***	0.268**
		(0.083)	(0.093)
% of patients with hybrid prosthetic surgery		0.044	0.100
		(0.055)	(0.052)
% of patients who died after surgery			−0.487
			(2.556)
Teaching hospital			−3.534
			(2.347)
Specialist hospital			−4.114
			(6.443)
Foundation trust			4.006*
			(1.980)
Market forces factor			−2.978
			(1.834)
% of pop. over 65 years old (catchment area)			−1.717**
			(0.609)
Mean deprivation rank (catchment area)			−0.000
			(0.000)
Mean distance to GP (catchment area)			7.633**
			(2.555)
R^2	0.080	0.150	0.197
Observations	892	892	892

Notes: Results from a Linear Probability Model with 0–100 response variable. The rate of participation to PROMs questionnaire is the number of patients who answered the PROMs questionnaire (one or both pre- and post-surgery questionnaires were answered) out of the total number of eligible patients per hospital, expressed as a percentage. Patient-level variables are aggregated at the hospital level and expressed as a percentage, so that a one-unit increase corresponds to one percentage point increase. The market forces factor is standardized by its sample standard deviation (0.07). The index of multiple deprivation ranks all small geographical areas in England (LSOAs) and goes from 1 (the most deprived) to 32,844 (the least deprived). A hospital catchment area comprises all the small homogenous geographic areas (Lower Super Output Areas, LSOAs) whose centroids fall within 30 km of the hospital's headquarters. In parentheses, robust standard errors clustered on hospitals.

* $p < 0.05$,.

** $p < 0.01$,.

*** $p < 0.001$.

References

Argote, L., Epple, D., 1990. Learning curves in manufacturing. Science 247 (4945), 920–924.

Avdic, D., Lundborg, P., Vikström, J., 2019. Estimating returns to hospital volume: evidence from advanced cancer surgery. J. Health Econ. 63, 81–99.

Ayoubi, C., Pezzoni, M., Visentin, F., 2017. At the origins of learning: absorbing knowledge flows from within the team. J. Econ. Behav. Organ. 134, 374–387.

Bartel, A.P., Beaulieu, N.D., Phibbs, C.S., Stone, P.W., 2014. Human capital and productivity in a team environment: evidence from the healthcare sector. American Econ. J. 6 (2), 231–259.

Bauer, H., Honselmann, K., 2017. Minimum volume standards in surgery - are we there yet? Visceral Med. 33 (2), 106–116.

Benkard, C.L., 2000. Learning and forgetting: the dynamics of aircraft production. Am. Econ. Rev. 90 (4), 1034–1054.

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

Blanco, B., Kothari, A., Blackwell, R., Brownlee, S., Yau, R., Attisha, J., Ezure, Y., Pappas, S., Kuo, P., Abood, G., 2017. Take the volume pledge" may result in disparity in access to care. Surgery 161 (3), 837–845.

Bloom, N., Propper, C., Seiler, S., Van Reenen, J., 2015. The impact of competition on management quality: evidence from public hospitals. Rev. Econ. Studies 82 (2), 457–489.

Brealey, S., Andronis, L., Dale, V., Gibbon, A.J., Gilbert, F.J., Hendry, M., Hood, K., King, D., Wilkinson, C., 2012. The effect of waiting times from general practitioner referral to MRI or orthopaedic consultation for the knee on patient-based outcomes. Br. J. Radiol. 85 (1019), e1134.

Brekke, K., Gravelle, H., Siciliani, L., Straume, O., 2014. Patient choice, mobility and competition among health care providers. In: Levaggi, R., Montefiori, M. (Eds.), Health Care Provision and Patient Mobility: Health Integration in the European Union. Springer Milan, Milano, pp. 1–26.

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

Chan, D.C., 2016. Teamwork and Moral Hazard: evidence from the Emergency Department. J. Political Econ. 124 (3), 734–770.

Chan, D.C., 2018. The efficiency of slacking off: evidence from the emergency department. Econometrica 86 (3), 997–1030.

Chan, D.C., Gruber, J., 2020. Provider Discretion and Variation in Resource Allocation: The Case of Triage Decisions. American Economic Review papers and proceedings 110, 279–283.

