

HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

WP 18/04

Hospital Choice In The NHS

Valentino Dardanoni; Mauro Laudicella and Paolo Li Donni

February 2018

<http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/>

HOSPITAL CHOICE IN THE NHS

VALENTINO DARDANONI, MAURO LAUDICELLA, AND PAOLO LI DONNI

ABSTRACT. We study hospital choice in the publicly funded National Health Service in England, using a two sample strategy to identify a structural model of demand for elective procedures. In the NHS patients are allowed to opt out from the market of free-of-charge public hospitals and choose a private provider; we find that the outside option has an important effect on competition, patient choice and elasticities compared with traditional models ignoring the private sector. Considering endogeneity of waiting-time, proper measures of quality and the existence of private sector, we find substantially different policy conclusions compared to existing hospital demand models.

JEL Classification Numbers D12, I11, I18, H51.

Keywords Hospital Demand, Patient Choice, Quality, NHS.

1. INTRODUCTION

There is a growing debate on the introduction of patient’s choice in publicly funded health care markets. Supporters of choice argue that, by introducing (more) choice, hospitals will be forced to respond to patients’ preferences and produce better care, since choice drives competition between providers. On the other hand, skeptics argue that patients, or their GP, don’t respond to quality signals as they are unable to observe or understand these signals: market incentives are too weak to be considered by hospital managers.¹

Researchers have studied the implications of hospital choice by estimating hospital demand functions. Hospital demand can be distinguished into two major branches: the demand for emergency procedures (e.g. AMI, stroke etc.), and the demand for elective procedures (e.g. hip replacement, cataract, etc.). The advantage of working with emergency procedures is that endogeneity issues (from selection) are greatly reduced. The disadvantage however, is that is not obvious what “choice” means in emergency. There are many papers in the literature which study the demand for elective procedures in the British NHS and in other National Health Services. The NHS hospital industry

Date: February 21, 2018.

* Valentino Dardanoni: Università di Palermo, Dipartimento SEAS, 90129, Palermo, Italy, valentino.dardanoni@unipa.it. Mauro Laudicella: City University, Northampton Square London, EC1V 0HB, UK, Mauro.Laudicella.1@city.ac.uk. Paolo Li Donni: Università di Palermo, Dipartimento SEAS, Viale delle Scienze, 90129, Palermo, Italy, paolo.lidonni@unipa.it.

¹On the discussion between “Death by market power” *vs.* “Competition kills” see the debate between Bloom *et al.* (2011) and Pollock *et al.* (2011).

is an interesting market since it operates under fix-price reimbursement, and it has no monetary cost for patients.

When studying hospital demand for elective procedures, researchers face two key challenges: the modeling of unobserved hospital quality, and the presence of an Independent Sector of health providers which exists together with public hospitals. Recent empirical IO literature has emphasized the importance of careful modeling of differentiated product markets for realistic policy analysis. The seminal Berry, Levinshon and Pakes (1995) paper in particular has shown the relevance of three key issues in estimation: *i*) allowing for realistic individual taste heterogeneity; *ii*) modeling unobserved product heterogeneity; and *iii*) including an ‘outside option’ in individuals’ choice set.²

In this paper we consider a structural model of hospital choice for elective procedures in the English NHS, taking into account the characteristics of the NHS hospital industry and the nature of the data typically available. Our empirical application considers the demand for primary hip-replacement in NHS hospitals in the period 2006-2009.³ Elective hip replacement is a relatively simple and planned procedure usually performed on elderly patients suffering from arthritis. Patients are free to choose the hospital for their treatment either at their GP practice or at home using the choose and book website. Normally, GPs offer a choice of four to five hospitals including information on waiting time and distance, similar information is available on the choose and book website.

The market for elective hip replacement is self-contained as there is no substitute operation, although the patient can opt for no operation. The market is served by a number of different providers, NHS public hospitals and NHS treatment centres, independent hospitals and independent sector treatment centres (ISTC). NHS public hospitals are large multi-service organization, while NHS treatment centres are public health centres specialized in few planned procedure performed routinely; we will refer to both as *public providers*. Independent hospitals are privately owned organizations offering few elective procedures to privately insured patients; ISTC are a growing subgroup of Independent hospitals that can provide services to both privately insured and publicly funded NHS patients. We will refer to these as *independent providers*. Virtually all public providers and a large share of independent providers are able to offer hip replacement operations, making it an ideal candidate for studying choice and competition in the health market. Hospital choice in hip replacement was implemented since 2006 (later for some other planned procedures). Finally, hip replacement is also less likely to be associated with a more complex health situation that might affect choice.

² See e.g. Einav and Levin (2010) for a review of the empirical IO literature and Nevo (2011) for a review of empirical models of consumer behavior.

³See Beckert *et al.* (2012); Beukers *et al.* (2014), Moscelli *et al.* (2016).

Our first contribution to the existing literature is to explicitly model the presence of an outside option in hospital demand. Most papers on hospital choice do not consider the possibility that patients choose any option different from the menu of public providers considered by the analyst, due to the lack of data on independent providers. However, there are in fact many reasons why a patient in need for an elective procedure may not choose from this menu. First of all, a large number of patients uses independent providers for hip replacement; in the period considered, an average of 18% of total procedures were performed by independent providers. Recent NHS policies have encouraged privately-owned hospitals to enter the market for publicly funded health care and directly compete with NHS hospitals; this trend is likely to continue since unsustainable health expenditure may call for further expansion of the role of the private sector as a top up/complement or substitute to public providers.⁴ Secondly, many analysts exclude from their sample NHS hospitals which perform a very small number of the procedures per year.⁵ Thirdly, a small fraction of patients may decide to have the hip replaced outside England (e.g. in other UK regions or overseas). Finally, some patients may simply decide not to have the surgery done, e.g. because they are discouraged by long waiting times, or receive conflicting medical advice, or decide to postpone the treatment for various reasons (e.g. level of severity or simply inertia).⁶

A second motivation of our study is that observed measures of hospital quality typically used in current literature are imperfect and much debated.⁷ Since it is likely that a large component of hospital quality is not captured by observable characteristics, there is potentially a large endogeneity problem in estimation. In particular, in fixed-price hospital industries, waiting time plays a similar equilibrium role as that of price in standard industries. If hospital quality is imperfectly controlled, we expect potentially large biases in waiting time elasticity estimates. We argue that control for endogeneity has been less than perfect in current literature, since it has relied on fixed effects, which are known not to be fully adequate for controlling for endogeneity (see e.g. Nevo, 2000). Our choice model allows us to recover an estimate of unobserved quality and the use of linear IV estimation to address the endogeneity issue.

Most of the studies on hospital choice in England are based on data from the Hospital Episode Statistics (HES), which includes only services provided free of charge

⁴See Beckert and Kelly (2017), and Kelly and Stoye (2015).

⁵This is similar to studies in empirical IO where products with very small market shares are excluded from the dataset.

⁶Mota *et al.* (2012) argue that age, gender, race and socio-economic disparities suggest that those who need total joint arthroplasty may not receive it: “it is not clear whether doctors limit treatment opportunities to patients, nor is it known the effect that patient beliefs and expectations about the operation, including their paid work status and retirement plans, have on the decision to get a surgery”.

⁷A recent study on NHS hospital quality by Gravelle *et al.* (2012) shows very low correlations between different routinely used quality measures.

to NHS patients.⁸ Thus HES based data is typically a *selected sample*, and hospital choice models based on it may imply wrong policy conclusions.⁹ We use a novel two-sample strategy, where we create a synthetic sample representing the universe of over 65 patients in need for a hip replacement using small area administrative data from the UK Census, and then match these administrative data with HES data which contain information on actual choices by NHS patients.

As will be shown in our empirical application, addressing the outside option and endogeneity issues is key for estimating sound models of the hospital industry, giving correct and useful policy guidance. Compared to standard models currently used in the literature, our approach shows that estimated own and cross elasticities for waiting time and observed quality are much greater than previously found. We also address the much debated issue on the relationship between competition and quality of care. We find that the correlation between overall quality and market concentration is strongly positive using the standard model, while it becomes strongly negative in our model, which addresses the endogeneity of quality and hospitals' location choices. A closure simulation exercise also predicts quite different counterfactuals using the two approaches.

