Institute for Fiscal Studies

Divided by choice? Private providers, patient choice and hospital sorting in the English National Health Service

IFS Working Paper W17/15

Walter Beckert

Elaine Kelly

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

Divided by Choice? Private Providers, Patient Choice and Hospital Sorting in the English National Health Service*

Walter Beckert† Elaine Kelly‡

August 9, 2017

Abstract

A common reform used to increase consumer choice and competition in public services has been to allow private providers to compete with public incumbents. However, there remains a concern that not all consumers are able to benefit equally from wider choice. We consider the case of publicly funded elective surgery in England, where reforms in the 2000s enabled privately owned hospitals to enter the market. We show that, post-reform, poor and ethnic minority patients were much less likely to choose private hospitals; and that dominant drivers of sorting between public and private providers are health based criteria for treatment by private providers and the geographic distribution of hospitals. Counterfactual simulations suggest differences in health explain 18% of the difference in the use of private providers between rich and poor patients, while the geographic distribution of hospitals explains 61% once other sorting mechanisms - ethnicity, patient preferences, physician referral patterns - are accounted for. Although much of the observed sorting does not appear to be the result of market frictions, limited variation in payments made to hospitals according to patient health means that sorting is estimated to cost public hospitals in excess of £426,426 ($625,000) per year.

JEL classification: I11, I18, L1, L44, D12

Keywords: Patient choice, demand for healthcare, healthcare reform, inequality

*We thank the Health and Social Care Information Centre for providing access to the Hospital Episode Statistics under data sharing agreement CON-205762-B8S7B. This paper has been screened to ensure no confidential information is revealed. We thank the ESRC through The Centre for the Microeconomic Analysis of Public Policy (CPP) (ES/H021221/1) and Kelly's Future Leaders grant (ES/K009060/1) for financial support. We are grateful to James Banks, Jennifer Dixon, Pierre Dubois, Rachel Griffith, Jon Gruber, Martin O'Connell, Carol Propper, Adam Roberts, Kate Smith, Ron Smith, and the participants of the Winter 2016 HESG in Manchester, RES and EEA 2016 Conferences. All errors are our own.

†Birkbeck College, University of London (w.beckert@bbk.ac.uk)

‡Institute for Fiscal Studies (e.kelly@ifs.org.uk)

1 Introduction

Recent reforms in many countries have sought to increase the role of consumer choice in public services such as education and health care. At a time when government finances are severely constrained, choice is viewed as a mechanism for driving competition between providers and thereby, in a system with fixed prices, delivering improvements in quality and efficiency. One type of reform employed to promote choice has been to increase the options available to consumers by allowing entry from private sector or for-profit providers (Besley and Ghatak, 2003; Blöchliger, 2008; Hoxby, 2003). Examples include Charter schools in the US and Sweden (Asth *et al.*, 2013; Böhlmark and Lindahl, 2015; Ladd, 2002); and publicly funded health care systems, like the English National Health Service (NHS) studied in this paper[1].

Policies to increase choice have often proved controversial. Debates typically surround two related sets of issues. The first and most widely studied is the impact of choice on competition, and the effect of that competition on hospital level outcomes (Cooper *et al.*, 2011; Deming *et al.*, 2014; Gaynor *et al.*, 2016). The second is the distributional impact of choice reforms, and in particular whether disadvantaged consumers are less able to engage in choice (Altonji *et al.*, 2015; Urquiola, 2005). It is this second set of concerns that we focus upon on this paper. Differential sorting across provider types may be of concern for at least three reasons. First, if disadvantaged groups are less likely to use new providers due to market imperfections such as incomplete information, sorting may imply welfare losses. Second, where markets for different types of providers become segmented, the degree of competition between providers may be muted. And third, when public reimbursements for services are largely capitated, then how different types of consumers sort across providers has important implications for the distribution of public funds, provider profits and competition.

The case we consider is how and why patients sort across hospital types following reforms to the English National Health Service (NHS) in the 2000s, which increased choice by allowing privately-owned hospitals, or Independent Sector Providers (ISPs), to enter the market for publicly funded health care. These privately-owned hospitals typically have shorter waiting times, higher patient satisfaction and arguably higher clinical quality[2].

[1] In the US system, except for the elderly and those on social benefits, provision is provided by private managed care organizations, comprising health insurance and health care delivery. In that system, the "public option", blocked by Congress in 2010, was intended as a constraint on the private market place.

[2] See, for example, NHS Partners Network (2015), Care Quality Commission (2012) and Browne *et al.* (2008). The Care Quality Commission (CQC) reports that 82 percent of private hospitals met all essentials standards (including care and welfare, cleanliness and infection control, medicines management, monitoring service quality, complaints) that were checked in inspections, compared to 77 percent of public hospitals. Browne *et al.* (2008) that found, once adjustments are made for pre-operative characteristics, patients undergoing hip replacement (and cataract surgery) at ISTCs achieved a slightly greater improvement in

Our analysis uses National Health Service administrative records on patients that had elective hip replacements in 2012/13. Hip replacements are used as the procedure is conducted in large volumes and ISPs have gained a substantial presence in the market, treating 20% of patients and accounting for 38% of hospitals that treated publicly funded patients in 2012/13. Our analysis focuses on the role of three different sorting mechanisms: (i) the geographic distribution of providers, (ii) differences in patients' underlying health, and (iii) the role of the patient's referring physician. Understanding the relative importance of these mechanisms is important both for assessing the welfare implications of sorting and in designing potential policy responses.

We start by identifying and quantifying differences in the use of privately owned providers by patient characteristics. Descriptive evidence is then provided on the extent to which each of distance, patient health and referring physician might explain these differences. For patients in our sample, those living in the richest fifth of local areas are almost twice as likely to choose an ISP (23%) as those living in the poorest fifth of areas (13%), while white patients are more than twice as likely to choose an ISP (20%) as ethnic minority patients (8%). However, patients from ethnic minorities and deprived areas live on average further from ISPs, and are in worse underlying health, both of which are associated with lower ISP use.

We next estimate a mixed multinomial logit (MMNL) model of hospital choice, where patients are able to choose from a set of both public NHS and privately owned hospitals. This allows us to take into account how characteristics of different patient types interact with attributes of the hospitals in patients' choice sets. The set of hospitals that patients are able to choose between is defined in two ways. First, in a standard approach, which defines choice sets based on distance and which we refer to as "distance choice sets"; and second, by defining the choice sets by the prior referral patterns of the patient's physician - or general practitioner (GP) -, which we refer to as "GP choice sets". This aims to capture the important role GPs play as both the "gatekeeper" to elective hospital care, and the main source of advice to patients (Boyle, 2013; Dixon *et al.*, 2010). Consistent with the findings in the hospital choice literature (Beckert *et al.*, 2012; Capps *et al.*, 2003; Gaynor *et al.*, 2016; Ho, 2006), the model estimates show that distance, waiting times and quality emerge as significant determinants of choice. Alongside the observable heterogeneity, we also identify significant unobserved heterogeneity across patients with regard to their sensitivity to distance and their preferences for ISPs, relative to NHS hospitals. These results are robust with respect to the two choice sets definitions we consider.

The final part of the paper uses parameter estimates from the model to conduct coun-

functional status and quality of life than those treated in NHS facilities.

terfactual simulations to assess the relative contributions of the mechanisms we highlight. We find that differences in underlying health, the geographical location of privately-owned hospitals, and restrictions in choices offered to patients by GPs all have a substantial role to play in the observed patterns of sorting. Our simulations suggest that differences in health explain 18% of the gradient in the use of privately owned providers between rich and poor patients and 56% of the difference between white and ethnic minority patients. After removing differences in health, as well as other possible sorting mechanisms such as ethnicity, patient preferences and GP referral patterns, what remains is the geographical distribution of privately owned providers that explains 61% of the original income gradient and 44% of the original differences in the use of private providers between white and ethnic minority patients. The referral patterns of GPs explain about a fifth of the income gradient but none of the differences by ethnicity. Very little of the sorting we observe is attributable to differences in preferences once other patient characteristics have been taken into account.

In terms of welfare, sorting by health and geography appears consistent with patient needs and preferences rather than market imperfections. The impact of restrictions placed on choices by GPs is less clear. These restrictions may reflect differences in patient characteristics unobserved to the econometrician, but may also reflect imperfect information, GP preferences or administrative frictions. We find evidence administrative distortions to GP referrals unrelated to patient health do affect referral patterns, but play only a small role in the overall pattern of sorting that is observed.

Our last counterfactual focuses on the costs to NHS hospitals of sorting on the basis of health. As in healthcare systems in most other developed countries, the majority of hospital care in England is reimbursed through a Diagnosis-related grouping (DRG) system (Langenbrunner *et al.*, 2009). These systems typically have only limited variation in payments by patient characteristics. Here we show that ISPs receive the same payment for most patients as NHS hospitals, but treat patients with better underlying health who are less costly to treat, based on predicted length of stay. We estimate that this raises public hospitals' cost by at least £426,426 ($625,000) in 2012/13.

Our work contributes to several existing literatures. Our principal contribution is to the hospital choice literature (Beckert *et al.*, 2012; Capps *et al.*, 2003; Gaynor *et al.*, 2012; Ho, 2006; Kessler and McClellan, 2000). Here we build on existing work by focusing on how patients sort across hospitals, and on the distribution of welfare gains, rather than aggregate changes. We also incorporate the introduction of private providers, which are not included in other models of hospital choice in England (Gaynor *et al.*, 2016)[3]. Our results provide three main insights. First, the primacy of distance in a patient's choice of hospital entails that

[3]Gaynor *et al.* (2016) consider CABG surgery and this is not a market where ISPs operate.

the introduction of new providers can have distributional consequences. Second, there are restrictions on choice that arise from frictions in the market and therefore distort consumer welfare and limit competition[4]. Finally, we provide an estimate of the financial cost of sorting when there is limited variation in payments to hospitals across payment types, which has relevance to all healthcare systems where payments are based on DRGs.

Beyond healthcare, we contribute to the literature that considers the relationship between choice and sorting in other services, such as school choice (Altonji *et al.*, 2015; Böhlmark *et al.*, 2016; Burgess *et al.*, 2015; Edmark *et al.*, 2014; Hastings *et al.*, 2010; Hastings and Weinstein, 2008; Urquiola, 2005). As in health care, reforms to increase school choice have included offering parents wider formal choice opportunities, providing information on school quality to aid choice (Hastings and Weinstein, 2008), and introducing charter schools and school vouchers in countries such as the US, Norway and Sweden (Ladd, 2002). Our results support the assertion that the geographic distribution and entrance criteria of new schools are likely to have important implications for how students sort across schools. Where the costs of educating pupils varies, any sorting of "cheaper" pupils to new providers could raise the average cost of existing schools.

Finally, we add to the extensive literature on socioeconomic inequalities in health care utilisation. In general, this literature finds pro-poor inequalities in the use of primary care and community health services and pro-rich inequalities in the use of hospital care (Cookson *et al.*, 2012; Doorslaer *et al.*, 2004; Morris *et al.*, 2005; O'Donnell and Propper, 1991). However, the extent and the direction of these inequalities typically vary by country, year and condition and are hard to generalize[5]. The literature on variation in the *quality* and *types* of care received by different types of patients is smaller, but typically shows that treatment is on average less intensive and of a lower quality for disadvantaged socioeconomic and racial minority groups (Fiscella *et al.*, 2000; Moscelli *et al.*, 2015).

The rest of the paper is organized as follows. Section 2 provides some background on the NHS and the relevant policy reforms. Section 3 describes the data. Section 4 outlines the model and empirical strategy. Section 5 presents the results and a discussion of their

[4]While our study focusses on the choices made by patients, given the institutional, socio-demographic and choice protocol setting they find themselves in, we note that an emerging literature is concerned with structurally modelling choice protocols in which choice (or consideration) sets are restricted or heterogeneous across decision makers, often in ways that are only partially observed by the econometrician. In the area of choices in health care, see for example Beckert and Collyer (2016) and Gaynor *et al.* (2016). Abaluck and Adams (2017) provide conditions under which utility and consideration set probabilities can be separately identified without auxiliary data.

[5]For example, Cookson *et al.* (2012) find no change in the inequality in the provision of hip replacements between 2001 and 2008 in England, but Kelly and Stoye (2016) find uneven growth in the number of hip replacements by local area deprivation from 2002 to 2011, largely explained by changes between 2008 and 2011.

robustness. Section 6 provides counterfactual simulations and cost calculations and Section 7 concludes.

2 Background

2.1 NHS Policy Reforms and Patient Choice

The majority of health care in England is provided through the taxpayer funded National Health Service (NHS), free at the point of use. In this paper, we study the market for NHS funded elective hospital care.

On the demand side, patients access specialist hospital care via a referral from their primary care doctor or General Practitioner (GP). In addition to acting as an agent for their patients, GPs also operate as a "gatekeeper" to hospital-based care to manage demand. Since the "patient choice" reforms of 2006 and 2008, GPs are required to offer patients a choice of hospital when making a referral[6]. GPs may influence where patients are treated either by pre-selecting the options that are presented to patients to choose from, or by providing advice to help patients choose between the options presented (Dixon *et al.*, 2010). The role of GPs in determining how patients sort across hospitals is returned to in more detail in Sections 3 and 4.