Chan, D.C., 2021a. Influence and information in team decisions: evidence from medical residency. Am. Econ. J. Econ. Policy 13 (1), 106–137.

Chan, D.C., 2021b. Influence and information in team decisions: evidence from medical residency. Am. Econ. J. Econ. Policy 13 (1), 106–137.

Chowdhury, M., Dagash, H., Pierro, A., 2007. A systematic review of the impact of volume of surgery and specialization on patient outcome. British J. Surg. Soc. 94 (2), 145–161.

Chung, I.H., 2015. School choice, housing prices, and residential sorting: empirical evidence from inter-and intra-district choice. Reg. Sci. Urban Econ. 52, 39–49.
Com-Ruelle, L., Or, Z., Renaud, T., 2008. The Volume-Outcome Relationship in hospitals: Lessons from the literature. Questions d'économie de La Santé. Institut de Recherche et Documentation en Economie de la Santé, Paris, France.
Cooper, Z., Gibbons, S., Jones, S., McGuire, A., 2012. Does competition improve public hospitals' efficiency? Evidence from a quasi-experiment in the English national health service. CEP discussion papers.. Centre For Economic Performance. London School of Economics and Political Science, London, UK.
Cooper, Z., Gibbons, S., Skellern, M., 2018. Does competition from private surgical centres improve public hospitals' performance? Evidence from the English National Health Service. J Public Econ 166, 63–80.
Dawson, J., Fitzpatrick, R., Carr, A., Murray, D., 1996. Questionnaire on the perceptions of patients about total hip replacement. J. Bone Joint Surg. 78 (2), 185.
De Vries, L., Sturkenboom, M., Verhaar, J., Kingma, J., Stricker, B., 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 Orthop. 82 (5), 545–552.
Department of Health, 2012. A methodology for applying casemix adjustment - Annex C: coefficients for hip replacement models. Patient Reported Outcome Measures (PROMs) in England 1–7. Retrieved from https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms/proms-methodologies#case-mix.
Dimick, J.B., Nicholas, L.H., Ryan, A.M., Thumma, J.R., Birkmeyer, J.D., 2013. Bariatric surgery complications before vs after implementation of a national policy restricting coverage to centers of excellence. J. Am. Med. Assoc. 309 (8), 792–799.
Dranove, D., 1998. Economies of scale in non-revenue producing cost centers: implications for hospital mergers. J. Health Econ. 17 (1), 69–83.
Elixhauser, R., Steiner, M., Harris, M., Coffey, M., 1998. Comorbidity measures for use with administrative data. Med. Care 36 (1), 8–27.
Feng, Y., Pistollato, M., Charlesworth, A., Devlin, N., Propper, C., Sussex, J., 2015. Association between market concentration of hospitals and patient health gain following hip replacement surgery. J. Health Serv. Res. Policy 20 (1), 11–17.
Ferguson, B., Posnett, J., Sheldon, T., 1997. Concentration and choice in the provision of hospital services. Summary Report. University of York, York.
Gaynor, M., Moreno Serra, R., Propper, C., 2013. Death by market power: reform, competition, and patient outcomes in the national health service. Am. Econ. J. Econ. Policy 5 (4), 134–166.
Gaynor, M., Seider, H., Vogt, W., 2005. The volume-outcome effect, scale economies, and learning-by-doing. Am. Econ. Rev. 95 (2), 243–247.
Gaynor, M., Town, R., 2011. Competition in health care markets. In: Pauly, M., McGuire, T., Barros, P. (Eds.), Handbook of Health Economics. Elsevier, pp. 499–637.
Gaynor, M.S., Kleiner, S.A., Vogt, W.B., 2015. Analysis of hospital production: an output index approach. J. Appl. Econ. 30 (3), 398–421.
Geweke, J., Gowrisankaran, G., Town, R., 2003. Bayesian inference for hospital quality in a selection model. Econometrica 71 (4), 1215–1238.
Giancotti, M., Guglielmo, A., Mauro, M., 2017. Efficiency and optimal size of hospitals: results of a systematic search. PLoS ONE 12 (3), e0174533 -e0174533.
Gowrisankaran, G., Town, R., 2003. Competition, payers, and hospital quality. Health Serv. Res. 38 (6p1), 1403–1422.