The paper is structured as follows. In the next section we describe the institutional details of the market for hip replacement in the England in the period 2006-2009, and explain our two-sample strategy employed in estimation. Section three introduces our hospital demand model and describes our estimating strategy. Section four illustrates the estimation results. Section five concludes.

2. INSTITUTIONAL DETAILS AND DATA

Patient's freedom of choice of health care provider for elective procedures was introduced by law in January 2006.¹⁰ Prior to this date, NHS patients had little or no choice; they were referred to a specific hospital by their GPs by selective contracting arrangements. After January 2006, patients were able to choose from a list of four to five different providers, including ISTC, provided by their GP. Alternatively, patients were able to select their providers from home by using the choose and book website managed by the NHS. Patients, and their GPs, were also given greater information

⁸Beckert and Kelly (2017) estimate a multinomial logit model of demand for NHS hospitals and independent providers using HES data, considering only NHS-funded inpatient treatments.

⁹This holds for procedures where there is the possibility of having the procedure performed outside an NHS hospital; procedure such as CABG considered by Gaynor *et al.* (2016) are performed only in a small number of HNS hospitals and the option of not to receive medical treatment is not really a feasible alternative.

¹⁰A seminal paper comparing hospital choice before and after the 2006 reform is Gaynor *et al.* 2016, which models choice before 2006 using GP's induced consideration sets, and a standard discrete choice model after 2006.

on which to make this choice, such as risk-adjusted mortality rates, waiting times, infection rates and hospital activity rates.

Hip replacements are delivered in almost all NHS hospitals. The market for hip replacement is characterised by the existence of a large independent sector performing the procedure in non emergency cases: the independent sector share for elective hip replacement procedures was about 30% in 2003 and 15% in 2009 (see Table 1 below). It is very interesting to note that NHS hospitals' waiting times for the procedure are strongly correlated with the independent sector share, as shown in the table below:

TABLE 1. Independent sector size and average waiting time by year

	2003	2004	2005	2006	2007	2008	2009
wt	6.1771	5.4013	4.8297	3.8171	2.7818	2.8911	3.0097
IS share	0.3147	0.2627	0.2234	0.2171	0.1857	0.1731	0.1540

2.1. A Tale of Two Samples. Suppose we want to estimate a parametric model of hospital choice

$$Pr(\text{patient } i \text{ chooses } j \mid \mathbf{x}_{ij}, \boldsymbol{\beta}), \quad j = 0, \dots, J \quad (1)$$

where $j = 1, \dots, J$ indicizes the set of J NHS hospitals which perform elective hip replacement, $j = 0$ denotes the outside option, \mathbf{x}_{ij} denotes the set of variables affecting patient's i choice of option j , and $\boldsymbol{\beta}$ is the parameter to be estimated.

Suppose we have data on \mathbf{x}_{ij} and on actual choices (denoted C_{ij} , which takes value one if patient i chooses hospital j) only for $j = 1, \dots, J$, that is, only for the set of patients who have chosen a NHS hospital. In other words, we have a selected sample of the population. Let $S = \{1, \dots, n\}$ denote this selected sample, and let $U = \{1, \dots, N\}$ denote the universe of all English over 65 patients in need of elective hip replacement. The problem is that we do not observe U .

Suppose we have external information of the total number of patients in U which did not choose a NHS hospital, say $N_0 = N - n$. If we knew the nature of the sample selection process, we could appropriately generate a synthetic sample of N_0 patients to append to the existing sample S to create a size N sample representative of U . In practice, since we are agnostic about the precise nature of the selection process, we use a different approach: we generate a synthetic sample of size N —say U' —using administrative data, reproducing the population of over 65 English patients in need of elective hip replacement.

Note that while under sufficiently rich administrative data, U' may contain all the variables \mathbf{x}_{ij} needed to calculate the choice probabilities (1), U' of course does not contain real patients, and thus does not contain data on actual choices C_{ij} . In section 3.3 we show how we can use observed sample moments in both the actual NHS sample S and in our synthetic sample U' for parameter estimate.

2.2. The samples we use. We use two different datasets: the HES, which collects the universe of inpatient discharges receiving hip replacement from every hospital in the NHS in England; and administrative data from Lower-Layer Super Output Areas (LSOAs), which are a geographic hierarchy designed to improve the reporting of small area statistics in England.

2.2.1. Hospital Episode Statistics. Data on patient admissions are extracted from the UK Department of Health’s Hospital Episode Statistics (HES), which comprise records of all publicly funded patients admitted to hospitals in England. We include in our study all hospital admissions during the fiscal year 2006 to 2009 of English patients aged 65 and over receiving a bilateral or primary hip replacement cemented or uncemented (HRG code: H01 H02 H80 H81) at the time of admission.

We include elective admissions from waiting lists and booked admissions, i.e. patients without date of admission and patients having a booked date of admission. We excluded a small share of planned admissions (4.2% of total admissions), i.e. admissions that are part of a planned sequence of clinical care determined mainly on clinical criteria, rather than hospital capacity.¹¹ We also drop patients with implausible waiting time (longer than 3 years), and consider only hospital trusts with more than 50 relevant admissions per year. Our sample includes 27,962, 29,604, 31,206, 31,875 patients respectively treated in each year from 2006 to 2009.

For each patients we collect two key variables: waiting time and place of residence. At the patient-level we observe the time elapsed between the referral and the actual treatment, and use this information to construct the hospital-level waiting time measured as the average number of month patients wait before having an inpatient admission. The HES also contains information on the postal code of the neighbourhood in which the patient lives, which identifies her LSOA using the GSS coding of the Office for National Statistics. Data from the NHS Organisation Data Service (ODS) provides each hospital’s address, which we use to measure the straight line travel distance (in km) between the centroid of the LSOA where patient lives and each hospital considered.

Following the majority of NHS hospital choice studies, since the HES does not collect data on patient’s economic variables, we use the socioeconomic Index of Multiple Deprivation of the LSOA of patient’s residence as a proxy for the patient’s economic condition. The Indexes of Multiple Deprivation are indicators of small area deprivation explicitly designed to capture the multidimensional aspects of socioeconomic deprivation at LSOA geographical level. The IMD income deprivation domain measures the proportion of the LSOA population living in low-income households reliant on one or

¹¹We only include elective patients because we know elective patients may plausibly choose their hospital; planned elective spells are excluded because for these spells the waiting time before an operation is for clinical reasons and not due to lack of capacity (HES Online, 2015).

more means-tested benefits, based on population census and benefit claims data (Noble *et al.*, 2006).

The IMD Health Deprivation and Disability Domain also identifies areas with relatively high rates of people dying prematurely, or whose quality of life is impaired by poor health, or who are disabled. This domain measures morbidity, disability and premature mortality at the LSOA level (Noble *et al.*, 2006). Henceforth, we will call ‘income deprivation’ the socioeconomic Index of Multiple Deprivation, and ‘health deprivation’ the health Index of Multiple Deprivation.

The HES data is linked to a number of other sources to provide additional information on hospital characteristics. We use data from the National Centre for Health Outcomes Development (NCHOD) and the Care Quality Commission (CQC) to collect information on hospital quality, capacity and costs.

As capacity measures we consider the number of sites in the hospital trust, the total number of beds available, the number of doctors, qualified and unqualified nurses and allied practitioners. Potential costs’ differences between hospitals are measured using the Market Forces Factor (MFF), which is an estimate of unavoidable cost differences between health care providers, based on their geographical location. MFF is used to adjust resource allocations in the NHS in proportion to these cost differences (Monitor, 2013).

Quality of health care is generally multidimensional and intrinsically not observable. For this reason a common strategy is to use a set of indicators to capture it. Although there is not a clear view on how properly these indicators capture health care quality (Gravelle *et al.*, 2012), we follow the literature and include: the Care Quality Commission’s (*CQC*) quality rating (available from the Department of Health); the incidence of Methicillin-Resistant Staphylococcus Aureus (*MRSA*) infections in 2006-7 (published by the Health Protection Agency); the standardized mortality rate (*SMR*) for hip fracture over period 2006-8; and a measure of hospital’s predicted performance based on readmission (*READ*) after a hip fracture (see Laudicella *et al.*, 2013).