On the supply side, NHS funded hospital care has historically been delivered by state owned and run NHS Acute Trusts, or hospitals[7]. This paper focuses on a set of reforms introduced alongside the patient choice reforms that further extended choice by increasing the number of providers or hospitals available to NHS funded patients. The NHS had purchased small volumes of care from the private sector on an ad hoc basis for many years, but reforms introduced between 2003 and 2008 formalized and greatly increased the access of private providers to markets for NHS funded care.[8]

The reforms created two types of ISP. The first type, Independent Sector Treatment Centres (ISTCs), were privately owned facilities under contract with the NHS to provide diagnostic tests and elective procedures to NHS funded patients free at the point of use. The introduction of ISTCs, from 2003 onwards, reflected the policy focus of the early 2000s on reducing very long waiting times for elective care through strict waiting time targets. ISTC

[6]These reforms were motivated by both, the belief that patients valued the choice over their care, and evidence that health care competition when prices were fixed could improve quality (Gaynor, 2006). A series of work has found that this reform led to improvements in quality (Cooper *et al.*, 2011; Gaynor *et al.*, 2012).

[7]A NHS Acute Trust may be comprised of a single hospital or multiple hospitals within the same geographic area.

[8]For hip replacements there also exists a small private pay sector, which accounted for a fifth of hip replacements in 2002 (Arora *et al.*, 2013); it is excluded from all analyses in this paper.

contracts were therefore awarded in part based on local capacity constraints and waiting times at local NHS hospitals. The objective was to use ISTCs to treat routine patients, allowing NHS trusts to focus on emergencies and patients with more complex needs. As ISTCs do not have intensive care facilities, there were restrictions placed on which patients were eligible for ISTC treatment. This meant patients treated by ISTCs were on average younger and richer than those treated by NHS providers (Bardsley and Dixon, 2011; Chard *et al.*, 2011).

The second type of ISP were existing conventional private hospitals, which had previously received almost all their income from private pay patients. Reforms in 2008 allowed these private hospitals to compete for NHS-funded patients, and treat private pay and NHS funded patients alongside one another. This second expansion in the role of private provision within the NHS reflected a greater policy focus on competition and choice post 2006, with ISPs now viewed as a mechanism for increasing consumer choice and thereby putting pressure on NHS hospitals to increase quality (Naylor and Gregory, 2009). Unlike ISTCs, the location of these new providers was not a policy response to local capacity constraints. However, all types of private provider had the same restrictions on eligibility based on underlying health, as most private hospitals are not equipped to treat patients that might need intensive care or have other complex needs.

During the period we study, both GPs and hospitals received NHS payments through 152 "Primary Care Trusts" (PCTs). These PCTs were publicly funded bodies who had the responsibility of paying for the healthcare of all patients living within their designated geographic area[9]. Payments to GPs were based on a capitation fee, plus a payment-for-performance supplement, and were not dependent on referrals. By 2012/13 all hospitals, whether NHS or privately owned, received payments based on activity, where payments were based on a diagnosis-related grouping system. These payments are set nationally and vary very little across providers[10]. However, the earliest ISTCs ("wave 1") initially received block contracts for a certain number of procedures and were paid even if these procedures did not take place. These block contracts had all expired by 2012/13, but during their operation GPs were encouraged by their PCTs to refer to these providers.

Figure 1 shows how ISPs spread across England between 2006/7 and 2012/13. In 2006/7 there were just 9 ISPs conducting at least 20 NHS funded procedures. By 2012/12 this had

[9]These organisations were established in 2002 to deliver a purchaser-provider split necessary to sustain a market for healthcare.

[10]These are known as Diagnosis Related Groups (DRGs) in the US and Healthcare Resource Groups (HRGs) in England. Prices or Tariffs are then set at a national level based on the average cost of providing the associated care. Small adjustments are made to the payments received, based on length of stay and local costs of living.

risen to over 119, spread from Newcastle in the North East to Cornwall in the South West. The number of NHS hospitals remained roughly stable at 200 throughout the period. The reforms therefore increased the hospitals available to patients by more than half, and by 2012/13 a fifth of NHS funded hip replacements were conducted by ISPs.

2.2 Mechanisms of Patient Sorting by Provider Type

The structure of the reforms points towards three mechanisms that might generate differences in the characteristics of patients by provider type, each of which suggests different policy responses.

First, as patients must meet certain health criteria in order to be eligible for ISP treatment, some differences in use by underlying health are inevitable[11]. As ill-health is correlated with other patient characteristics, such as poverty and advancing age, we may expect these groups to be less likely to use ISPs. Insofar as differences in underlying health reflect need, this type of sorting may be an efficient allocation of patients across hospitals. The same type of sorting might be expected in selective secondary or tertiary education, where children from richer families typically have higher entry test scores.

Second, the geographic distribution of ISPs is non-random and is likely to result in differential access to ISPs. In particular, after 2008 most new ISPs were existing private hospitals. These were typically located in richer areas, close to the private-pay patients they primarily serve. Again, given that patients always show a preference for shorter distances, any resulting sorting may be efficient, taking the geographic distribution of ISPs as given. Whether the geographic distribution is itself efficient is a separate question.

Finally, there may be market imperfections that affect where patients are treated. One potential source of market imperfection are restrictions in the choices offered to patients by their referring GP. As ISPs were new and introduced very quickly, GPs may lack information about the additional providers, at least in the short run. The structure of the first wave of ISTC contracts also provided an incentive for PCTs to encourage GPs to refer to ISTCs, to avoid paying for procedures that did not take place. These types of frictions are at least potentially inefficient, both in terms of restricting access of patients to ISPs and limiting competition between providers. We will return to the issue of the options presented to patients in Section 4.2.

In addition to concerns about potential welfare losses resulting from market frictions, there are at least two further reasons why policy makers may be concerned about the sorting of patients across providers.

[11] These regulations were laid down by the government. Whether ISPs then imposed additional eligibility requirements that did amount to 'cherry picking' or 'cream skimming' remains open to debate.

First, even if ISP use were based on complete information and absent administrative constraints, policy makers may be concerned if choice leads to too much segmentation, or indeed segregation, in public service utilization, given it is paid for by, and designed to serve, all. Moreover, this segmentation may limit the extent of competition between NHS hospitals and ISP, reducing the pressure on NHS hospitals to improve quality.

Second, the characteristics of patients carry implications for hospital costs. In particular, the cost of treating patients with worse underlying health is typically higher, but there is limited variation in payments that hospitals receive. For hip replacements, as with many other procedures, there exist different clinical groupings and a higher payment for treating patients that have complications and comorbidities. However, relatively few patients qualify for the higher payment.[12] Moreover, the costs of treating patients are likely to vary more continuously with underlying health. Low cost patients moving from NHS hospitals to ISPs and thereby leaving NHS hospitals to treat high cost patients may be regarded as an adverse selection issue. This will generate negative spillovers for NHS hospitals, as it limits their ability to cross-subsidize across patients, leading to the average cost of patients treated by the NHS to rise while ISPs receive a surplus.

All these concerns depend upon the extent and type of sorting that takes places. Existing evidence from early wave 1 ISTCs points towards ISPs treating younger, healthier patients (Bardsley and Dixon, 2011; Chard *et al.*, 2011; Cooper *et al.*, 2015).

The next section details our data and describes the patterns of sorting in 2012/13, by which time almost all ISPs had been introduced.

3 Data

We use data on NHS-funded elective hip replacements. The data come from the NHS inpatient Hospital Episode Statistics (HES). They provide an administrative record of all NHS-funded inpatient treatments in England, including treatments provided by both NHS hospitals and ISPs. Each patient record contains information on where the patient was admitted, the date of admission and discharge, up to 20 ICD-10 diagnoses, and information on any procedures that took place. For each patient record, HES data also identify the referring GP practice, albeit not the individual GP. We extract hip replacements using the relevant orthopaedic procedure codes, and obtain a sample of 68,769 patients.[13]. For

[12] In our sample, 75% of patients fall under HRG HB12C "Major Hip Procedures for non Trauma Category 1 without CC".

[13] Hip replacements include those operations with Office of Population Censuses and Surveys (OPCS) Classification of Interventions and Procedures Codes (4th Edition) beginning W37, W38, W39, W93, W94 and W95. Each operation code defines a different type of hip replacement. For a full list of OPCS codes see

modelling reasons discussed below, we restrict the sample to those treated by hospitals that conduct at least 20 hip replacements in a year, and those choosing a hospital from among the nearest 10 providers. After improving these restrictins, we are left with a sample of 62,695 patients.

3.1 Patient Characteristics

Table 3 shows mean demographic and health characteristics by chosen provider type. The average age of patients treated by both NHS hospitals and ISPs is 68. The share of ethnic minority patients, which has not been examined by existing studies, is much lower among ISP patients (1.3%) than among NHS patients (3.9%)[14]. This is consistent with qualitative evidence on how patient choice operated during this period, where GPs voiced concerns that language barriers may limit the ability of minority ethnic populations from exercising choice (Dixon *et al.*, 2010).

Two sets of measures are used to capture underlying health of the patient. First, we consider the Charlson Index of comorbidities[15] We group patients into three categories: a score of zero for no comorbidities; a score of 1 for "mild comorbidities"; and a score of more than one for "severe comorbidities".

Second, we extract all prior admissions for patients in our estimation sample, and create indicators for whether the patient had at least one (NHS funded) elective or emergency admission in the three years (1095 days) prior to the hip replacement admission, for any cause. All our measures confirm that ISP patients are on average less complex and have better underlying health than NHS hospital patients[16]. It is however important to note that the market is not completely segmented by underlying health: a substantial fraction of ISP patients do have comorbidities or prior admissions.[17].

here: http://www.surginet.org.uk/informatics/opcs.php.

[14]These shares are much lower than the share of people of an ethnic minority patients in the population, due to the age structure of the ethnic minority population in England

[15]It is calculated using the comorbidities recorded at the point of the hip replacement admission. The Charlson Index predicts ten-year mortality using 22 comorbidity conditions. Each condition is scored a 1, 2, 3 or 6, depending on the severity of the condition, and is calculated on the basis of all diagnoses recorded in hip replacement admission. See Sundararajan *et al.* (2004) for more details on the Charlson Index.

[16]Comparing these measures with the reported underlying health recorded for the 60% of the sample that responded the Patient Recorded Outcome Measures (PROMs) survey illustrates that the health measures we use pick up different elements of ill health. Of those that report ever having cancer in PROMs, 79% have had an elective admission to hospital over the previous 3 years, compared to 53% for all other patients, emergency admissions were 10 percentage points higher (29% verses 19%), and cancer patients were twice as likely to have a Charlson index score of 2 or more (15% verses 7%). By contrast, for those reporting high blood pressure, the shares with prior emergency and elective admissions are both only 2 percentage points higher than the rest of the sample, whereas the share of those with a Charlson Index of 2 or more is 6 percentage points higher.

[17]This is also true when we use the more detailed Patient Reported Outcome Questionnaire available for

HES data do not contain any patient level socioeconomic information, but we are able to embed characteristics at the neighborhood level via the patient's postcode district and LSOA.[18] Socioeconomic status is measured using the neighborhood level Index of Multiple Deprivation (IMD) as compiled by the Office for National Statistics.[19] This measure allows us to rank neighborhoods from the least to the most deprived. We rescale the IMD, henceforth referred to as 'deprivation', to lie between zero and one. Higher values imply higher deprivation. As documented by Chard *et al.* (2011) and elsewhere, ISP patients are on average less deprived than patients that are treated by NHS hospitals. In our sample, the average NHS patient lived in an area with a deprivation rank of 0.45, compared to 0.39 for the average ISP patient.

The final set of characteristics is the historic referral patterns of the patient's GP. This reflects the likely importance of the GP in the referral decision. From HES outpatient records detailing GP practice referrals in 2011/12 in the Orthopaedics and Trauma specialty, which is the largest specialty by volume in the NHS and contains consultants who would see joint replacement patients, we calculate a Herfindahl-Hirschman Index (HHI) of the concentration of referrals across providers for each GP practice.[20] We also use all referrals from 2009/10 to 2011/12 to calculate the share of referrals to ISPs over those three years. Table 3 shows that patients who choose ISPs are registered at GP practices with lower concentrations of referrals. The average patient treated by an ISP was registered with a GP practice that referred 13.2% of patients to ISPs, compared to an average of 7.6% for those treated by an NHS hospital.

3.2 Hospital Characteristics

We construct hospital attributes for 314 hospitals in our sample, which conducted at least 20 NHS-funded hip replacements in 2012/13. Of these, 119 (or 38%) are ISPs. This share is higher than the share of patients treated by ISPs of just over 20%, because ISPs treat fewer patients per hospital (103 on average, compared to an average of 253 for NHS hospitals).

two thirds of the sample. Even for those who report having a previous stroke or heart attack, 10% have a hip replacement conducted by an ISP.

[18]Lower Super Output Areas are statistical geographical aggregation units with no administrative jurisdiction, similar to a census tract, and are designed to be as homogeneous as possible with respect to population composition. They contain an average of 1,500 individuals. There are approximately 32,500 LSOAs in England.

[19]The Index of Multiple Deprivation (IMD) is an local area based measure of deprivation produced by the UK government that includes measures of income, employment, health deprivation and disability, education skills and training, barriers to housing and services, crime and the living environment. We use the version produced in 2010. Please see https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010 for more details. This is the same measure used to capture the economic conditions as Gaynor *et al.* (2016)

[20]This is given by the sum of squared referral shares of each hospital that the GP practice refers to.