Gravelle, H., Santos, R., Siciliani, L., 2014. Does a hospital's quality depend on the quality of other hospitals? A spatial econometrics approach. Reg. Sci. Urban Econ. 49, 203–216.
Grieco, P., Mcdevitt, R., 2017. Productivity and quality in health care: evidence from the dialysis industry. Rev. Econ. Studies 84 (3), 1071–1105.
Gutacker, N., Bloor, K., Cookson, R., Gale, C.P., Maynard, A., Pagano, D., Pomar, J., Bernal-Delgado, E., 2017. Hospital surgical volumes and mortality after coronary artery bypass grafting: using international comparisons to determine a safe threshold. Health Serv. Res. 52 (2), 863–878.
Gutacker, N., Siciliani, L., Moscelli, G., Gravelle, H., 2016. Choice of hospital: which type of quality matters? J Health Econ. 50, 230–246.
Halm, E., Lee, C., Chassin, M., 2002. Is volume related to outcome in health care? a systematic review and methodologic critique of the literature. Ann. Intern. Med. 137 (6), 511–520.
Hamilton, B., Hamilton, V., 1997. Estimating surgical volume—Outcome relationships applying survival models: accounting for frailty and hospital fixed effects. Health Econ. 6 (4), 383–395.
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., Mennicken, R., 2018. The volume–outcome relationship revisited: practice indeed makes perfect. Health Serv. Res. 53 (1), 15–34.
Ho, V., 2002. Learning and the evolution of medical technologies: the diffusion of coronary angioplasty. J. Health Econ. 21 (5), 873–885.
Ho, V., 2014. Learning by Doing. In: Culyer, A.J. (Ed.), Encyclopedia of Health Economics. Elsevier, San Diego, pp. 141–145.
Hockenberry, J., Helmchen, L., 2014. The nature of surgeon human capital depreciation. J. Health Econ. 37, 70–80.
Horowitz, J., Keil, S., Spector, L., 2009. Do charter schools affect property values? Rev. Region. Studies 39 (3), 297–316.
Jacobs, R., 2006. Measuring Efficiency in Health care: Analytic Techniques and Health Policy. Cambridge University Press, Cambridge.
Judge, A., Chard, J., Learmonth, I., Dieppe, P., 2006. The effects of surgical volumes and training centre status on outcomes following total joint replacement: analysis of the Hospital Episode Statistics for England. J. Public Health (Bangkok) 28 (2), 116–124.
Kahn, J.M., Ten Have, T.R., Iwashyna, T.J., 2009. The relationship between hospital volume and mortality in mechanical ventilation: an instrumental variable analysis. Health Serv. Res. 44 (3), 862–879.
Kelly, E., Stoye, G., 2016, New joints: private providers and rising demand in the English National Health Service. IFS Working Papers, Institute for Fiscal Studies (IFS), London.
Kessler, D., McClellan, M., 2000. Is hospital competition socially wasteful? Q. J. Econ. 115 (2), 577–615.
Kittelsen, S.A.C., Anthun, K.S., Häkkinen, U., Kruse, M., Rehnberg, C., 2018. Scale and quality in Nordic hospitals. Nordic J. Health Econ. 6 (2), 29–44.
Kizer, K., 2003. The volume–outcome conundrum. The New Eng. J. Med. 349 (22), 2159–2161.
Livingston, E.H., 2009. Bariatric surgery outcomes at designated centers of excellence vs nondesignated programs. Arch. Surg. 144 (4), 319–325.
Losina, E., Barrett, J., Mahomed, N.N., Baron, J.A., Katz, J.N., 2004. Early failures of total hip replacement: effect of surgeon volume. Arthritis Rheumatism 50 (4), 1338–1343.
Luft, H., Hunt, S., Maerki, S., 1987. The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? Health Serv. Res. 22 (2), 157.
McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Economics. Academic Press, New York, pp. 105–142.
Mesman, R., Faber, M.J., Berden, B.J.J.M., Westert, G.P., 2017. Evaluation of minimum volume standards for surgery in the Netherlands (2003–2017): a successful policy? Health Policy (New York) 121 (12), 1263–1273.
Moscelli, G., Gravelle, H., Siciliani, L., Gutacker, N., 2018a. The effect of hospital ownership on quality of care: evidence from England. J. Econ. Behav. Organ. 153, 322–344.