2.2.2. Lower-Layer Super Output Areas. The second source of data comes from the UK Census which provides detailed information on LSOAs’ characteristics. LSOA are geographical units developed by the National Office of Statistics to improve the reporting of small area statistics for the in UK (Briggs *et al.*, 2007). There are about 32,482 LSOAs in England for the period we consider, with a minimum population size of 400 households and 1,000 individuals, and a maximum population of 1,200 households and 3,000 individuals. For each LSOAs we collect the total number of population over 65: the average total population over 65 is 252. We also collect the socioeconomic and health IMD indices for each LSOA, and using the geographical location of its centroid, the distance between each LSOA and each NHS hospital considered.

Our strategy is to couple the sample extracted from the HES dataset, with a synthetic sample which uses LSOA's data to mimic the over 65 English population seeking hip replacement surgery. For this purpose, for robustness, we build two synthetic LSOA samples under different scenarios:

- (1) a smaller sample, where the total number of patients seeking a hip replacement equals to the number of elective hip replacement procedures performed in all England by the over 65 in those years, as reported by the Annual reports of the National Joint Registry (see Table 2.3 of the National Joint Registry 2010 for 2009).¹² Table 1 reports the share of hip replacements received by privately funded patients in England. Using this information, the size of the outside option can be easily recovered by adding all patients privately funded to all patients who are NHS funded, but received an hip replacement either in a private hospital or in an ISTC or in NHS treatment centre.¹³ Notice however that, although the NJR report is an important source of information to appraise the size of the IS in England, the average compliance in the period considered is about 85%, giving an underestimate of the total performed procedures.¹⁴
- (2) a larger sample, which uses, after review of epidemiological studies on the incidence of hip replacement across different countries and times, and after personal consultation with various health care professionals, an informed guess on the incidence of hip replacement for over 65 patients of 5 procedures for 1000 individuals.

We extract the total size of the over 65 English population from the LSOA data. The size of the first synthetic sample, which –for $t = 2006, 2007, 2008, 2009$, we denote $N_t^{S,1}$ – is equal to 35,716, 36,355, 37,739 and 37,677. The size of the second synthetic sample, which we denote $N_t^{S,2}$, is equal to 40,429, 40,797, 41,426 and 42,001. For each time t and in each of the two cases, we sample, with replacement, $N_t^{S,1}$ and $N_t^{S,2}$ units in proportion to the total population over 65 in each LSOA.

2.3. Using the two samples. Both the HES and the synthetic LSOA samples we use contain the observable variables affecting patients' hospital choice as specified by her indirect utility function, with a qualification: individual economic and health status is in both samples is proxied by the socioeconomics and health IMD indices.

¹²The NJR provides an annual report containing a rich set of aggregate information on the amount of hip and knee replacements delivered in England and Wales based on information submitted by private and public hospitals. In particular it reports the number of total hip replacements performed by types of provider (NHS hospital, independent hospital, NHS treatment centre and Independent Sector Treatment Centres (ISTC)) and source of funding (Independent or NHS).

¹³We include patients treated in a NHS treatment centre in the outside sector since, differently from the hospital trusts, these centres may also depend to the local health authorities.

¹⁴NJR defines compliance rate as the rate, expressed as a percentage, of procedure records submitted to the NJR compared with the levy returns for the number of implants sold.

The HES dataset does not record patients' economic status variables, and our approach proxying socioeconomic status is standard. On the other hand, only the HES dataset contains information on patient individual health; a drawback of our two-sample strategy is that we have to proxy individual patients' health status by the health IMD index without using actual individual health information - as it is not contained in the LSOA sample. We test the robustness of our approach by comparing, in our NHS sample, logit estimates obtained using individual level information on patient's health status (the patient's number of secondary diagnoses) rather than the health IMD index. We found that estimates give very similar conclusions on all our policy results.

Finally, it is worth recalling that, since the LSOA is a synthetic sample with fictitious patients, only the HES sample has information on patient hospital choices.

3. HOSPITAL DEMAND

Patient's choice depends on hospital characteristics such as the distance from the patients' residence, the time she has to wait to get the procedure, and the quality of hospital care. The indirect utility of patient i at time t for NHS hospital $j = 1, \dots, J_t$ is given by

$$U_{ijt} = \beta_i w_{jt} + \gamma_i d_{ij} + \eta a_{ij} + q_{jt} + \epsilon_{ijt}, \quad (2)$$

where, at time t , q_{jt} denotes hospital j 's quality, which is unobserved; w_{jt} is average waiting time (in months) for hospital j ; d_{ij} is (the log of) the distance (in kilometres) between the residence of patient i and hospital j ; a_{ij} is a dummy variable which takes value one if hospital j is in the 'attention area' of patient i (namely, j is in the attention area of patient i if it is either within a distance of 20 km or is one of the 5 closest hospitals to patient i); and ϵ_{ijt} is an i.i.d. extreme value individual preference shifter.

In this formulation, there are two types of heterogeneity in patients' preferences: the purely idiosyncratic shifter ϵ_{ijt} , and the marginal (dis)utility for distance and waiting time β_i and γ_i . We model taste heterogeneity for distance and waiting time in terms of observable patients' characteristics and an idiosyncratic random term:

$$\beta_i = \beta_0 + \beta_I I_i + \beta_H H_i + \sigma_w R_{w,i}$$

$$\gamma_i = \gamma_0 + \gamma_I I_i + \gamma_H H_i + \sigma_d R_{d,i}$$

where I_i and H_i denote patient i economic and health deprivation, and $R_{w,i}$ and $R_{d,i}$ are distributed in the population as standard normal variates. As well known in the discrete choice literature, modelling taste heterogeneity is key for estimating realistic substitution patterns between products, since shares do not satisfy the restrictive I.I.A. assumption.

Contrary to most literature on hospital demand,¹⁵ we assume that patient can choose an outside option (which we denote $j = 0$). As discussed above, the outside option contains all possible alternatives that a patient needing a hip replacement procedure may choose besides one of the J_t NHS hospitals we consider. The utility of patient i from choosing the outside option ($j = 0$) at time t is

$$U_{i0t} = \alpha_i + \epsilon_{i0t} \quad (3)$$

where ϵ_{i0t} is an i.i.d. extreme value preference shifter and

$$\alpha_i = \alpha_{0t} + \alpha_{I1}I_i + \alpha_{I2}I_i^2 + \alpha_{H1}H_i + \alpha_{H2}H_i^2 + \sigma_0 R_{0,i}$$

with $R_{0,i}$ standard normal distributed.

It turns out that it is quite useful to decompose the utility of choosing a NHS hospital $j = 1, \dots, J_t$ into: *i*) a component which does not vary among patients', say $\delta_{jt} = \beta_0 w_{jt} + q_{jt}$; *ii*) a component $\mu_{ijt} = (\beta_I I_i + \beta_H H_i + \sigma_w R_{w,i})w_{jt} + (\gamma + \gamma_I I_i + \gamma_H H_i + \sigma_d R_{d,i})d_{ij} + \eta a_{ij}$ which captures individual patients' heterogeneity (excluding the error term); *iii*) the purely idiosyncratic logit error ϵ_{ijt} . Therefore,

$$U_{ijt} = \delta_{jt} + \mu_{ijt} + \epsilon_{ijt}, \quad j = 1, \dots, J_t, \quad (4)$$

while

$$U_{i0t} = \delta_{0t} + \mu_{i0t} + \epsilon_{i0t}, \quad (5)$$

with $\delta_{0t} = \alpha_{0t}$ and $\mu_{i0t} = \alpha_{I1}I_i + \alpha_{I2}I_i^2 + \alpha_{H1}H_i + \alpha_{H2}H_i^2 + \sigma_0 R_{0,i}$.

This formulation makes it clear that in this choice model the only source of endogeneity is included in the constants δ_{jt} , which are defined for each hospital j at each time t , and so they absorb all hospitals' unobservable characteristics which may be correlated with the observable variables contained in the utility function. In particular, unobservable quality –which is typically strongly correlated with waiting time– has been subsumed into the constants δ_{jt} .

3.1. Choice Probabilities in the Two Samples. As discussed above, we use two samples: the HES sample collecting the universe of NHS financed patients which have chosen one of the J_t NHS hospitals, and a (synthetic) LSOA samples which mimicks the over 65 population in England which seek a hip replacement procedure.