Previous analyses of hospital choice in England and elsewhere have shown that distance is the principal hospital attribute driving patient decisions (Beckert *et al.*, 2012; Gaynor *et al.*, 2016).[21] Figure 5 shows the distribution of patient choices, with hospitals ordered by distance.[22] The black bars indicate that 45% of patients chose their closest hospital and 82% chose amongst their closest three. When we exclude ISPs - which in some cases are the nearest provider - and just look at patients that chose NHS hospitals, shown in the grey bars, 66% chose their closest NHS hospital and 91% chose from among their three closest. The closest NHS hospital and ISP are on average 8.9km (s.d. 7.3km) and 12.7km (s.d. 10.8km) away, respectively.

Further hospital attributes driving patients' decisions are summarized in Table 2.[23] We control for hospitals' clinical quality using the ratio of 30-day all-cause emergency readmissions for hip replacement relative to expected readmissions at the hospital level, given the hospital's case mix.[24] A ratio of unity indicates that the rate of readmissions is as expected, higher ratios imply higher than expected readmissions, i.e. lower clinical quality. The mean readmission ratio is higher for NHS hospitals than ISPs. However, there is substantial overlap in the distributions of readmission ratios across hospital types.

We also control for hospital quality more summarily, in terms of broad hospital type categorisations. The first category comprises "early FTs", i.e. NHS hospitals that became a "Foundation Trust" (FT) up to and including 2006. Foundation Trust status allows hospitals a degree of independence from the Department of Health. The first hospitals were granted Foundation Trust status in 2004. These hospitals were typically of higher quality in terms of both, management and clinical outcomes. In subsequent years, the majority of hospitals have become Foundation Trusts, but as a consequence the average quality of FT hospitals has declined. We use the cut-off of 2006 in our definition of early FTs as a measure of the highest quality hospitals. 16% of NHS hospitals are classified as early FTs.

The second category comprises Specialist Orthopaedic hospitals. There are eight in total, seven NHS hospitals and one ISP. Specialist orthopaedic hospitals treat a larger number of orthopaedic patients, and they may be a particularly relevant alternative, not only for

[21] The same pattern exists for education choices and other public services (Burgess *et al.*, 2015).

[22] Distance is measured in a straight line from the centroid of the patient's Lower Super Output Area to the hospital postal code.

[23] A large range of quality measures is recorded for NHS hospitals, but very few of these are available for ISPs. All the quality measures we use are therefore constructed using the information available in HES. For example, while PROMS data are relatively abundant for treatments at NHS hospitals, they are sparse for treatments at ISPs. We therefore decided not to construct quality measures from PROMS data.

[24] Readmissions include any emergency readmission to any hospital for any cause within 30 days. Expected admissions are constructed by regressing readmissions on age, sex, and prior admissions, and underlying co-morbidities of hospital patients. We calculate average predicted readmission rates for each hospital and then divide by the observed readmission rate.

patients living nearby.

3.3 Descriptive Evidence on Sorting

Table 3 reveals that ISP patients are on average healthier, richer, more likely to be white than those treated by NHS hospitals. Differences by health are to be expected, given that there are health restrictions on eligibility for ISP treatment. Reasons for differences by ethnicity and deprivation are less clear. In this Section we provide evidence on how the three mechanisms we highlight - the geographic distribution of ISPs, underlying health, and market imperfections - might explain the patterns we observe.

In Table 4, we consider the relationship between the location of ISPs and the gradient in ISP use by deprivation. The Table shows the share of patients that have an ISP as the closest provider and the share of patients that are treated by an ISP by local area deprivation quintile for each year between 2006/07 and 2011/12. The share of patients who had an ISP as their closest provider increased over time in all quintiles, with the largest increases between 2007/08 and 2008/09 and between 2009/10 and 2010/11. However, in all years the share of patients living close to an ISP decreases with local area deprivation, with both relative and absolute increases in the gradient after 2008/09. By 2013/13, 37.1% of patients living in the richest fifth of areas have an ISP as their closest provider, compared to 22.1% of those in the most deprived fifth.

Differences in ISP use by local area deprivation may be a consequence of where ISPs are located or the patients that choose ISPs conditional on location. Table 4 shows that the share of patients treated by ISPs increases over time, with particularly large increases coinciding with the growth in the number of providers. In 2006/07 and 2007/08, ISP use was fairly flat across the deprivation distribution. From 2008/09 onwards, ISP use declines with local area deprivation, with ISP use substantially lower in the bottom two (poorest) quintiles of distribution relative to the top (richest) three quintiles. The deprivation gradient steepens somewhat over time, with the share of patients in the richest quintile 35% higher than those in the most deprived quintile in 2008/09, rising steadily to 79% higher in 2012/13. Calculating the share of patients who choose an ISP conditional on having an ISP as the closest provider is stable over time and fairly flat across the deprivation distribution until 2010/11, after which a negative gradient developed. In 2012/13, 32% of those in the richest quintile who had an ISP as the closest provider chose an ISP, compared to 22% in the most deprived quintile.

There is a similar pattern in how ISPs are distributed across England with respect to ethnicity. In 2009/10, 15.2% of white hip replacement patients have an ISP as their nearest

hospital, compared to 9.4% of ethnic minority patients. By 2012/13, this had increased to 31.9% of white patients and 25.1% of ethnic minority patients. Over the same period the share of patients that were treated by ISPs increased from 12.6% to 20.4% for white patients and from 5.7% to 8.2% for ethnic minorities. The difference between ISP use by ethnic minority status is therefore not just the result of the geographical distribution of providers. Ethnic minorities are less likely to choose ISPs conditional on location.

The second mechanism we highlight is the potential contribution of differences in underlying health to sorting. Figure 6 shows the share of patients that had a previous emergency admission, mild co-morbidities and severe comorbidities, by deprivation quintile.[25] As expected, underlying health declines with local area deprivation. For previous emergency admissions and mild comorbidities, the declines in health are largely confined to the most deprived half of the distribution. There is a small difference in the underlying health by ethnicity: the share of ethnic minority patients with prior emergency admissions is 2 percentage points higher than for the white population, and ethnic minority patients exhibit slightly more comorbidities. Differences in health may therefore explain a part of observed sorting, particularly by deprivation. The sizable differences by ethnic minority status appear harder to explain. For example, the share of ethnic minorities that choose an ISP (7.5%) is approximately equal to share of patients with both low income (living in the poorest fifth of local areas) and poor underlying health (have a prior emergency admission) who choose an ISP.

The final mechanism is potential imperfections, or frictions, in the market for NHS-funded care. In particular, we focus on the role the GP may play in creating imperfections. Table 3 highlights that there is a strong correlation between prior GP referral patterns and choosing an ISP. Comparing the concentration of GP practices referrals across patients in our sample shows that concentration does increase slightly with deprivation, with a mean of 0.62 for those in the most deprived fifth of areas, compared to 0.58 in the least deprived quintile [26]. There is very little difference in mean concentration by ethnic minority status (0.6 for the white population and 0.59 for ethnic minorities). Interpreting these differences by GP referral characteristics is particularly difficult, as referrals are a function of multiple mechanisms, including the size and composition of the local health economy, the needs of patients, and potentially GP preferences and information sets.

Taken together the descriptive evidence in this Section indicates that the geographical distribution of ISPs and the differences in population health are likely to explain some of the differences in ISP use. However, understanding the relative importance of the mechanisms

[25] Here, the value 1 represents the least deprived quintile.

[26] The exception to this pattern is London, where concentration decreases with local area deprivation

we highlight and accounting for differences in the characteristics of patients and the hospitals available to them requires a model of hospital choice.

4 Econometric Choice Model

4.1 Patient Level Choice Model

We use a random utility model (RUM) to describe the patient's discrete hospital choice problem. We consider a mixed multinomial logit (MMNL) model that allows us to capture a wide spectrum of patient level heterogeneity, exhibits unrestricted substitution patterns and does not impose a correlation structure across choice alternatives.[27].

Consider hip replacement patient i. Let $g(i)$ denote i's GP (practice).[28] And suppose that $g(i)$ offers i to choose among a set of NHS hospitals $\mathcal{N}_{g(i)}$ and a set of ISPs $\mathcal{I}_{g(i)}$. Then, patient i's choice set is given by $J_{g(i)} = \mathcal{N}_{g(i)} \cup \mathcal{I}_{g(i)}$. Let U_{ij} denote i's indirect conditional utility from having the procedure carried out at hospital j, $j \in \mathcal{J}_{g(i)}$, and consider the specification

$$U_{ij} = \mathbf{x}'_{ij}\beta_i + \epsilon_{ij},$$

where $\mathbf{x}_{ij}$ is a K-vector of hospital attributes that may vary across patients, such as distance between patient and hospital. The vector β_i is a vector of possibly random coefficients,

$$\beta_{ik} = \beta_k + \mathbf{z}'_i\theta_k + \sigma_k\nu_{ik}, \quad k = 1, \cdots, K,$$

where $\mathbf{z}_i$ is a vector of patient level characteristics, $\sigma_{ik} > 0$ for random coefficient and zero otherwise, and ν_{ik} is an independent standard normally distributed random variable. In this model, $\beta_k + \mathbf{z}'_i\theta_k$ captures the conditional mean of the random coefficient β_{ik} on hospital attribute k, given patient characteristics $\mathbf{z}_i$, or the observed heterogeneity in i's valuation of attribute k. The contribution $\sigma_k\nu_{ik}$ to β_{ik}, in turn, captures unobserved heterogeneity in i's valuation of attribute k. The term ϵ_{ij} captures unobserved taste variation across hospitals that is not quantified by hospital attributes $\mathbf{x}_{ij}$. The collection $\{\epsilon_{ij}, j \in \mathcal{J}_{g(i)}\}$ is assumed to be i.i.d. $EV(0,1)$. Patient i chooses the hospital associated with the highest indirect conditional utility. Let $D_{ij} = 1$ if patient i is observed to choose alternative j, and $D_{ij} = 0$

[27] More tightly specified alternatives in the logit family, such as conditional or nested logit models, while yielding more efficient estimates, embed the risk of being misspecified and consequently inducing inconsistent estimators. As demonstrated by McFadden and Train (2000), an appropriately rich MMNL specification can arbitrarily closely approximate any RUM for discrete choice. This flexibility renders it an attractive econometric framework for analysis.

[28] In line with the informational content of our data, which identify a patient's GP practice, but not the individual GP, in much of our discussion we refer to GP and GP practice synonymously.

otherwise. Then,

$$D_{ij} = 1 \quad \Leftrightarrow \quad U_{ij} = \max\{U_{in}, n \in \mathcal{J}_{g(i)}\}.$$

This model can be estimated by Maximum Simulated Likelihood (Hajivassiliou, 2000).

We include an ISP dummy among those attributes in $\mathbf{x}_{ij}$ that carry a random coefficient, i.e. $x_{ijk} = 1_{\{j \in \mathcal{I}_{g(i)}\}}$ and $\sigma_k \geq 0$. Heterogeneity in sorting into ISPs then operates through the interactions of x_{ijk} with $\mathbf{z}_i$. By controlling for i's health and GP $g(i)$'s referral pattern among $\mathbf{z}_i$, the model allows us to identify differential sorting, conditional on access and health, with respect to other patient socio-demographics, such as ethnicity and neigborhood deprivation. Our MMNL model endows two other hospital attributes with random coefficients: distance, and the 30-day emergency readmissions ratio. Interactions with all continuous variables are constrained to be linear, with the exception of local area deprivation where we include separate terms for the most and least deprived half of the distribution. We also include pairwise interactions between the hospital attributes.

4.2 Choice Sets

The choice model, as specified, assumes that choice sets $\mathcal{J}_{g(i)}$ may vary across GP practices, but do not vary across patients within GP practice. For our application to hospital choice for elective hip replacements, a definition of these choice sets is required. Even though hospital choice for various elective procedures has been considered in numerous earlier studies, the literature does not offer a clear cut approach to such a definition. This is notwithstanding the fact that some authors emphasize that choice sets as considered by decision makers do matter.[29]

We consider two definitions of the choice sets $\mathcal{J}_{g(i)}$. In line with standard practice (Beckert *et al.*, 2012; Ho, 2006), the first approach defines $\mathcal{J}_{g(i)}$ by distance to the GP practice as patients predominantly choose hospitals that are nearby. The challenge is how to define the appropriate distance cut-off or number of hospitals. The smaller the choice set, the more patients will be dropped for choosing hospitals outside the choice set, but larger choice sets may provide a large number of hospitals that are never considered by patients. Our definition is based on the ten nearest hospitals conducting at least 20 procedures, plus all specialist hospitals within 50km; we refer to choice sets according to this definition as "distance choice set"[30]. The definition we use generates a sample of 62,695 patients, the average number of

[29] In the context of health care choice, see for example Beckert and Collyer (2017); Gaynor *et al.* (2016).

[30] Distances are measured in a straight line from the centroid or central point of the patient's Lower Super Output Area to the post code of the hospital. We include only hospitals that perform at least 20 hip operations in 2012/13, as hospitals that perform very low volumes may not be in patient choice sets. This is a particular problem for ISPs where a relatively large fraction of sites perform very few procedures. For

ISPs in their distance choice sets is 3.9, and 80 per cent of them have between 3 and 5 ISPs. Since most patients in our data, in fact, choose from the nearest five hospitals, the implicit assumption underlying our "distance choice set" definition is that they look a bit beyond when considering their options.