Moscelli, G., Gravelle, H., Siciliani, L., Santos, R., 2018b. Heterogeneous effects of patient choice and hospital competition on mortality. Soc. Sci. Med. 216, 50–58.
Mukoyama, T., 2006. Rosenberg's "learning by using" and technology diffusion. J. Econ. Behav. Organ. 61 (1), 123–144.
Murphy, K.M., Topel, R.H., 1985. Estimation and inference in two-step econometric models. J. Bus. Econ. Stat. 3 (4), 370–379.
Nguyen, N.T., Hohmann, S., Slone, J., Varela, E., Smith, B.R., Hoyt, D., 2010. Improved bariatric surgery outcomes for Medicare beneficiaries after implementation of the medicare national coverage determination. Arch. Surgery 145 (1), 72–78 Chicago, Ill.: 1960.
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, 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 Publishing, Paris.
Ostendorf, M., van Stel, H.F., Buskens, E., Schrijvers, A.J.P., Marting, L.N., Verbout, A.J., Dhert, W.J.A., 2004. Patient- reported outcome in total hip replacement. A comparison of five instruments of health status. J. Bone Joint Surg. Br. 86 (6), 801–808.
Paterson, J.M., Williams, J.I., Kreder, H.J., Mahomed, N.N., Gunraj, N., Wang, X., Laupacis, A., 2010. Provider volumes and early outcomes of primary total joint replacement in Ontario. Canadian J. Surg. 53 (3), 175–183.
Preyra, C., Pink, G., 2006. Scale and scope efficiencies through hospital consolidations. J. Health Econ. 25 (6), 1049–1068.
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., Jenkinson, R., Austin, P., Croxford, R., Wasserstein, D., Escott, B., Paterson, J., Kreder, H., Hawker, G., 2014. Relation between surgeon volume and risk of complications after total hip arhroplasty: propensity score matched cohort study. Br. Med. J. 348 (7961), 14.
Reagans, R., Argote, L., Brooks, D., 2005. Individual Experience and Experience Working Together: predicting Learning Rates from Knowing Who Knows What and Knowing How to Work Together. Manage Sci 51 (6), 869–881.
Schilling, M.A., Vidal, P., Ployhart, R.E., Marangoni, A., 2003. Learning by doing something else: variation, relatedness, and the learning curve. Manage. Sci. 49 (1), 39–56.
Sfekas, A., 2009. Learning, forgetting, and hospital quality: an empirical analysis of cardiac procedures in Maryland and Arizona. Health Econ. 18 (6), 697–711.
Shojania, K., Forster, A., 2008. Hospital mortality: when failure is not a good measure of success. Can. Med. Assoc. J. 179 (2), 153–157.
Singh, J., Kwoh, C., Boudreau, R., Lee, G., Ibrahim, S., 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.
Skellern, M., 2017. The hospital as a multi-product firm: the effect of hospital competition on value-added indicators of clinical quality. CEP Discussion Papers. Centre for Economic Performance. London School Econ. Political Sci..
Street, A., Gutacker, N., Bojke, C., Devlin, N., Daidone, S., 2014. Variations in outcome and costs among NHS providers for common surgical procedures: econometric analyses of routinely collected data. Health Services Delivery Res. 2 (1), 1–90.
Thompson, P., 2010. Chapter 10 - Learning by Doing. In: Hall, B.H., Rosenberg, N. (Eds.), Handbook of the Economics of Innovation. North-Holland, pp. 429–476.
Train, K., 2009. Discrete Choice Methods With Simulation. Cambridge University Press, Cambridge.
Tsai, A.C., Votruba, M., Bridges, J.F.P., Cebul, R.D., 2006. Overcoming bias in estimating the volume-outcome relationship. Health Serv Res 41 (1), 252–264.
Tuominen, U., Sintonen, H., Hirvonen, J., Seitsalo, S., Paavolainen, P., Lehto, M., Hietaniemi, K., Blom, M., 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., 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 Qual. Saf. 24 (3), 195.
Varagunam, M., Hutchings, A., Black, N., 2015b. Relationship between patient-reported outcomes of elective surgery and hospital and consultant volume. Med. Care 53 (4), 310–316.