Omitting t for simplicity, the probability that individual h in the HES sample chooses hospital j is

$$P_{hj}^N = \frac{\exp(\delta_j + \mu_{hj})}{\sum_{k=1}^J \exp(\delta_k + \mu_{hk})}, \quad j = 1, \dots, J, \quad (6)$$

¹⁵This holds not only for the HNS type models discussed above, but also for those referring to different hospital industries such as Capps *et al.* (2003), Ho (2006) and Ho (2009).

while the probability that an individual s in a LSOA sample chooses option j is

$$P_{sj}^S = \frac{\exp(\delta_j + \mu_{sj})}{\sum_{k=0}^J \exp(\delta_k + \mu_{sk})}, \quad j = 0, 1, \dots, J. \quad (7)$$

Notice that δ_j , which captures hospital characteristics, is common in the two samples.

3.2. Estimation. We follow Goolsbee and Petrin (2004) and Train and Winston (2007) and estimate the model in two stages. In the first stage we estimate the mean utilities δ and the parameters included in μ_{ijt} . These are collected as

$$\theta_t = [\underbrace{\delta_t}_{P^S, P^N}, \underbrace{\alpha_0, \alpha_{I1}, \alpha_{I2}, \alpha_{H1}, \alpha_{H2}}_{P^S}, \underbrace{\beta_I, \beta_H, \gamma, \gamma_I, \gamma_H, \eta}_{P^N, P^S}, \underbrace{\sigma_0}_{P^S}, \underbrace{\sigma_I, \sigma_d}_{P^N, P^S}]. \quad (8)$$

illustrating how the θ_t parameters belong to P^N and P^S . Estimation of θ_t is implemented by simulated GMM. In the second stage we use estimated hospitals' mean utilities $\hat{\delta}_t$ to estimate waiting time and observable quality parameters, correcting for endogeneity by TSLS.

3.3. Moments. To simplify notation, let $\bar{P}_{sjt}^S(\theta_t; \mathbf{z}_s^S)$ denote the expected probability for patient s in the LSOA sample to choose hospital j at time t , where \mathbf{z}_s^S collects the variables which enter the utility function in the LSOA sample. $\bar{P}_{sjt}^S(\theta_t; \mathbf{z}_s^S)$ is the integral of $P_{sjt}^S(\theta_t; \mathbf{z}_s^S)$ over the distribution of the random variables R_w, R_h, R_0 . In practice we approximate $\bar{P}_{sjt}^S(\theta_t; \mathbf{z}_s^S)$ by simulation, using 100 antithetic Halton draws of the standard normal variables R_w, R_h, R_0 .¹⁶

We use *three sets* of moments:

- (1) *The BLP Moments:* we equate the observed aggregate hospital shares S to the average probabilities in the LSOA sample:

$$S_{jt} = \frac{1}{N^S} \sum_s \bar{P}_{sjt}^S(\theta_t; \mathbf{z}_s^S). \quad (9)$$

Berry (1994) shows that the predicted shares can be inverted to get the vector δ_t , for any value of the remaining parameters in θ_t .

- (2) *The HES Moments:* in the HES sample we set standard observation-specific moments:

$$\mathbf{g}_{hjt}^N(\theta_t) = \left(C_{hjt} - \bar{P}_{hjt}^N(\theta_t; \mathbf{z}_h^N) \right) (\mathbf{z}_{hjt}^H, \mathbf{v}_{jt}), \quad (10)$$

where C_{hjt} is the choice variable which takes value one if individual h at time t choose hospital j and \mathbf{v} is an appropriate vector of hospital characteristics.

In our application we use hospital dummies for teaching, acute and London

¹⁶The similar notation $\bar{P}_{hjt}^N(\theta_t; \mathbf{z}_h^N)$ is used for the expected probability of patient h in the HES sample to choose hospital j at time t .

hospitals, and hospital capacity variables such as the number of beds, doctors, qualified and unqualified nurses and health practitioners, with squares and interactions.

- (3) *The LSOA Moments:* we link the HES and LSOA samples by matching observed attributes in the HES sample with those in the LSOA.¹⁷ Consider for example income deprivation I . From observed choices in the HES sample, we derive the average income deprivation of patients using hospital j at time t , namely $\frac{1}{N_t^N} \sum_{h=1}^{N_t^N} I_h C_{hjt} / S_{jt}^N$, and, using Bayes' rule, we match this with the expected deprivation of patients using hospital j in the LSOA sample under the theoretical choice probabilities $\bar{P}_{s jt}^S(\boldsymbol{\theta}_t; \mathbf{z}_s^S)$:

$$\mathbf{g}_{s jt}^I(\boldsymbol{\theta}_t) = \left(\frac{1}{N_t^N} \sum_{h=1}^{N_t^N} I_h C_{hjt} / S_{jt}^N - I_s \bar{P}_{s jt}^S(\boldsymbol{\theta}_t; \mathbf{z}_s^S) / S_{jt}^S \right) \bar{\mathbf{v}}_{jt}, \quad (11)$$

where $\bar{\mathbf{v}}$ denotes the vector of hospital characteristics \mathbf{v} above, plus a constant.

By a similar reasoning, we derive another set of sample moments by matching health deprivation H

$$\mathbf{g}_{s jt}^H(\boldsymbol{\theta}_t) = \left(\frac{1}{N_t^N} \sum_{h=1}^{N_t^N} H_h C_{hjt} / S_{jt}^N - H_s \bar{P}_{s jt}^S(\boldsymbol{\theta}_t; \mathbf{z}_s^S) / S_{jt}^S \right) \bar{\mathbf{v}}_{jt}. \quad (12)$$

3.4. First Stage Estimation. Stack these individual moments to get

$$\mathbf{g}^N(\boldsymbol{\theta}) = \sum_t \sum_j \sum_h \mathbf{g}_{h jt}^N(\boldsymbol{\theta}_t); \quad \mathbf{g}^S(\boldsymbol{\theta}) = \sum_t \sum_j \sum_s (\mathbf{g}_{s jt}^I(\boldsymbol{\theta}_t), \mathbf{g}_{s jt}^H(\boldsymbol{\theta}_t))$$

$$\mathbf{S}^N(\boldsymbol{\theta}) = \sum_t \sum_j \sum_h \mathbf{g}_{h jt}^N(\boldsymbol{\theta}_t) \mathbf{g}_{h jt}^N(\boldsymbol{\theta}_t)'; \quad \mathbf{S}^S(\boldsymbol{\theta}) = \sum_t \sum_j \sum_s (\mathbf{g}_{s jt}^I(\boldsymbol{\theta}_t), \mathbf{g}_{s jt}^H(\boldsymbol{\theta}_t)) (\mathbf{g}_{s jt}^I(\boldsymbol{\theta}_t), \mathbf{g}_{s jt}^H(\boldsymbol{\theta}_t))'$$

and define

$$\mathbf{g}(\boldsymbol{\theta}) = (\mathbf{g}^N(\boldsymbol{\theta}), \mathbf{g}^S(\boldsymbol{\theta})), \quad \mathbf{S}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{S}^N(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^S(\boldsymbol{\theta}) \end{pmatrix}. \quad (13)$$

To estimate $\boldsymbol{\theta}$ we use two-step Simulated GMM: in the first step we estimate $\boldsymbol{\theta}_1 = \argmin \mathbf{g}(\boldsymbol{\theta})' \mathbf{g}(\boldsymbol{\theta})$; in the second step we find $\hat{\boldsymbol{\theta}} = \argmin \mathbf{g}(\boldsymbol{\theta})' \mathbf{S}(\boldsymbol{\theta}_1)^{-1} \mathbf{g}(\boldsymbol{\theta})$. At each step, and within each iteration of the minimization problem, we also use the aggregate constraints (9) which equate observed market shares with theoretical ones. Estimated s.e. are corrected using Pollard and Pakes (1989) procedure taking into account the simulation.

3.5. Second stage. In the second stage we recover the parameters which enter the mean utility vector $\boldsymbol{\delta}$. Let $\hat{\delta}_{jt}$ denote the estimated mean utilities from the first stage. Decompose hospital quality q_{jt} into a time and hospital¹⁸ fixed effects Δ_t and Δ_j and

¹⁷Imbens and Lancaster (1994) discuss using macro moments in micro models with choice based samples (see also Petrin (2002) and Berry, Levinsohn and Pakes (2004)). We create a synthetic sample representing the universe of patients, reproducing at the micro level patients' characteristics for all options (including the outside one).