As a robustness check, we consider a variant of this definition of distance choice set. This alternative definition considers the patient's 30 closest hospitals, up to a maximal distance of 30km relative to the centroid of the patient's LSOA. With this definition, patients in metropolitan areas with a high hospital density have choice sets with many alternatives and a relatively small geographic spread, while patients in areas with low hospital density have choice sets with relatively few alternatives and maximal geographic spread, i.e. up to the imposed 30km boundary.

The second approach defines $\mathcal{J}_{g(i)}$ as the set of hospital alternatives that the GP referred patients to over the last three years; we refer to choice sets according to this definition as "GP choice sets".[31]. The approach is motivated by the central importance of GPs in guiding patients when they exercise choice in the English NHS.[32] We assume that this approach should form a close proxy for the choices offered to patients, as referral to a provider indicates either pre-existing knowledge or subsequent knowledge obtained following feedback from patients (Dixon *et al.*, 2010).

The use of prior outpatient referrals should be a reasonable approximation of the alternative that may have been considered, and avoids any restrictions on the size of the choice set from using only hip replacement patients. Again, we restrict to hospitals that conducted 20 or more hip replacements in 2012/13. There is a possibility that this approach falsely excludes providers that are never chosen, but given the costs of transmitting information about additional providers to patients, it seems unlikely that GPs would continue to offer providers that patients never chose. The aim is to proxy for the set of hospitals that the patient is likely to have discussed with their GP. Estimating the model with these choice sets means that patient choices can vary across GPs through both the hospital attribute/GP referral practices interactions included in the "distance choice set" model, and differences in

example, reducing the minimum threshold from 20 to 5 procedures increases the number of relevant ISPs by 22%, but these smaller sites accounted for just 2.7% of ISP patients in 2010/11 and 0.5% of all NHS funded patients. We include additional Specialist Orthopaedic hospitals within 50km, as these are hospitals that patient predominately choose when not picking one of their nearest 10. Patients that chose a hospital outside their nearest 10, plus nearby specialist hospitals, are dropped, which removes 7% of the patient sample.

[31] To construct these choices, we take all referrals by that GP within Orthopaedics and Trauma over the period 2009 to 2012 (with an average of 420 referrals), and include hospitals where the GP referred more than 0.5% of patients, plus any sites where any hip replacement patients were referred to in our hip replacement sample. We exclude patients that were not included in the "distance choice set" sample, to ensure that model estimates are comparable

[32] Dixon *et al.* (2010)

the sets of hospitals that are presented to patients by GPs. We exclude patients who are not included in the distance choice set sample, to ensure that paramter estimates are comparable across model specifications.

The sets of hospitals presented to patients by their GP can be thought of as outcome of at least three different competing processes: (i) GPs acting as an altruistic agent for the patient; (ii) costs of acquiring and disseminating information limiting the options that are presented to patients; and, (iii) incentives of GPs not aligned with those of the patient. Appendix A discusses each mechanism in more detail and presents estimates from a model that examines the determinants of the GP's binary decision whether or not to include the hospital in the GP choice set[33]. We find that higher quality hospitals are more likely to be included in GP choice sets, but the magnitude of the quality effect is small. By contrast, the inclusion of a hospital in a GP choice set is strongly associated with features of local health care organisation unrelated to patient health[34]. And these determinants dominate the hospital quality effects or population health characteristics.

Figures 2 - 4 compare the composition of distance and GP choice sets. There is large variation in the number of choices that are offered across GP practices in the "GP choice set"; also, the majority of GP practices refer to far fewer than the 10 hospitals we consider in the distance choice set model. Figure 2 shows that the distribution of the number of alternatives in the GP choice sets is roughly normal. The median number of ISP alternatives is 4 in the distance choice set but only 1 in the GP choice set.

5 Results

5.1 Baseline Results

Tables 5 and 6 show parameter estimates from the mixed logit models based on the distance and GP choice sets[35]

[33]While a formally incorporating the GP level choice set formation process into our model is beyond the scope of this paper, the paper provides descriptive evidence for the factors that are likely to determine how the choice set is formed. See Beckert and Collyer (2017), Goeree (2008) and Gaynor *et al.* (2016) for examples fo structural modelling of choice set design; Crawford *et al.* (2016) study demand estimation in the absence of accurate and quantifiable information on the true choice sets.

[34]In particular, GPs are much more likely to refer to NHS hospitals within the same PCT. This may reflect some inertia in referral practices dating back to block contracting in the early 2000s, or a desire to maintain the revenues of hospitals that provide emergency care for their patients. For ISPs, the hospitals most likely to be included are the first Wave of ISTCs. The hospitals received a block payment for a fixed number of procedures, whether or not these procedures were carried out. GPs were therefore encouraged to make sure these procedure slots were used.

[35]The tables provide the most important estimates within the context of our substantive discussion. The remaining parameter estimates are available on request.

The parameter estimates for the mean valuations of hospital attributes are presented in Table 5. The parameter estimates from the "distance choice set" model on the left hand side of the Table provide similar results to existing work on patient choice. Patients have a preference for shorter travel distances, shorter waiting times and higher quality. We find that specialist hospitals are more likely to be chosen and ISPs less likely to be chosen. The random coefficient parameters indicate significant heterogeneity in valuations of distance and ISPs, but no unobserved variations in the emergency readmission rate. This finding might be explained by patients deferring to their GP with regard to quality assessments (Dixon and Robertson, 2009; Monitor, 2015). In an incomplete information setting like the one considered here, quality is likely assessed via the patients' GPs who possess superior information. GPs, in turn, may have relatively homogeneous information on hospital quality and thus are unlikely to vary significantly in their quality assessments.

The GP choice set model produces a similar pattern of estimates. Responses to quality, as measured by emergency readmissions and early FT status are slightly smaller (relative to other attributes such as distance). This might be explained by GP pre-selection eliminating lower quality hospitals. Specialist hospitals are also valued more highly under the GP choice set model. This model also identifies patient health, captured by prior emergency admissions, as characteristic that tends to preclude treatment at specialist hospitals.

Table 6 presents parameter estimates for interactions between hospital type and patient and GP characteristics. Starting again with the distance choice set parameter estimates, we find that ethnic minorities and those with underlying ill-health are less likely to seek treatment at an ISP. The interactions between deprivation and ISP use are positive and statistically significant at the 10% level for the least deprived half of the the distribution and negative but not statistically significant for the poorest half. Patients who are registered with GP practices with high referral concentrations or low prior referral shares to ISPs are less likely to choose an ISP, which is consistent with an important role played by GPs in the decision making process.

The results of our robustness check on the underlying distance choice set definition, using the variant of up to the 30 nearest hospitals within 30km, are reported in Table 7. The estimates obtained from this alternative choice set definition are closely aligned with those obtained on the basis of the original distance choice set definition, with the exception of the magnitude of some of the coefficient estimates on interaction terms and of the ISP coefficient.[36] Qualitatively, the estimates display the same picture as those from the model

[36]This concerns the magnitude, but not the sign or statistical significance, of the coefficient estimates on the interaction of the ISP dummy with ethnic minority, deprivation and the GP's share of prior referrals to ISPs.

based on the original distance choice set definition.

Turning now the the parameter estimates of the GP choice set model, they produce a similar pattern of results with respect to ethnicity and health. In both models, the magnitude of the interaction between ISP and ethnic minority is approximately equal to the interaction between ISP and having a previous emergency hospital admission. These parameters indicate that ISP patients are healthier even accounting for distance and the hospital choice alternatives that are available, which is unsurprising given the eligibility criteria for ISPs. Ethnic minorities are less likely to use ISPs, even when controlling for distance to ISPs, differences in deprivation, or observable measures of health.

The move to the GP choice set model has a greater impact on the parameters of the interactions between ISP and local area deprivation. The interaction for the least deprived areas is positive and not statistically significant, whereas whereas the parameter for the poorest half is negative and statistically significant at the 5% level. This pattern is similar to that observed in the raw data, and indicates that lower ISP use by the most deprived is not fully explained by distance, observable differences in underlying health, or the characteristics of GPs that treat more deprived patients. GP prior referral characteristics continue to play a strong role in the GP choice set model, although the relative importance of the prior concentration of GP referrals is reduced. This is presumably because concentration in part captures for the number of hospitals in the GP's choice set.

6 Counterfactual Simulations

6.1 Preliminaries and Underlying Assumptions

In this section, we present results of a series of counterfactual simulations. The simulations aim at exploring the relative contributions of potentially important drivers of patient choice that are unrelated to hospital attributes and instead derive from patient characteristics and guidance that patients receive from their GP. We emphasize that we do not interpret the results presented here as predictions from policy simulations but rather as stress tests of the model, in order to assess the relative substantive importance of the respective drivers of choice.

The main reason why we caution against over-interpreting the level predictions of the counterfactual simulations is that they are likely to be contaminated by several sources of potential bias. In general, our maximum likelihood (ML) estimation methodology assesses goodness of fit in terms of the value of the likelihood function at the ML estimates, not in terms of a criterion based on prediction errors.

That aside, discrete choice models like the ones estimated here assign non-zero probability to every choice alternative in the choice set. In our context, this is relatively unproblematic as far as NHS hospitals are concerned, but potentially problematic when considering ISPs because some patients may not be considered eligible for treatment by ISPs, due to health conditions unobserved by the econometrician. Non-zero predicted choice probabilities for ISPs induce an upward bias in counterfactual predictions of ISP choice.

Moreover, when predicting ISP choice conditional on a specific patient characteristic, for example, then the formally correct approach to calculate the predicted conditional choice probability is to integrate out all other patient characteristics, conditional on the specific characteristic under consideration. But the multivariate conditional distribution of such other characteristics is difficult to estimate, even with the rich data we have, and attempts to estimate it entail risks of bias and non-robustness of their own.

A straightforward way to estimate predicted ISP uptake, conditional on a patient characteristic, say deprivation, is to sum the predicted choice probabilities conditional on that characteristic. If deprived patients have relatively bad health and hence their predicted ISP choice probabilities are relatively low, then in the formally correct approach these low predicted probabilities would have higher weight in the calculation of expected ISP choice. Hence, simply summing unweighted predicted ISP choice probabilities will add to the aforementioned upward bias. Conversely, if wealthy patients are in relatively good health and hence their predicted ISP choice probabilities are relatively high, then these high predicted probabilities would have higher weight in the correct approach to calculation of expected ISP choice, and the sum of unweighted predicted ISP probabilities will tend to underestimate the expected ISP uptake. Therefore, comparing predicted ISP uptake by sums of predicted choice probabilities for high and low deprivation patients will tend to underestimate the discrepancy between these patient groups. Acknowledging this limitation of our approach to prediction, we accordingly limit our interpretation to the comparison of proportionate changes in ISP uptake in a number of counterfactual simulations.

Our assumptions for these simulations are as follows. When considering counterfactual expansions in choice sets, we assume that the costs to the GP of providing wider choice are minimal, i.e. there is no GP-level capacity constraint. Furthermore, we assume that there is no capacity constraint at the hospital level, so that additional patients treated at a hospital under the counterfactual do not change the attributes of that hospital. Given that hospital attributes such as waiting times may change, the predicted demand shift is an upper bound of the expected effects.[37]

[37]Hospitals that are predicted to face increased patient demand under the counterfactual with given attributes are likely to increase waiting times if they are capacity comnstrained, which in equilibrium would

6.2 Simulation Results

We use the model to conduct two sets of simulations to examine factors that might explain the patterns of sorting we observe by deprivation and ethnicity. The first set sequentially removes differences between patient groups to ascertain the relative importance of differences in health, preferences and GP behaviour. The second considers the role of frictions that we observe affecting the inclusion of sites in GP choice sets.

6.2.1 Mechanisms

Table 8 shows the predicted ISP shares by income quintile and ethnic minority status when we successively shut down mechanisms that could contribute to the differences in ISP use.

The first two columns show the total number of hip replacements that took place over the observation horizon and the share of those hip replacements that were performed by ISPs. The third column gives the predicted shares of ISP patients using our baseline model. The predicted shares decrease as deprivation increases, but the gradient is less steep than in the data, for the reasons outlined above. The remaining five columns present results from simulations where we successively remove potential channels that could generate this sorting.

In step 1, we equalise health across the socio-economic distribution by giving all individuals the mean health across all four health measures. The aim is to eliminate differences in ISP use by ethnicity and deprivation that are derived from differences in health. The effect is to increase the total number of hip replacements at ISPs in all income quintiles. This is because many individuals will be moved away from having a pre-existing health condition to a relatively low probability of that condition. The absolute increases are relatively similar across the income quintiles, but the relative growth is larger for the bottom quintile (22%) than for the richest three income quintiles (all around 13%). The gradient in ISP use between richest and poorest patients shrinks from 5.7 percentage points (26.4 % - 20.7 %) to 4.7 percentage points (29.9 % - 25.2 %), This equates to an absolute decrease of 1 percentage point, which is 17.5% of the original gradient. Step 2 equalises ethnicity, eliminating all interactions with ethnicity, so that all patients are assumed to behave like those who are White British. This has minimal effects on the distribution of ISPs by deprivation, as there are only a small number of ethnic minority patients. In Step 3, we equalise preferences by eliminating interactions between hospital attributes and deprivation. Again, there is very little change in the predicted number of ISP patients and the deprivation gradient. This

dampen predicted demand. Conversely, hospitals that are predicted to face diminished demand under the counterfactual with given attributes are likely to be able to reduce waiting times, which in turn would enhance predicted demand. Beckert and Schiraldi (2017) provide an equilibrium analysis in which waiting times respond endogenously to changes in demand.

implies that given the set of hospitals available, lower ISP use is not attributable to patients from poorer areas 'preferring' to stay with the local NHS hospital.

After removing differences in health and preferences, the remaining differences in ISP use should be explained by ISP availability. ISPs may be unavailable to patients because (i) they are not offered the option of choosing an ISP in their local area or (ii) there are no ISPs within their local area. Simulation step 4 removes all interactions with GP characteristics. As a consequence, we assume that all choices available will be given equal weight. This results in a small increase in ISP use for all income quintiles, on the order of 2% for all income quintiles. The final step replaces the GP choice set with the distance choice set, and assumes that all patients choose between the 10 closest providers. This has a sizable impact on the share of patients treated by ISPs in all income quintiles, as the distance choice set is typically larger than the GP choice set, particularly in terms of ISPs. The gradient in ISP use between richest and poorest patients shrinks again, to 3.5 percentage points (42.2 % - 38.7 %). As health, ethnicity and GP referrals have been equalised, the only remaining differences come from differences in the ten nearest hospitals. This remaining 3.5 percentage points, or 61% of the original difference from the model is therefore attributable to differences in the geographic distribution of hospitals.

A similar set of results are observed when we consider the same set of simulations by ethnicity. The model predicts a gap of 9.6 percentage points in ISP use between white and ethnic minority patients (25 % vs 15.4 %). Step 1, which equalises health, reduces the difference in ISP use by ethnicity from 9.6 percentage points to 4.2 percentage points, or 56% of the original difference. Step 2 and 3 are equivalent in this simulation and have almost no impact on ISP use. Step 4, which removes GP preferences, reduces ISP use by minorities. As for the income simulations, Step 5 also has a large impact on predicted ISP use. For white patients, applying the distance choice set increases ISP by 41% compared to 63% for ethnic minority patients. After step 4, the predicted difference in ISP use remains 4.2 percentage points or 44 % of the total.

It may be worth noting that any ethnic minority effect is likely to be accentuated by London. There are very few ISPs in London, but a high share of ethnic minority patients. In the raw data, the share of ethnic minorities that use ISPs is 7.7 percent when London is included, and 10.8 percent when London is excluded. [38]

Taken together, the simulations reveal three points of note. First, differences in health do explain a sizable and important portion of the gradient in ISP use by local area deprivation and ethnicity. Second, patients in poorer areas are presented with a narrower range of hospitals when making choices, even when taking the local health economy into account.

[38] The same is not true for deprivation: The share of ISP use with and without London is very similar.

Third, our simulations show that even removing differences in health, preferences, and the choice of providers offered, ethnic minority patients and those living in more deprived areas are less likely to choose an ISP. This can only be attributed to the geographical location of ISPs. Conditional on hospital placement, this response from patients may well be efficient. The question in policy terms, is whether such a geographic distribution of service provision is desirable.

6.2.2 Market Imperfections

Interpreting the role of GPs is difficult because GPs fulfil a dual role, as advocates for their patients on the one hand and as gatekeepers for the NHS on the other. Any observed differences in referrals across practices could therefore reflect differences in the needs of patients, which may not be observable to the econometrician, GP's responding to incentives or pressures unrelated to patient need, or information imperfections.

In this section we therefore focus on two important market features that are potentially related to the incentives that GPs face when referring patients and that are unrelated to patient health: the type of ISP contract and the role of administrative boundaries.

Table 9 shows the share of patients from each provider type by local area deprivation quintile observed in the data. The first column shows the distribution for NHS hospitals. Patients in the richest 3 quintiles are over-represented, and those in the poorest two quintiles are under-represented. This is because we define deprivation based on the whole of England rather than the sample, and patients living in richer areas are more likely to have hip replacements. The higher rate of hip replacements in poorer areas in part reflects the age structure of the poorest parts of the country, although there is evidence that these areas are also under-served relative to need (Judge *et al.*, 2010). The second column shows the distribution for private hospitals (not ISTCs). As expected, private hospital patients are most likely to be in the richest two quintiles of the country. Only 7.3% of private hospital patients lived in the poorest quintile of areas, relative to 13.7% of NHS hospital patients. This is unsurprising, as private hospitals are typically located in richer areas. However, Wave 1 ISTCs, which were centrally planned and initially had block contracts, show a similar pattern, treating only 5.5% of patients in the most deprived quintile of areas. Wave 2 ISTCs, which were also centrally planned but without guaranteed volume contracts, have a distribution which is closest to NHS providers, but only treat 1,462 patients, so the impact on the overall ISP gradient is relatively minor.

To understand the relative roles of ISTCs and private hospitals in generating the observed gradient in ISP use, we simulate the predicted probability of treatment by an ISTC when all private hospitals are eliminated. We then examine how the gradient in use at ISTCs changes.

This is important, because the establishment of ISTCs involved negotiations between the NHS and the private health care providers about the locations in which these contracts would be issued, with need being a key determinant. By contrast, by 2012/13, almost all private hospitals had entered the market, with location determined purely on the locations of pre-existing private hospitals.

Of the 11,000 private hospital patients to be redistributed, slightly more than 1,700 are predicted to attend an ISTC, with the rest redistributed to NHS hospitals. However, the redistribution of these patients to ISTCs does not change the overall deprivation gradient. This shows that the gradient is not the result of allowing pre-existing private hospitals to enter the market for NHS funded care in terms of the relative shares treated in the richest and poorest quintiles. However, because among NHS-funded patients treated by ISPs in 2012/13 private hospitals treated the majority of patients, access to private hospitals does increase the absolute difference between the numbers treated by ISPs in the richest and the poorest quintile. This also suggests the incentives to refer to wave 1 ISTCs are not what is generating the local area deprivation gradient.

The determinants of the GP choice in Appendix A1 indicate that administrative boundaries have an important role in determining whether a hospital is included in a GP choice set. In particular, hospitals are more likely to be included if they are located in the same PCT as the GP practice. To examine the impact of these boundaries on ISP use, we take a sample of GP practices that are within 1500 metres of each other but lie on opposite sides of the PCT boundary. The median absolute difference in predicted ISP share between practices on either side of PCT boundaries is 3.7 percentage points within PCT, but 5.6 percentage points between PCTs. However, on the 75th and 90th percentile, the predicted ISP shares between PCT practices are closer together. This suggests that the impact of administrative boundaries on ISP shares is relatively small. This is what would be expected in a system in which payment is on a prospective basis.

6.3 The costs of sorting

Our parameter estimates and simulations illustrate that differences in underlying health are important sources of sorting. This may represent an efficient allocation of resources as ISPs do not have the specialist facilities needed to appropriately treat those with complex needs. However, the payments the NHS makes to providers have only very limited gradations in payments to take account of differences in patient health that affect the costs of treatments patients receive.

Table 10 shows the main NHS tariffs or payments that are applied to hip replacement

patients in our sample. These tariffs are set to reflect the average cost of providing a particular bundle of care. The first and second column show the number of patients in each HRG category and the associated tariff. Three quarters of the sample fall into a single category (Major Hip Procedures for non Trauma Category 1 without Complications or Comorbidities HB12C), which had a price of £5,832 in 2013/13 (NHS England, 2013). Only around 10% fall into HRGs for those with complications and comorbidities, which are associated with higher tariffs.

The fourth column of Table 10 shows the share of patients within each tariff that are treated by ISPs. ISPs treat 22% of patients in the "simple" HRG, but only 9.1% of those within the tariff for those with "complications and comorbidities" and 7.6% within the tariff for those with "major complications and comorbidities". This sorting across tariffs is compatible with the goals of the ISP reforms, as competition even within a segment of the market may be sufficient to generate pressure to improve quality for all. The distribution is also consistent with improving patient welfare, as NHS hospitals are likely to be better equipped to deal with patients with worse underlying health. NHS hospitals may see a higher share of complicated patients, but they are compensated for the associated costs by higher tariff payments.

In this section we consider the implications for NHS hospitals and for competition when sorting occurs within HRGs. The third column of Table 10 shows the share of patients with previous emergency admissions to hospitals. There are two points to note here. The first is that, as might be expected, the share of patients with previous emergency admissions for the "simple" tariff is approximately half (0.17) that of patients for tariffs with some and major complications (0.345 and 0.34 respectively). Second, although the share of patients with previous admissions is lower for those in the "simple" HRG, there is a considerable share that do have underlying health conditions. In additional to the 17% who have had a previous emergency admission, 53% have had a previous elective admission, 27% have a non-zero Charlson Index, and 10% are over 80. This suggests that the true costs to hospitals of treating patients within the same HRG are likely to vary. Any sorting according to underlying health and the associated costs would not be reflected in tariff payments. In particular, if healthier patients choose ISPs, the payments received by NHS hospitals and ISPs will be the same, but the average cost per patient treated by an NHS hospital will rise. This limits the ability of NHS hospitals to cross-subsidise across patients.

Our data do not allow us to calculate the precise cost to the hospital of treating a particular patient. We therefore adopt a simplified approach where we assume that differences in cost are proxied for by length of stay. Estimates of costs per bed day vary, but we use the per day long stay payment in 2012/13 of £231 per day (NHS England, 2013).

Restricting our attention to just those patients who have a "simple" HRG (HB12C), the data show that patients treated by NHS hospitals have a mean stay of 4.5 days, compared to 3.7 days for patient treated by ISPs. This shorter length of stay for ISP patients may be explained by two factors. First, ISPs may deliver care faster and discharge patients earlier. Second, patients treated by ISPs may be healthier and less costly to treat.

To establish the importance of this second channel, we run a linear model where length of stay for patients treated by NHS hospitals is estimated as a function of a quadratic in age, of sex, previous admissions and Charlson index, ethnic minority status and local area deprivation. We then use this model to predict the length of stay for all patients. For ISP patients, we therefore generate the expected length of stay had they chosen an NHS hospital. The distribution of predicted lengths of stay by provider type is presented in Figure 7. The distribution is more concentrated for ISP patients, with a higher proportion with short expected lengths of stay (up to 5 days) and a lower proportion having expected lengths of stay of 6 days or more. The mean is 4.44 for NHS patients and 4.26 for patients treated by ISPs. This shows that the majority of the differences in the raw data are driven by ISPs discharging patients sooner, but that patients treated by ISPs are also slightly less complex [39].

Using these estimates we conduct a counterfactual calculation where the 10,728 ISP patients replace a random sample of the same number of patient treated in an NHS hospital. To do this we use the sum of the predicted lengths of stay across all ISP patients, giving a total number of bed days of 45,653 days. We then draw a random sample of 10,728 patients from NHS hospitals from within the same HRG, and again calculate the predicted number of bed days. For the sample we draw, this is equal to 47,499 patient days. Subtracting one from the other gives the total number of days saved by NHS hospitals if ISP patients had been treated by NHS hospitals, which is equal to 1,846 patient days across the NHS or 14 days per year per NHS Acute Trust. Applying the cost per bed day of £231, this gives a total cost of £426,426 ($625,000). This is equivalent of to the cost of an additional 77 hip replacements or 0.1% of the total number of hip replacements that tool place in 2012/13.

In summary, these calculations suggest that the limited gradation in HRG payments means that NHS hospitals and ISPs receive the same payment for treating "simple" patients, but that ISPs face lower patient contingent per treatment costs. This would imply that ISPs have higher profit margins, which are not due to ISPs being intrinsically more efficient, and that as a consequence of sorting ISPs receive an implicit subsidy from the taxpayer. Any patient sorting that is correlated with treatment costs will have similar implications in all

[39] We check the robustness of this result by also running a model that includes the comorbidities recorded in PROMS data, which are not present in HES. The results do not change

DRG systems. These costs must be taken into account when regulating healthcare markets or designing policy, and must be weighed against the advantages of a limited number of DRG codes.

7 Conclusions

In this paper we examine mechanisms of patient sorting between private and public providers of publicly funded elective medical treatments in the English National Health Service, using elective hip replacements as the example procedure. There are at least three reasons why differential sorting may be of concern. First, inequality in access to, and uptake of, private provision is potentially important for welfare, especially when private providers are able to deliver care much faster than public providers, and where patient satisfaction and quality are arguably superior (NHS Partners Network (2015), Care Quality Commission (2012), Browne *et al.* (2008)). Second, policies to expand market access to private providers are often introduced to generate competitive pressure on public incumbents, with the aim of improving efficiency, quality and innovation. Unequal access implies the threat of patients switching provider is below its full potential, and hence public providers may be expected to experience less competitive pressure than intended by the policy reform. Finally, in a system of national prices that do not necessarily fully compensate for differences in the severity of patient illness, different patient types with different costs entail profit implications for providers, and these matter acutely when budgets and capacity are constrained.

We find that those patients who live in more deprived areas and those from ethnic minorities are less likely to choose private providers. Our model estimates and simulations show that the most important mechanisms in driving these differences are underlying health, which affects eligibility for treatment by a private provider, and the geographic distribution of providers. However, the restrictions that GPs place on the choices that they offer their patients also have a substantive role to play. The dual role of the GP as both an agent for the patient and a gatekeeper for NHS services makes these relationships harder to interpret. On the one hand, in a market where patients have imperfect information and find it difficult to assess quality, let alone to act on it (Dixon *et al.*, 2010), GP pre-selection and mandated advice improve welfare. On the other hand, given the budgetary implications that GPs' referral decisions have, their incentives may not be perfectly aligned with those of their patients, and this may bias their advice and thereby reduce welfare (Beckert and Collyer, 2017).