¹⁸We use fixed effects for hospitals which are present in the sample for at least two periods.

observed and unobserved residual components of hospital quality, to get the regression equation

$$\hat{\delta}_{jt} = \beta_0 w_{jt} + \mathbf{x}_{jt}' \boldsymbol{\phi} + \Delta_t + \Delta_j + \xi_{jt} \quad (14)$$

where \mathbf{x}_{jt} denotes a vector of observed quality measures which includes: *CQC*, *SMR*, *MRSA* and *READ*.

Clearly equation (14) is subject to a large endogeneity problem, since waiting times are much correlated with hospital quality, and hospitals fixed effect do not necessarily address the endogeneity problem adequately (see e.g. Nevo, 2000). We estimate equation (14) by TSLS. The instruments we use are the Market Factor Forces variable (*MFF*) with its square, and the average number of beds and sites of the other hospitals in the market.

4. RESULTS

In this section we report estimated parameters. We estimate three models:

- (1) *The Standard Logit Model*. As benchmark, we estimate the standard logit model which is the workhorse of most NHS hospital demand studies. To estimate the first stage, we use GMM with the set of moments given by (9) and (10), without random coefficients and outside option.¹⁹ To appraise the effect of waiting time endogeneity, in the second stage we estimate two regressions, one where parameters are estimated by OLS, and one using TSLS for better control for endogeneity. The Logit model with fixed hospital effect (our OLS model) can be considered the prototype of many current papers on NHS hospital demand (e.g. Sivey (2012), Beckert *et al.* (2012), Gaynor *et al.* (2016).).
- (2) *Our Two-Sample Model (2SM)*. We estimate our full model using our two sample strategy with two different LSOA samples, namely the NJR-LSOA and the Epidemiological-LSOA ones, as described in Section 2.2.2.

4.1. First stage. Table 2 reports estimates of the $\boldsymbol{\theta}$ parameters in (8). Estimates for the Logit Model are reported in the first two columns of table 2, while the remaining columns describe the parameters of the NJR-2SM and Epidemiological-2SM ones.

In keeping with most studies of hospital demand for elective procedures,²⁰ we find that in all models distance strongly affects patient choice; patients are significantly likely to choose a closer hospital, with the 'attention area' dummy being strongly significant. There is also significant demographic heterogeneity on patients' disutility for distance: the disutility from travelling is increasing for individuals coming from higher

¹⁹For robustness we have also estimated the first stage by Maximum Likelihood, with very similar results.

²⁰Capps *et al.* (2003), Ho (2006), Sivey (2012) Beckert *et al.* (2012), Beckert and Kelly (2017), Gaynor *et al.* (2016), Moscelli *et al.* (2016).

income deprived areas, but decreasing for those coming from more health deprived areas, indicating that poorer patients tend to choose closer hospital, while more sick individuals have a higher willingness to travel. Demographic heterogeneity on patients' disutility from waiting time is lower for those coming from more income deprived areas and higher for those coming from more health deprived areas, but these effect are generally not very significant.

In our two models we also estimate the probability of choosing the outside option, and idiosyncratic waiting time and distance taste heterogeneity. In both models it emerges that the outside option is more likely to be chosen by patients coming from wealthier and more health deprived areas. Differences are not substantial between the IS model and the epidemiological one, which gives some evidence that these results are quite robust to changes in measuring the size of the outside sector. Looking at the three random coefficient estimates, it emerges that there is a large taste heterogeneity for distance and for the outside option, but not for waiting time.

TABLE 2. First Stage

	Logit		NJR 2SM		Epi. 2SM	
	Coef.	SE	Coef.	SE	Coef.	SE
wt*I	0.0137	0.0142	-0.0898	0.0336	-0.0377	0.0339
d*I	-0.2556	0.0232	-0.7894	0.0433	-0.8898	0.0451
wt*H	-0.0176	0.0146	0.0190	0.0308	0.0596	0.0306
d*H	0.3197	0.0127	0.8584	0.0386	0.8917	0.0386
d	-3.3187	0.0132	-8.3739	0.1175	-8.1910	0.1075
a	1.6117	0.0314	2.9380	0.1471	2.9304	0.1546
I			-3.3548	0.3740	-3.3379	0.3170
I ²			1.1276	0.1332	1.0687	0.1170
H			5.2416	0.2070	5.2135	0.1697
H ²			-0.4306	0.1206	-0.2364	0.0918
$R_{0,s}$			3.6250	0.6618	3.5686	0.6168
$R_{w,s}$			0.5445	0.0520	0.1700	0.0510
$R_{d,s}$			2.9146	0.0987	2.9928	0.0934

4.2. Second stage. Table 3 reports estimated coefficients for the second stage regression as described in Section 3.5 with and without endogeneity control by IV. Waiting time has a significant negative effect on patients' utility, so that hospitals with longer waiting time are less likely to be chosen. The key observation is that this effect is much bigger when one controls for endogeneity by both fixed effects and IV, rather than by fixed effects only as commonly done in this literature.²¹

²¹The F-test statistics of the joint significance of the instruments in the standard first-stage IV regression equals to 9.90. This indicates that the null of weak identification of the endogenous variable can be rejected.

Focusing on observed quality measures, as expected estimated quality q_j is positively correlated with observed quality indicators. However, only MRSA infections have a consistently significant effect across models. This is not surprising since CQC hospital classification shows little variation (very few hospital are classified as poor) and also looks at many different aspects of quality that are likely to be driven by the standard of care in acute and emergency services. SMR are driven by the standards of care for operations with a significant mortality risk, which is not the case of hip replacement. MRSA appears to be the indicator with a strong logical link to the quality of care for hip replacement patients, since patients are at risk of this type of infections.

TABLE 3. Second Stage

	OLS		TSLS	
	Coef.	SE	Coef.	SE
<i>Panel A:</i>				
<i>Logit</i>				
CQC	0.0512	0.0532	0.0224	0.0547
SMR	-0.0003	0.0227	0.0012	0.0188
Read	-0.4749	0.7916	-0.8051	0.7271
MRSA	-0.2193	0.1586	-0.2821	0.0991
Waiting Time	-0.1294	0.0350	-0.4320	0.1147
Fixed Effects	Yes		Yes	
<i>Panel B:</i>				
<i>NJR 2SM</i>				
CQC	0.1083	0.1926	0.0244	0.2168
SMR	0.0442	0.0776	0.0486	0.0744
Read	-1.7540	2.8844	-2.7163	2.8827
MRSA	-0.8596	0.4573	-1.0426	0.3927
Waiting Time	-0.6494	0.1491	-1.5315	0.4546
Fixed Effects	Yes		Yes	
<i>Panel C:</i>				
<i>Epi. 2SM</i>				
CQC	0.0391	0.1755	-0.0617	0.1971
SMR	0.0168	0.0750	0.0221	0.0677
Read	-3.3398	2.6267	-4.4956	2.6204
MRSA	-0.9503	0.4751	-1.1701	0.3570
Waiting Time	-0.3964	0.1255	-1.4558	0.4132
Fixed Effects	Yes		Yes	

4.3. Elasticities. In this section we calculate own and cross elasticities of hospital demand with respect to waiting time and an observed quality measure (namely MRSA, the only consistently significant indicator). For each market t we first calculate the $J_t \times J_t$ matrix of elasticities, with the J_t -sized diagonal containing the own elasticity.

We then report the mean and standard deviation of these J_t own elasticities for each time t .

Regarding cross elasticities, we first compute the maximum of the off-diagonal elements for each row of the $J_t \times J_t$ matrix (that is, the cross elasticity between each hospital j and its greatest competitor). Then for each time t , we find the mean and standard deviation of these values which are reported in the table.