Our findings have several important implications for policy, both within the English NHS and in other country and policy settings. First, eligibility criteria and the geographic

distribution of providers have important implications for how consumers sort across providers. In our context, we show how health criteria and the opening up of pre-existing private hospitals to public patients led to ISPs treating a higher share of rich and white patients. However, the same mechanisms apply in other contexts. In the UK, a close paralell is the sorting of more advantaged pupils into selective secondary education or grammar schools [40]. This type of sorting is not necessarily inefficient when the set of providers and eligibility criteria is taken as given, but may still be of concern to policy makers. In terms of equity, we show that certain groups of patients have less access to ISP providers, which offer lower waiting times than NHS hospitals. This may be regarded as undesirable by politicians or the public. In terms of payments and public finances, our results illustrate that when payments to hospitals are largely fixed, the sorting of healthier patients to ISPs led to an increase in cost for NHS hospitals and limited their ability to cross-subsidise across patients. Such concerns are particularly relevant at present, when post Great Recession austerity means that budgets for the NHS and other public services are particularly tight.

Second, if choice is to be the driving force behind competition and a drive to improve quality, more work is needed to understand information imperfections and other frictions that may limit choice for all, and for some groups in particular. Our results show that referral patterns of GPs are strongly correlated with subsequent patient choices. However, whether this reflects GPs acting as a perfect agent for the patient, information imperfections, or alternative incentives unrelated to patient health is less clear. We show that the choice sets presented to patients are a function of attributes of hospitals that are unrelated to patient need. Yet the degree of sorting as the result of frictions, at least into provider type, is minor relative to the importance of patient characteristics and the geographical distribution of providers.

Finally, our results support existing work that highlights ethnic disparities in health care access (Cookson *et al.*, 2016; Dixon *et al.*, 2010; Fiscella *et al.*, 2000). We find that ethnic minority patients are less likely to use an ISP than a patient who both lives in the most deprived fifth of areas and has poor underlying health. While part of this disparity by ethnicity is accounted for by differences in health and the location of ISPs, a large fraction is unexplained. There may be some differences in health that are unobserved. However, surveys of GPs and studies of differential take up of screening suggest that there are likely to be further barriers (Fisher *et al.*, 2014; Moser *et al.*, 2009). As in many other public services, reducing ethnic disparities will require identifying these barriers and developing strategies to overcome them.

[40] These schools are disproportionately located in richer than average areas, and students meeting the test score eligibility criteria are from more advantaged background.

References

Abaluck, J. and A. Adams [2017]. "What do consumers consider before they choose? Identification from asymmetric demand responses". NBER Working Paper No.23566

Altonji, J. G., C.-I. Huang, and C. R. Taber [2015]. "Estimating the Cream Skimming Effect of School Choice". *Journal of Political Economy*, 123(2): 266–324. URL http://www.jstor.org/stable/10.1086/679497

Arora, S., A. Charlesworth, E. Kelly, and G. Stoye [2013]. "Public Pay and Private Provision: The changing landscape of healthcare in the 2000s". Nuffield Trust and Institute for Fiscal Studies Research Report

Asth, J., E. Andersson, and B. Malmberg [2013]. "School Choice and Increasing Performance Difference: A Counterfactual Approach". *Urban Studies*, 50(2): 407–425. URL http://EconPapers.repec.org/RePEc:sae:urbstu:v:50:y:2013:i:2:p:407-425

Bardsley, M. and J. Dixon [2011]. "Quality of care in independent sector treatment centres". British Medical Journal 343

Beckert, W., M. Christensen, and K. Collyer [2012]. "Choice of NHS-Funding Hospital Services in England". *Economic Journal*, 122(560): 400–417. DOI: 10.1111/j.1468-0297.2012.02496.x

Beckert, W. and K. Collyer [2016]. "Choice in the Presence of Experts: The Role of GPs in Patients' Hospital Choice". IFS Working Paper W16/21

Beckert, W. and K. Collyer [2017]. "Choice in the presence of experts: the role of general practitioners in patients' hospital choice". URL https://www.ifs.org.uk/uploads/publications/wps/WP201621.pdf

Beckert, W. and P. Schiraldi [2017]. "Demand Estimation with Externalities: Hospital Choice and the Role of Waiting Times"

Besley, T. and M. Ghatak [2003]. "Incentives, Choice, and Accountability in the Provision of Public Services". *Oxford Review of Economic Policy*, 19(2): 235–249. doi:10.1093/oxrep/19.2.235. URL http://oxrep.oxfordjournals.org/content/19/2/235.abstract

Blöchliger, H. [2008]. "Market Mechanisms in Public Service Provision". OECD Economics Department Working Papers, No. 626. Organization for Economic Cooperation and Development, Paris, France.

Böhlmark, A., H. Holmlund, and M. Lindahl [2016]. "Parental choice, neighbourhood segregation or cream skimming? An analysis of school segregation after a generalized choice reform". *Journal of Population Economics*, 29(4): 1155–1190. URL http://EconPapers.repec.org/RePEc:spr:jopoec:v:29:y:2016:i:4:d:10.1007_s00148-016-0595-y

Böhlmark, A. and M. Lindahl [2015]. "Independent School and Long run Education Outcomes: Evidence from Sweden's Large Scale Voucher Reform". *Economica*

Boyle, D. [2013]. "The Barriers to Choice Review. How are people using choice in public services?" Independent Report for the UK Cabinet Office. URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/80070/Barriers_to_choice_review_0.pdf

Browne, J., L. Jamieson, J. Lewsey, J. van der Meulen, L. Copley, and N. Black [2008]. "Case-mix & patients' reports of outcome in Independent Sector Treatment Centres: Comparison with NHS providers". *BMC Health Services Research*, 8:78. doi:doi:10.1186/1472-6963-8-78

Burgess, S., E. Greaves, A. Vignoles, and D. Wilson [2015]. "What Parents Want: School Preferences and School Choice". *The Economic Journal*, 125(587): 1262–1289. doi:10.1111/ecoj.12153. URL http://dx.doi.org/10.1111/ecoj.12153

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

Care Quality Commission [2012]. "Market Report 2012". URL http://www.cqc.org.uk/sites/default/files/documents/20120626_cqc_market_report_issue_1_for_website_final_0.pdf

Chard, J., M. Kuczawski, N. Black, and J. van der Meulen [2011]. "Outcomes of elective surgery undertaken in indepdnent sector treatment centres adn NHS providers in England: audit of patient outcomes in surgery". British Medical Journal 343

Cookson, R., M. Laudicella, and P. L. Donni [2012]. "Measuring change in health care equity using small-area administrative data – Evidence from the English NHS 2001–2008". *Social Science & Medicine*, 75(8): 1514–1522

Cookson, R., C. Propper, M. Asaria, and R. Raine [2016]. "Socio-Economic Inequalities in Health Care in England". *Fiscal Studies*, 37(3-4): 371–403. doi:10.1111/j.1475-5890.2016.12109. URL http://dx.doi.org/10.1111/j.1475-5890.2016.12109

Cooper, Z., S. Gibbons, S. Jones, and A. McGuire [2011]. "Does Hospital Competition Save Lives? Evidence From The English NHS Patient Choice Reforms". *Economic Journal*, 121(554): F228–F260. URL http://ideas.repec.org/a/ecj/econjl/v121y2011i554pf228-f260.html

Cooper, Z., S. Gibbons, and M. Skellern [2015]. "Does Competition from Private Surgical Centres Improve Public Hospital' Performance? Evidence from the English National Health Service". Mimeo, London School of Economics

Crawford, G. S., R. Griffith, and A. Iaria [2016]. "Demand estimation with unobserved choice set heterogeneity"

Deming, D., J. S. Hastings, T. Kane, and D. Staiger [2014]. "School Choice, School Quality, and Postsecondary Attainment". *American Economic Review*, 104(3): 991–1013

Dixon, A. and R. Robertson [2009]. "Choice at the Point of Referral". URL http://www.kingsfund.org.uk/publications/choice-point-referral

Dixon, A., R. Robertson, J. Appleby, P. Burge, N. Devlin, and H. Magee [2010]. "Patient choice: how patients choose and how providers respond". Kings Fund Report

Doorslaer, E. v., X. Koolman, and A. M. Jones [2004]. "Explaining income-related inequalities in doctor utilisation in Europe". *Health Economics*, 13(7): 629–647

Edmark, K., M. Frlich, and V. Wondratschek [2014]. "Sweden's school choice reform and equality of opportunity". *Labour Economics*, 30(C): 129–142. doi:10.1016/j.labeco.2014.04. URL https://ideas.repec.org/a/eee/labeco/v30y2014icp129-142.html

Fiscella, K., P. Franks, M. Gold, and C. Clancy [2000]. "Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care." *JAMA*, 283(29): 2579–2584

Fisher, H., S. Audrey, J. Mytton, M. Hickman, and C. Trotter [2014]. "Examining inequalities in the uptake of the school-based HPV vaccination programme in England: a retrospective cohort study". *Journal of Public Health*, 36(1): 36–45

Gaynor, M. [2006]. "What Do We Know About Competition and Quality in Health Care Markets?" *Foundations and Trends in Microeconomics*, 2(6): 441–508

Gaynor, M., R. Moreno-Serra, and C. Propper [2012]. "Death by Market Power: Reform, Competition and Patient Outcomes in the National Health Service". mimeo, Carnegie Mellon University. URL http://ideas.repec.org/p/nbr/nberwo/16164.html

Gaynor, M., C. Propper, and S. Seiler [2016]. "Free to Choose? Reform and Demand Response in the English National Health Service." *American Economic Review*, 106(11): 3521–3527

Goeree, M. [2008]. "Limited information and advertising in the US personal computer industry". *Econometrica*, 76(5): 1017–1074

Hajivassiliou, V. [2000]. *Simulation Based Inference in Econometrics*, chap. Some practical issues in maximum simulated likelihood., pp. 71–99. Cambridge University Press

Hastings, J. S., T. Kane, and D. Staiger [2010]. "Heterogeneous Preferences and the Efficacy of Public School Choice". NBER Working Paper 12145

Hastings, J. S. and J. M. Weinstein [2008]. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments". *The Quarterly Journal of Economics*, 123(4): 1373–1414. URL http://www.jstor.org/stable/40506212

Ho, K. [2006]. "The Welfare Effects of Restricted Hospital Choice in the US Medical Care Market". *Journal of Applied Econometrics*, 21: 1039–1079

Hoxby, C. M. [2003]. *The Economics of School Choice*, chap. School Choice and School Productivity: Could school choice be a tide that lifts all boats, pp. 287–342. National Bureau of Economic Research, Cambridge MA

Judge, A., N. Welton, J. Sandhu, and Y. Ben-Shlomo [2010]. "Equity in access to total joint replacement of the hip and knee in England: cross sectional study". *BMJ*, 341

Kamenica, E. [2008]. "Contextual Inference in Markets: On the InInformation Content of Product Lines". *American Economic Review*, 98(5): 2127–2149

Kelly, E. and G. Stoye [2016]. "New Joints: Private providers and rising demand in the English National Health Service". IFS Working Paper (W16/15). doi:10.1920/wp.ifs.2016.1615. URL https://www.ifs.org.uk/uploads/publications/wps/WP201615.pdf

Kessler, D. P. and M. B. McClellan [2000]. "Is Hospital Competition Socially Wasteful?" *Quarterly Journal of Economics*, 115(2): 577–615

Ladd, H. F. [2002]. "School Vouchers: A Critical View". *Journal of Economic Perspectives*, 16(4): 3–24. doi:10.1257/089533002320950957. URL http://www.aeaweb.org/articles.php?doi=10.1257/089533002320950957

Langenbrunner, J. C., S. O'Duagherty, and C. S. Cashin [2009]. *Designing and Implementing Health Care Provider Payment Systems*. The World Bank. doi:10.1596/978-0-8213-7815-1. URL https://elibrary.worldbank.org/doi/abs/10.1596/978-0-8213-7815-1

McFadden, D. L. and K. Train [2000]. "Journal of Applied Economics". *Mixed MNL Models for Discrete Response*, pp. 447–470

Monitor [2015]. "Choice in Adult Hearing Services: Exploring How Choice is Working for Patients". URL https://www.gov.uk/government/

Morris, S., M. Sutton, and H. Gravelle [2005]. "Inequity and inequality in the use of health care in England: an empirical investigation". *Social Science & Medicine*, 60(6): 1251–1266. URL http://ideas.repec.org/a/eee/socmed/v60y2005i6p1251-1266.html

Moscelli, G., L. Siciliani, N. Gutacker, and R. Cookson [2015]. "Socioeconomic inequality of access to healthcare: Does patients' choice explain the gradient? Evidence from the English NHS". Centre for Health Economics Working Paper 112

Moser, K., J. Patnick, and V. Beral [2009]. "Inequalities in reported use of breast and cervical screening in Great Britain: analysis of cross sectional survey data?" *British Medical Journal*, 339: b2025

Naylor, C. and S. Gregory [2009]. "Briefing{:} Independent Sector Treatment Centres". Kings Fund http://www.kingsfund.org.uk/publications/briefing-independent-sector-treatment-centres