TABLE 4. Average Waiting Time Elasticities

Year	Elasticities				Cross Elasticities			
	OLS		TSLS		OLS		TSLS	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
<i>Panel A:</i>								
<i>Logit</i>								
2006	-0.2146	0.0745	-0.7182	0.2429	0.0630	0.0339	0.2101	0.1100
2007	-0.1701	0.0705	-0.5711	0.2352	0.0490	0.0300	0.1640	0.0981
2008	-0.1272	0.0580	-0.4241	0.1899	0.0364	0.0233	0.1209	0.0750
2009	-0.1341	0.0638	-0.4471	0.2124	0.0386	0.0236	0.1280	0.0758
<i>Panel B:</i>								
<i>NJR 2SM</i>								
2006	-0.2986	0.2843	-1.1072	0.3538	0.0955	0.1625	0.2551	0.1631
2007	-0.2634	0.2550	-0.9200	0.3861	0.0782	0.1437	0.2044	0.1385
2008	-0.1786	0.2004	-0.6560	0.2825	0.0550	0.1211	0.1354	0.0982
2009	-0.1825	0.1891	-0.6972	0.2979	0.0541	0.0979	0.1424	0.0996
<i>Panel C:</i>								
<i>Epi. 2SM</i>								
2006	-0.2886	0.0978	-1.3311	0.4402	0.0525	0.0336	0.2555	0.1554
2007	-0.2379	0.1026	-1.0714	0.4356	0.0432	0.0301	0.2028	0.1302
2008	-0.1696	0.0745	-0.7830	0.3465	0.0297	0.0230	0.1429	0.0991
2009	-0.1792	0.0781	-0.8334	0.3395	0.0311	0.0230	0.1513	0.0977

A quick glance at Table 4 reveals that both estimated own and cross waiting time elasticities tend to decrease over time in all model specifications. Estimated own elasticities with the standard logit model with fixed effect only control for endogeneity (OLS in Panel A) are in the range between -0.13 and -0.21, and are roughly on line with previous UK hospital demand literature. However, with better control for endogeneity and individual heterogeneity, and taking into account the outside option, estimated waiting time own elasticities in our models range between -0.65 and -1.33, that is, tend to be much bigger.²² Considering cross elasticities, the standard logit model (OLS on Panel A) predicts estimated cross elasticities in the range 0.04-0.06,

²²Substantial effects of IV use on price elasticity estimates is well documented in the empirical IO literature. Riganti *et al.* (2017) using a reduced form model of demand and supply for elective hospital procedures report IV estimated waiting time elasticities six times bigger than those obtained by OLS.

while our models predict estimated cross elasticities in the range 0.14-0.25, suggesting a similar pattern.

TABLE 5. Average MRSA Elasticities

Year	Elasticities				Cross Elasticities			
	OLS		TSLS		OLS		TSLS	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
<i>Panel A:</i>								
<i>Logit</i>								
2006	-0.1216	0.0664	-0.1564	0.0853	0.0354	0.0226	0.0455	0.0290
2007	-0.0887	0.0547	-0.1141	0.0703	0.0247	0.0188	0.0318	0.0241
2008	-0.0893	0.0544	-0.1148	0.0699	0.0250	0.0190	0.0321	0.0244
2009	-0.0880	0.0525	-0.1132	0.0675	0.0252	0.0192	0.0324	0.0247
<i>Panel B:</i>								
<i>NJR 2SM</i>								
2006	-0.2558	0.1197	-0.3101	0.1450	0.0604	0.0412	0.0732	0.0499
2007	-0.1923	0.1167	-0.2332	0.1414	0.0424	0.0322	0.0514	0.0391
2008	-0.1889	0.1077	-0.2290	0.1306	0.0402	0.0318	0.0488	0.0385
2009	-0.1923	0.1118	-0.2332	0.1356	0.0410	0.0329	0.0498	0.0399
<i>Panel C:</i>								
<i>Epi. 2SM</i>								
2006	-0.3068	0.1423	-0.3776	0.1751	0.0598	0.0409	0.0736	0.0503
2007	-0.2252	0.1274	-0.2772	0.1568	0.0425	0.0338	0.0523	0.0416
2008	-0.2232	0.1210	-0.2747	0.1489	0.0408	0.0320	0.0502	0.0394
2009	-0.2277	0.1253	-0.2803	0.1542	0.0421	0.0339	0.0518	0.0417

Table 5 reports the estimated elasticities for the number of MRSA Infections. Infections remains an important complication after hip replacement which can seriously affect the treatment (see for a meta-analysis of the literature Senthil *et al.*, 2011). Table 5 reveals that the estimated own elasticities using the 2SM with IV control for endogeneity are about three times bigger than those obtained with the standard logit model. Again, these differences are substantial also when one considers the cross elasticities, which, on average, double in size.

4.4. The relationship between market structure and quality in the hospital industry. There is a large debate on the relationship between competition and quality in the hospital industry.²³ To address this issue, researchers need a good measure of the competitive pressure faced by each hospital, and the level of quality it produces. The major problem here is that there are many ways to measure quality and competition, often giving conflicting results. In particular, two endogeneity issues are important

²³See Gaynor (2007) and Gaynor *et al.* (2015) for reviews. The relationship between hospital competition and quality in fixed price industries such as the NHS is discussed in Propper *et al.* (2008), Gaynor *et al.* (2013), Bloom *et al.* (2015), Moscelli *et al.* (2016) and Moscelli *et al.* (2016).

for a correct analysis: the endogeneity of quality affecting patient demand; and the endogeneity of hospitals' location choice.

4.4.1. *Measuring Competition.* In general, there are two major methods for measuring competition at the hospital level: choosing a geographical area defining hospital markets -say by the fixed or variable radius method- and calculating either how many hospitals operate in that area or a measure of market concentration such as the Herfindahl-Hirschman Index (HHI); or using patients flows, either actual or predicted, calculating the HHI index at the patient level, and then aggregating it to the hospital level.

Using predicted patient flows to calculate patient level HHI and aggregate it at the hospital level (as proposed by Kessler and McClellan, 2000; and Gowrisankaran and Town, 2003) has the advantage of addressing the quality endogeneity issue, and taking into account population density. To calculate predicted patients' flows, that is, patients' demand not influenced by hospitals' management activities, the analysts finds counterfactual patient choices which determine exogenous captive demand expected by hospitals with average quality and waiting times.

In our application, we exploit estimated structural preference parameters to derive counterfactual patient choices; this may give better estimate of captive demand compared to the standard use of *ad hoc* demand systems, since the counterfactual approach uses consistent estimates of the marginal disutility of distance. Furthermore, since our model takes into explicit account the outside option, it controls for possible location endogeneity, a notoriously difficult problem to handle. In the UK hospital industry, while location of NHS hospitals is very much regulated and generally beyond management choice, there is a host of ISTC and other private providers and accounting for the outside option -which, depending on the area, might have a strong influence on a hospital's competitive environment- gives a better picture of market pressure faced by each hospital.

4.4.2. *Measuring Quality.* As argued in Section 3.5, a host of observable measures are used to measure quality: mortality, infections, readmissions, patient satisfaction etc. It is well documented that there is often little correlation between these measures, and results may depend on the measure used. We take advantage of our structural model which allows to estimate overall unobservable quality q_j for each hospital. Estimated q_j can be seen as an overall index of hospital quality, which includes all observable and unobservable (to the econometrician) factors which affect patient choice (utility) after conditioning on distance and waiting times. Thus, q_j may include things which go beyond observed medical quality but are valued by patients, such as parking facilities, room amenities, staff behaviour, etc. which may be important, especially for routine elective procedures. Unobserved quality q_j is important in the context of the English

hospital market as hospitals are generally very large multi-service organizations treating on average 50,000 patients per year. Therefore, SMR, infection rates, readmissions and other hospital level indicators are likely to be driven by the standards in emergency services and might fail to capture important attributes that are relevant for patients' choice of routine procedures.

4.4.3. Results. We calculate quality and competition using the standard logit and the epidemiological 2SM using data from 2009. We first calculate overall quality q_j under the logit model, and use the HHI with actual patient flows as the (lack of) competition measure. We find a strong *positive* correlation of 0.602 between market concentration and quality. However, since in actual patient flows quality is endogenous, the correlation is actually much inflated. Correcting for quality endogeneity using predicted 'captive' patient flows lowers the correlation by more than two third to 0.188.

The standard logit model –even after quality endogeneity correction using captive demand– has some remaining problems, namely the lack of consideration of the outside option, and the failure of proper control for waiting time endogeneity, which may bias the estimate of overall quality q_j , since a portion of the disutility of waiting time is wrongly attributed to q_j . Our model with the outside option and proper waiting time endogeneity control finds a strong *negative* correlation -0.453 between market concentration (lack of competition) and quality.²⁴

4.5. Hospital closure simulation. A recurrent major policy issue in the UK health-care market has been the rearrangement of the NHS hospital industry by mergers and closures.²⁵ When studying the effects of closing a given hospital, it is key to evaluate the changes in hospital demand after the closure. Using a structural model it is possible to compute the counterfactual size of market share of each hospital to guide policies assessing closure effects. Using standard logit models to calculate these counterfactuals may paint a wrong picture for at least two main reasons: the presence of an outside option implies that the estimated increases in market share of other hospitals tend to be overestimated, and the logit I.I.A. structure tends to give wrong estimates of substitution effects.

Consider the following simple example showing the relevance of omitting the outside option. Suppose there are 800 patients and 3 NHS hospitals, a , b and c , with mean utilities $U_a = \log(1/4)$, $U_b = \log(1/2)$, $U_c = \log(3/4)$. Let the utility from the outside option be 0. Under logit errors, market shares are 0.1, 0.2, 0.3 for the NHS hospitals,

²⁴A negative correlation between market concentration and quality seems quite natural in an industry where demand has a very strong geographical component, the geographical distribution of NHS hospitals is heavily regulated, and hospitals have limited capacity flexibilities.

²⁵See Dafny (2009), Capps *et al.* (2003), Gaynor *et al.* (2012 and Gaynor *et al.* 2015), on hospital consolidation, mergers and closures.

and 0.4 for the outside option. The number of patients treated in the three NHS hospitals before closure is 80, 160, 240.

Suppose now hospital b closes. Calculating the new market shares, the demand for hospital a will increase from 80 to 100, and the demand for hospital c will increase from 240 to 300. However, if we estimated the logit model conditional on choosing a NHS hospital –that is, ignoring the outside option– the demand for hospitals a and c after b 's closure is 120 and 360 respectively. Thus, omission of the outside option implies a much higher increase in estimated demand after closure. If preferences are not homogeneous, using the standard logit model implies a further source of bias due to incorrect estimation of the substitution patterns between hospital b and a and c .

4.5.1. An Application. A major UK newspaper recently revealed that to cut expenses the NHS is considering to downgrade some hospitals. These hospitals may lose their A&E units and other acute services. This is in practice equivalent to a closure since the services would not be anymore delivered by that specific hospital. We draw inspiration from this news to illustrate the possible application of our model to understand how hospital market shares change when an hospital is closed (or simply downgraded for a specific procedure).²⁶

We consider the downgrade (closure) of a teaching hospital located in a highly competitive market area, where both public and private providers act extensively. We denote the closing hospital as hospital H, depicted in Figure 1 in the center of the picture. In 2009, the number of over-65 patients treated for elective hip replacement in hospital H was 137. We use the standard logit and our Epidemiological 2SM to predict the increase in patients demand in other hospitals following the closure. Table 6 shows the estimated increase in patients' demand in the 5 major competitors, denoted A,B,C,D,E, located around the closing hospital in Figure 1. Interestingly both models predict the same hospital having the greatest increase in demand after closure, namely hospital A. Hospital A is in the 3 km catchment area of the closed hospital, hospitals B and C are within 5 km of H, while D and E are within 10 and 15 km of H respectively.

The pattern of predictions across the two models is sharply different: the standard logit model predicts a much greater increase in demand in all hospitals, since it does not take into account the rather large number of patients choosing the outside option. On the other hand the two models imply quite different substitution patterns. In fact, as it can be seen in the table 6, the increase in the demand in hospital A is almost double in the 2SM, the outside option notwithstanding.

²⁶Our example is only hypothetical and it is intended to show the potential benefit of using our approach to formulate policy relevant counterfactuals.

TABLE 6. Hospital Closure Simulation: Estimated number of patients

Hospital	Before	Logit			Epi. 2SM		
		After	Increase	% Increase	After	Increase	% Increase
A	98	134.16	36.16	.37	169.67	71.67	.73
B	35	48.91	13.91	.40	45.66	10.66	.30
C	78	98.52	20.52	.26	82.42	4.42	.06
D	86	93.70	7.70	.09	88.01	2.01	.02
E	603	617.11	14.11	.02	606.90	3.90	.01

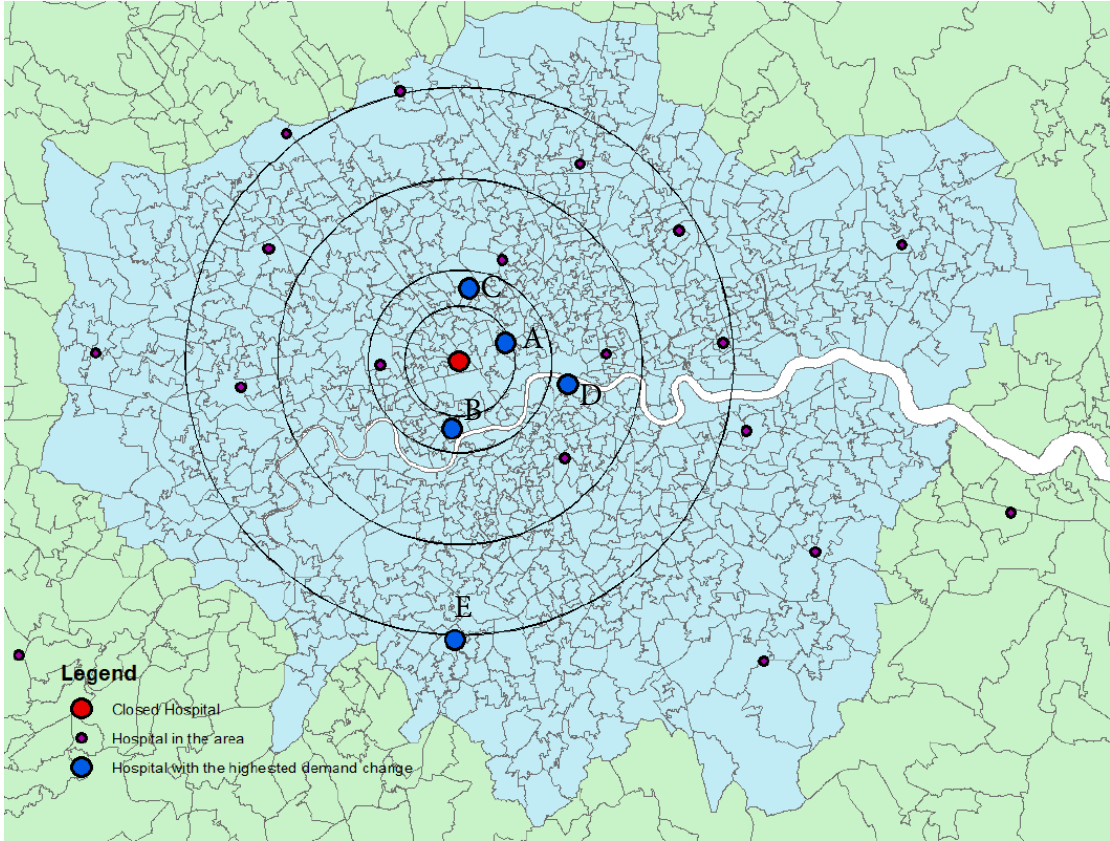


FIGURE 1. Geographic Distribution of Hospitals in Hypothetical Closure

5. CONCLUSIONS

Empirical IO literature has shown the importance of estimating choices in differentiated product markets allowing for individual taste heterogeneity, addressing unobserved quality, and including an ‘outside option’ in individuals’ choice set. We considered a structural model of hospital choice for elective hip replacement in the English NHS in the period 2006-2009. We showed how to properly address these issues, overcoming available data constraints, implementing a two-sample model by a novel use of administrative data on English small areas to supplement the standard HES dataset, which is universally used in NHS demand estimation. Our strategy was to couple the HES

sample, with a synthetic sample using LSOA's data to mimic the over 65 English population seeking hip replacement surgery, and matching observed attributes in the HES sample with those in the LSOA.