NHS England [2013]. "PbR tariff information spreadsheet for 2012 to 2013". available at https://www.gov.uk/government/publications/confirmation-of-payment-by-results-pbr-arrangements-for-2012-13

NHS Partners Network [2015]. "Independent sector providers caring for NHS patients". URL http://www.nhsconfed.org/~/media/Confederation/Files/public%20access/NHSPN%20headline%20indicator%20summary%20-%20June%2016.pdf

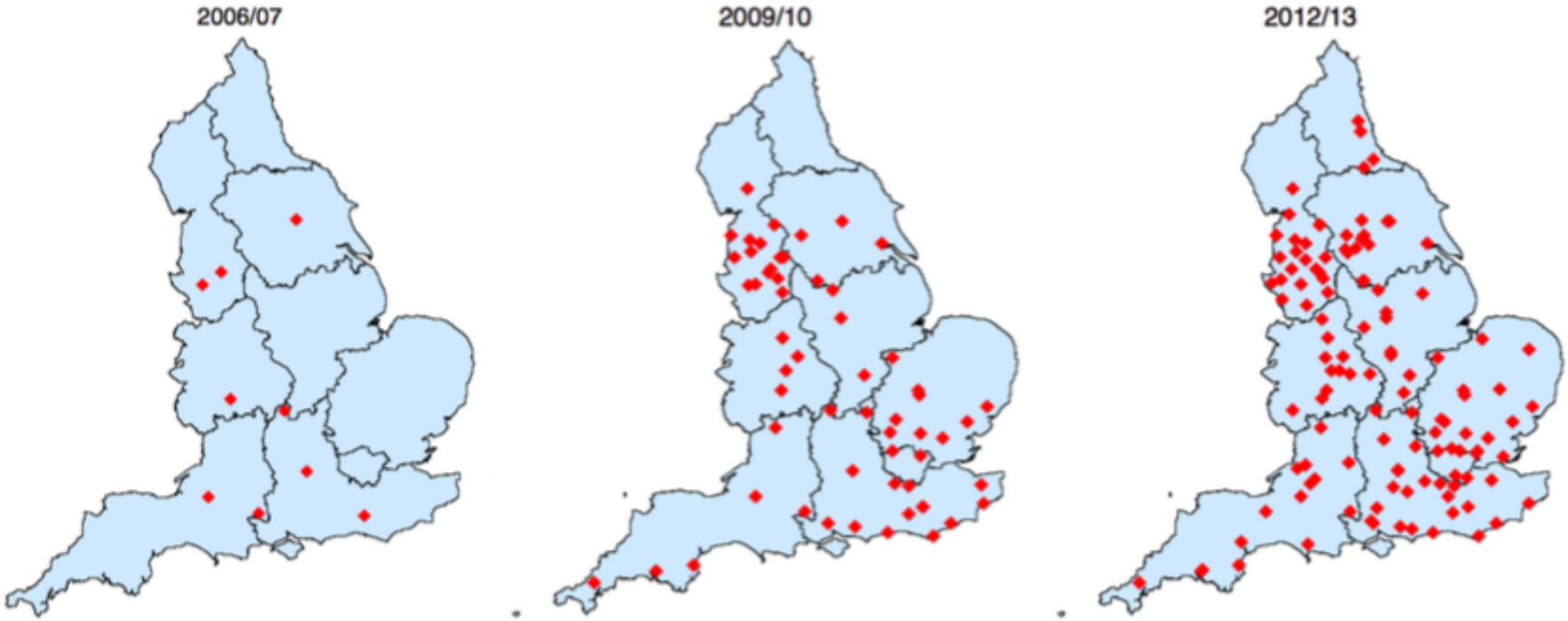
O'Donnell, O. and C. Propper [1991]. "Equity and the Distribution of UK National Health Service - Reply". *Journal of Health Economics*, 10: 247–249. doi:10.1016/0167-6296(91)90008-B

Sundararajan, V., T. Henderson, C. Perry, A. Muggivan, H. Quan, and W. A. Ghali [2004]. "New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality". *Journal of Clinical Epidemiology*, 57(12): 1288 – 1294. doi:http://dx.doi.org/10.1016/

j.jclinepi.2004.03.012. URL http://www.sciencedirect.com/science/article/pii/S0895435604001647

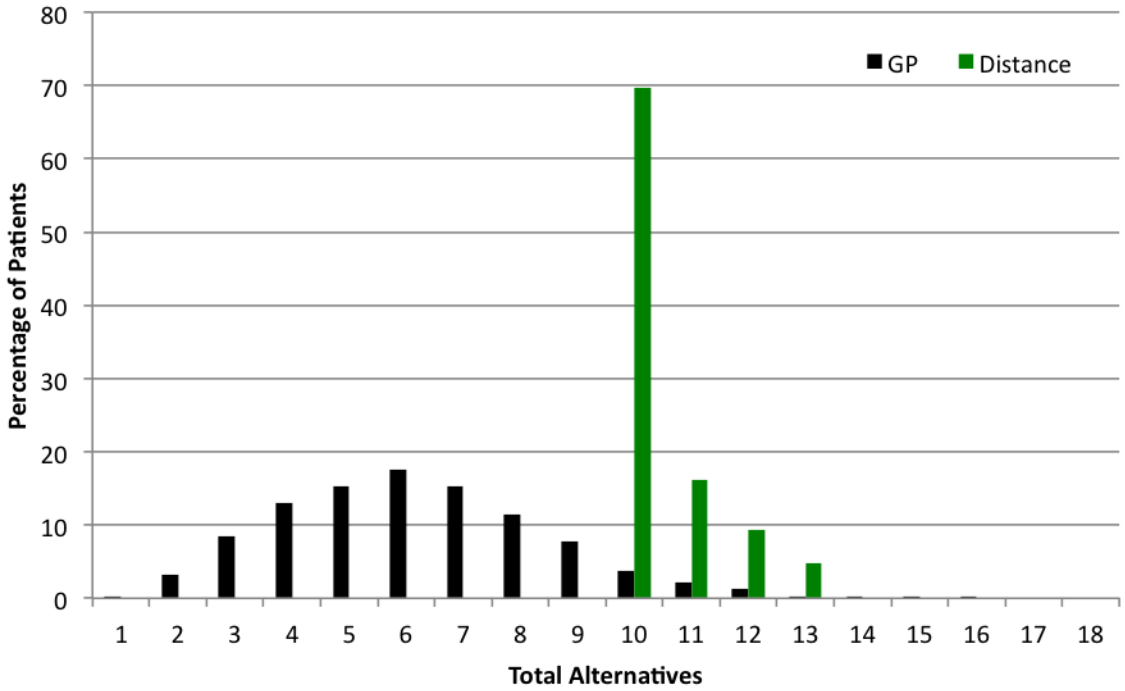
Urquiola, M. [2005]. "Does School Choice Lead to Sorting? Evidence from Tiebout Variation". *The American Economic Review*, 95(4): 1310–1326. URL http://www.jstor.org/stable/4132718

Figure 1: **The spread of ISPs across England (2006/07, 2008/09 and 2010/11)**



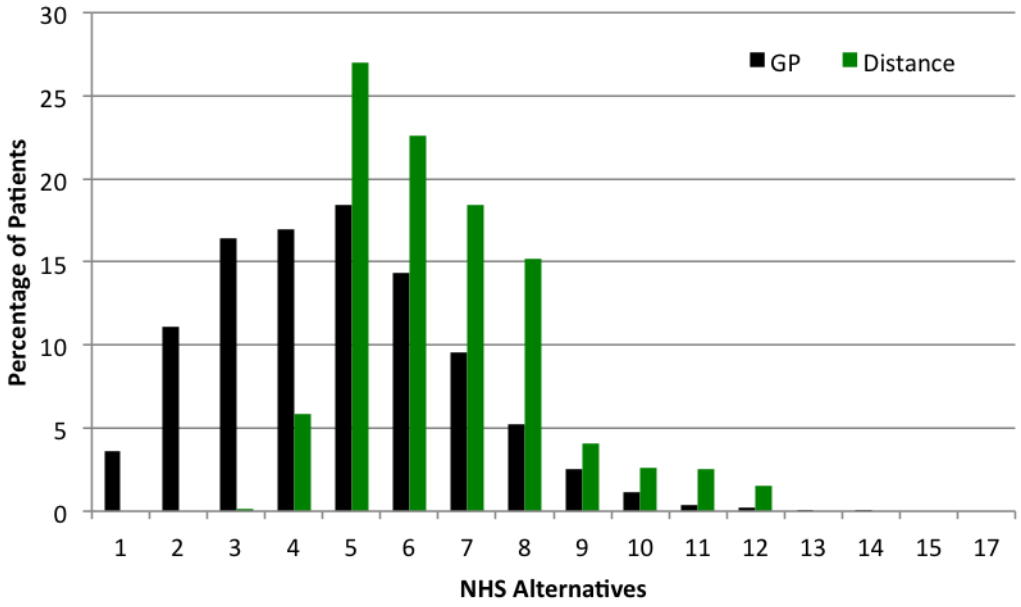
Notes Sample is restricted to those Independent Sector Providers that are recorded as conducting at least 20 NHS-funded elective hip replacements in the NHS Hospital Episode Statistics in the given

Figure 2: **The distribution of the number alternatives in patient choice sets**



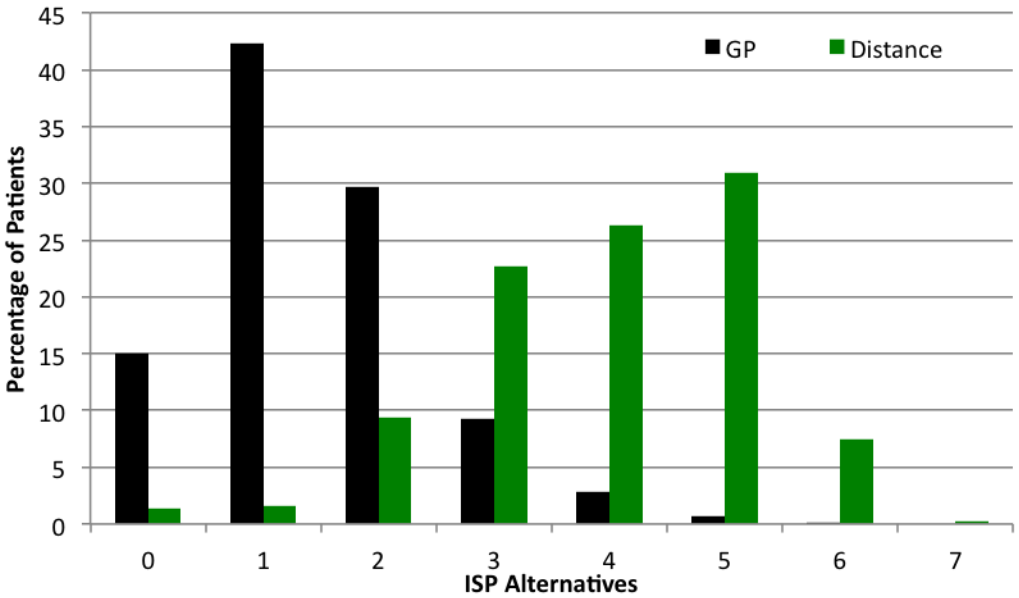
Notes Sample is restricted to those hospitals that are recorded as conducting at least 20 NHS-funded elective hip replacements in the NHS Hospital Episode Statistics in the given

Figure 3: **The distribution of the number NHS alternatives in patient choice sets**



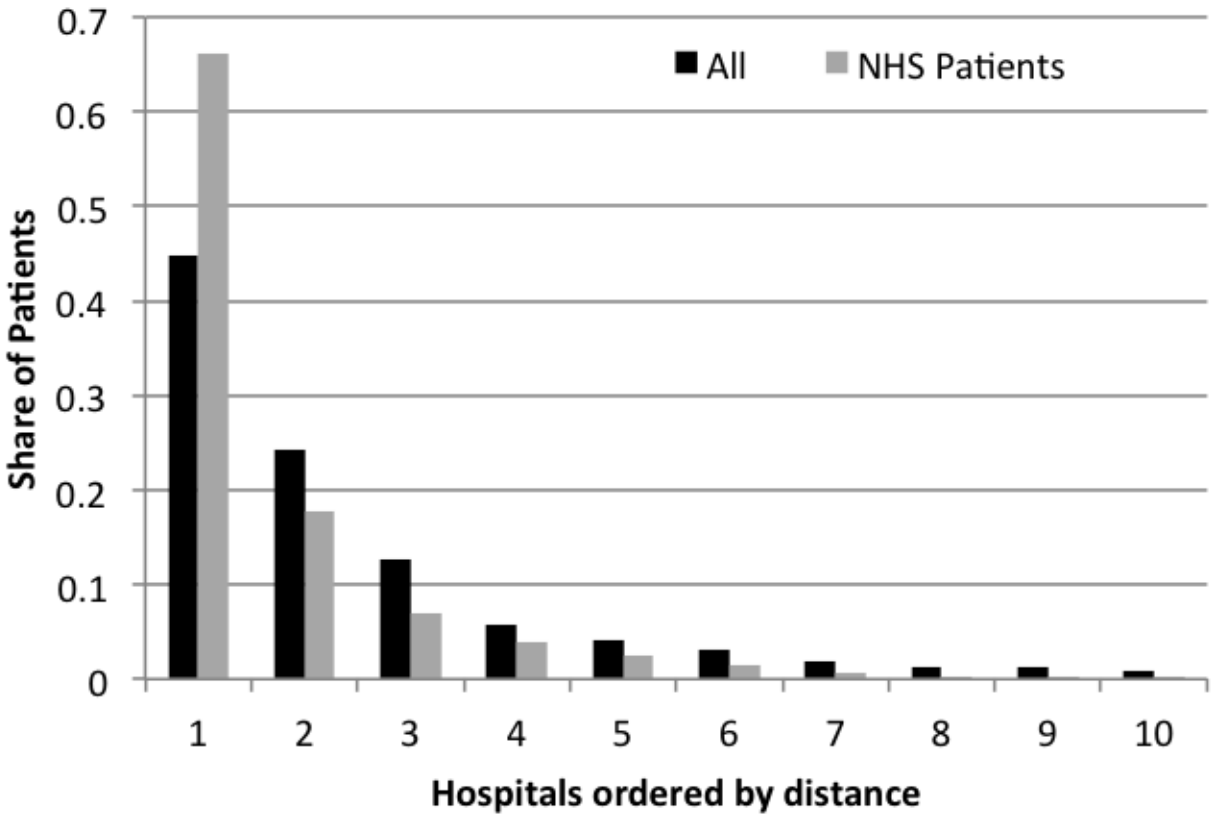
Notes Sample is restricted to those NHS hospitals that are recorded as conducting at least 20 NHS-funded elective hip replacements in the NHS Hospital Episode Statistics in the given