Our structural model starts with a standard utility function for NHS hospitals depending on distance, waiting time and the quality of hospital care with heterogeneity in patients' preferences. Our model adds the possibility that patient may opt-out from public funded NHS hospital by choosing an outside option, containing all possible other alternatives that a patient needing a hip replacement procedure may choose. To empirically estimate this model we also use a TSLS approach to take into account the endogeneity caused by unobservable quality.

Our approach gives a clearer understanding on how the hospital market reacts to policy changes. In particular, we find that waiting time own elasticities using a standard logit model are generally quite low (-0.16 on average, comparable with estimates for elective procedures in NHS of Sivey (2012) and Moscelli *et al.* (2016) of about -0.12 and -0.04), while estimated waiting time elasticities in our models are substantially bigger, on average equal to -0.92. Similarly, estimated waiting time cross-elasticities are much bigger in our richer models (0.18 on average) compared to 0.046 on average in the standard model. Own and cross quality (MRSA) elasticities are on average twice larger in our model.

We also address the much debated issue of the relationship between market structure and quality of care in the hospital industry. We find that this relationship is starkly different when addressing the endogeneity of quality and hospitals' location choices. Estimates of the correlation between overall quality and market concentration range between a *positive* correlation of 0.602 (using using actual patient flows without taking into account the outside option and estimating quality without endogeneity correction), and a *negative* correlation -0.453 when using our model.

Finally, in a counterfactual simulation of a hospital closure, we also find rather different results. Although both the logit model and ours predict the same hospital having the greatest increase in demand after closure, the logit model predicts a much greater total increase in demand in all the other hospitals due to the lack of consideration of the outside option. In other words, the logit model predicts a greater total inflow of patients to other NHS hospital, and at the same time predicts different patterns of hospital demand cross-substitution.

REFERENCES

- Beckert, W., M. Christensen, and K. Collyer (2012). Choice of NHS-funded hospital services in England. *The Economic Journal* 122(560), 400–417.
- Beckert, W. and E. Kelly (2017). Divided by choice? private providers, patient choice and hospital sorting in the English National Health Service. Working Papers W17/15, Institute for Fiscal Studies.
- Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63(4), 841–890.
- Berry, S., J. Levinsohn, and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy* 112(1), 68–105.
- Beukers, P. D. C., R. G. M. Kemp, and M. Varkevisser (2014, Dec). Patient hospital choice for hip replacement: empirical evidence from the Netherlands. *The European Journal of Health Economics* 15(9), 927–936.
- Bloom, N., Z. Cooper, M. Gaynor, S. Gibbons, S. Jones, A. McGuire, R. Moreno-Serra, C. Propper, J. Van Reenen, and S. Seiler (2011). In defence of our research on competition in England’s National Health Service. *The Lancet* 378(9809), 2064–2065.
- Bloom, N., C. Propper, S. Seiler, and J. Van Reenen (2015). The impact of competition on management quality: Evidence from public hospitals. *The Review of Economic Studies* 82(2), 457–489.
- Briggs, D., D. Fecht, and K. De Hoogh (2007). Census data issues for epidemiology and health risk assessment: experiences from the Small Area Health Statistics Unit. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170(2), 355–378.
- Capps, C., D. Dranove, and M. Satterthwaite (2003). Competition and market power in option demand markets. *The RAND Journal of Economics* 34(4), 737–763.
- Dafny, L. (2009). Estimation and identification of merger effects: An application to hospital mergers. *The Journal of Law & Economics* 52(3), 523–550.
- Einav, L. and J. Levin (2010). Empirical industrial organization: A progress report. *Journal of Economic Perspectives* 24(2), 145–62.
- Gaynor, M. (2007). Competition and quality in health care markets. *Foundations and Trends in Microeconomics* 2(6), 441–508.
- Gaynor, M., K. Ho, and R. J. Town (2015). The industrial organization of health-care markets. *Journal of Economic Literature* 53(2), 235–84.
- Gaynor, M., M. Laudicella, and C. Propper (2012). Can governments do it better? merger mania and hospital outcomes in the English NHS. *Journal of Health Economics* 31(3), 528–543.

- Gaynor, M., R. Moreno-Serra, and C. Propper (2013). Death by market power: Reform, competition, and patient outcomes in the National Health Service. *American Economic Journal: Economic Policy* 5(4), 134–66.
- Gaynor, M., C. Propper, and S. Seiler (2016). Free to choose? Reform, choice, and consideration sets in the English National Health Service. *American Economic Review* 106(11), 3521–57.
- Goolsbee, A. and A. Petrin (2004). The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica* 72(2), 351–381.
- Gowrisankaran, G. and R. J. Town (2003). Competition, payers, and hospital quality. *Health Services Research* 38(61), 1403–1422.
- Gravelle, H., R. Santos, L. Siciliani, and R. Goudie (2012). Hospital quality competition under fixed prices. Working Papers 80, Centre for Health Economics, University of York.
- Ho, K. (2006). The welfare effects of restricted hospital choice in the US medical care market. *Journal of Applied Econometrics* 21(7), 1039–1079.
- Ho, K. (2009). Insurer-provider networks in the medical care market. *The American Economic Review* 99(1), 393–430.
- Hospital Episode Statistics Analysis, Health and Social Care Information Centre (2015). Hospital episode statistics admitted patient care, England 2014–15. Technical report.
- Imbens, G. W. and T. Lancaster (1994). Combining micro and macro data in microeconomic models. *The Review of Economic Studies* 61(4), 655–680.
- Kelly, E. and G. Stoye (2015). New joints: private providers and rising demand in the English National Health Service. Working Papers W16/15, Institute for Fiscal Studies.
- Kessler, D. P. and M. B. McClellan (2000). Is hospital competition socially wasteful. *The Quarterly Journal of Economics* 115(2), 577–615.
- Laudicella, M., P. Li Donni, and P. C. Smith (2013). Hospital readmission rates: signal of failure or success? *Journal of health economics* 32(5), 909–921.
- Monitor (2013). A guide to the Market Forces Factor. Technical Report IRG 31/13, NHS England Publications Gateway Reference 00883.
- Moscelli, G., H. Gravelle, and L. Siciliani (2016). Market structure, patient choice and hospital quality for elective patients. Working Papers 139, Centre for Health Economics, University of York.
- Moscelli, G., L. Siciliani, N. Gutacker, and H. Gravelle (2016). Location, quality and choice of hospital: Evidence from England 2002–2013. *Regional Science and Urban Economics* 60(Supplement C), 112 – 124.

- Mota, R. E. M., R. Tarricone, O. Ciani, J. F. Bridges, and M. Drummond (2012). Determinants of demand for total hip and knee arthroplasty: A systematic literature review. *BMC health services research* 12(1), 225.
- National Joint Registry (2010). National Joint Registry for England and Wales: Annual report. Technical report, The National Joint Registry Centre, Hemel Hempstead.
- Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy* 9(4), 513–548.
- Nevo, A. (2011). Empirical models of consumer behavior. *Annual Review of Economics* 3(1), 51–75.
- Noble, M., G. Wright, G. Smith, and C. Dibben (2006). Measuring multiple deprivation at the small-area level. *Environment and planning A* 38(1), 169–185.
- Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* 57(5), 1027–1057.
- Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy* 110(4), 705–729.
- Pollock, A., A. Majeed, A. Macfarlane, I. Greener, G. Kirkwood, H. Mellett, S. Godden, S. Boyle, C. Morelli, and P. Brhlikova (2011). In defence of our research on competition in England's National Health Service—authors' reply. *The Lancet* 378(9809), 2065–2066.
- Propper, C., S. Burgess, and D. Gossage (2008). Competition and quality: Evidence from the NHS internal market 1991–9. *The Economic Journal* 118(525), 138–170.
- Riganti, A., L. Siciliani, and C. V. Fiorio (2017). The effect of waiting times on demand and supply for elective surgery: Evidence from Italy. *Health Economics* 26(S2), 92–105.
- Senthi, S., J. T. Munro, and R. P. Pitto (2011). Infection in total hip replacement: Meta-analysis. *International orthopaedics* 35(2), 253–260.
- Sivey, P. (2012). The effect of waiting time and distance on hospital choice for English cataract patients. *Health Economics* 21(4), 444–456.
- Train, K. E. and C. Winston (2007). Vehicle choice behavior and the declining market share of U.S. automakers. *International Economic Review* 48(4), 1469–1496.