Figure 4: **The distribution of the number ISP alternatives in patient choice sets**



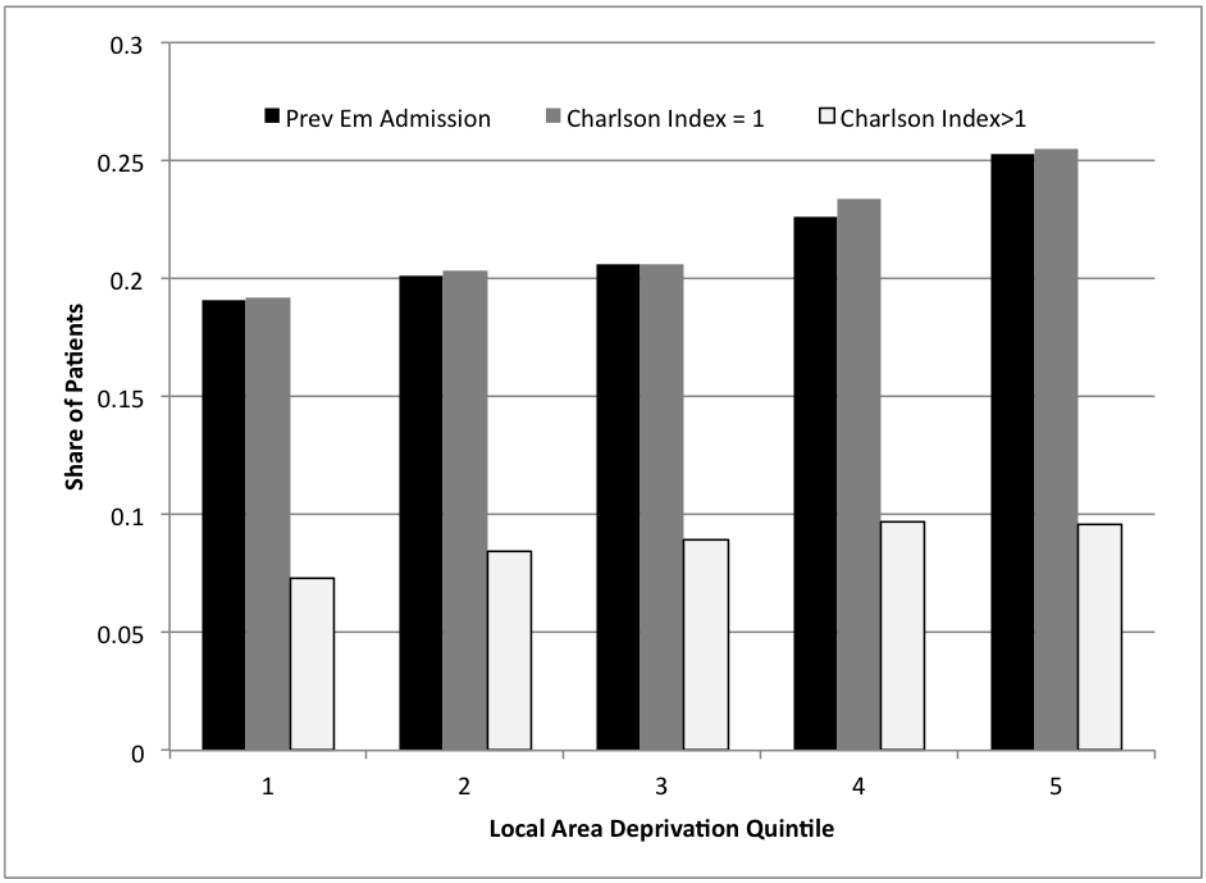
Notes Sample is restricted to those Independent Sector Providers that are recorded as conducting at least 20 NHS-funded elective hip replacements in the NHS Hospital Episode Statistics in the given

Figure 5: **Distribution of Patient Choices, Hospitals Ordered by Distance**



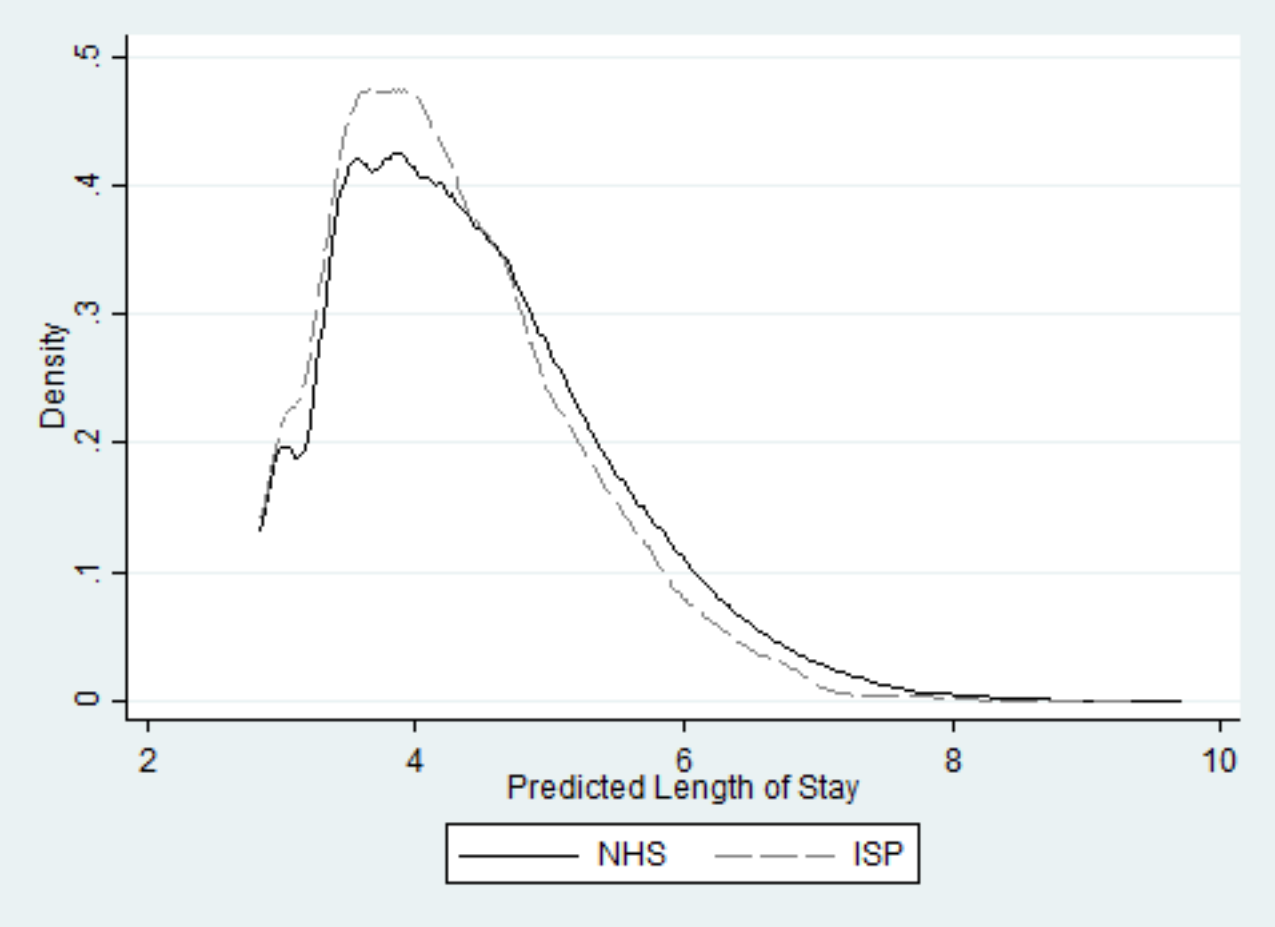
Notes: Distances are calculated in a straight line between the centroid of the patient's local area (LSOA) and the postcode of the hospital. The grey bars exclude patients that chose treatment at ISPs.

Figure 6: **Share of Patients with Underlying Health Problems by Local Area Deprivation Quintiles.)**



Notes: Quintile 1 corresponds the least deprived fifth of areas on the basis of IMD. Quintile 5 corresponds the most deprived quintile of areas. Local area deprivation measures by Index of Multiple Deprivation rank of the patient's lower super output area

Figure 7: **The Distribution of Predicted Length of Stay for Hip Replacement Patients, by Provider Type**



Notes: Kernel density function

Table 1: **ISP Types**

	Wave 1	Wave 2	
	Wave 1 ISTC	Wave 2 ISTC	AQP
Year commenced	2003	2006	2006
Year last contract ended	2011	2015	-
Payment	Contracted procedure numbers, full payment guaranteed	Payments for fixed costs guaranteed, per procedure thereafter.	Per procedure
Patients	NHS patients only	NHS patients only	NHS patients and private pay patients

Table 2: **Means of Hospital Attributes, by Provider Type**

	NHS	ISP	All
Attributes with RC			
30 Day Em Readmit Ratio (2012)	0.91	1.09	1.02
(1=expected)	(0.42)	(0.64)	(0.53)
Attributes without RC			
Median Waiting Time)	87	45.3	71.2
(days, 2012/13	(23.3)	(33.2)	(34.1)
Share Early Found Trusts	0.160	N/A	0.10
	(0.37)	N/A	(0.30)
Share Specialist Hosps	0.0205	0.008	0.016
	(0.142)	(0.091)	(0.125)
Patients	253.3	103.4	196.5
	(174.0)	(96.0)	(166.0)
Hospitals	195	119	314

Notes: The sample includes hospitals that treated at least 20 patients in 2012/13. Waiting times are measured from the date of the decision to admit for a procedure and the date of the admission for the procedure. The 30 day emergency admission rate is given by predicting readmissions based on a regression of readmissions on the age, sex and underlying health of patients and dividing the actual readmissions, by the expected number. An ISP is defined as a hospital site that has a code in HES that begins with an "N".

Table 3: **Mean patient characteristics by chosen provider type**

	ISP	**NHS**	**Difference**
Age	68.2	68.6	-0.4***
	(10.0)	(11.6)	(0.1)
Ethnic Minority	0.013	0.039	-0.025***
	(0.115)	(0.193)	(0.002)
Female	0.598	0.601	-0.002
	(0.49)	(0.49)	(0.005)
Local Area Deprivation	0.391	0.45	-0.058***
(Scaled 0-1)	(0.253)	(0.275)	(0.003)
Moderate Comorbidity	0.167	0.225	-0.058***
(CI=1)	(0.373)	(0.418)	(0.004)
Severe Comorbidity	0.044	0.096	-0.052***
(CI>1)	(0.205)	(0.295)	(0.003)
Prev Emergency Admission	0.132	0.230	-0.098***
	(0.338)	(0.421)	(0.004)
Prev Elective Admission	0.481	0.568	-0.088***
	(0.500)	(0.495)	(0.005)
GP ref HHI (2011)	0.548	0.607	-0.059***
	(0.178)	(0.197)	(0.002)
GP 3 year ISP ref share	0.13	0.077	0.054***
(2009/10-2011/12)	(0.106)	(0.086)	(0.001)
N	12,357	50,525	

Notes: Local area deprivation is measured using the (inverse) rank of the patient's Lower Super Output Area's Index of Multiple Deprivation in 2001. This measure is then rescaled to fall between zero and one. Ethnic minority are those not classed as White British or Irish. Comorbidities measured using the Charlson Index calculated using the information in the hip replacement admission. Previous admissions in the previous 3 years (1095 days) for any cause. Sample includes patients that had an elective hip replacement in 2012/13 and were treated by a hospital that treated at least 19 other patients and that was in the closest 10 providers from the centroid of the patient's LSOA or a specialist hospital within 50km.

Table 4: **Mean patient characteristics by chosen provider type**

	2006/07		2007/08		2008/09		2009/10		2010/11		2011/12		2012/13	
	% of Patients		% of Patients		% of Patients		% of Patients		% of Patients		% of Patients		% of Patients	
	ISP Close	Use ISP	ISP Close	Use ISP	ISP Close	Use ISP	ISP Close	Use ISP	ISP Close	Use ISP	ISP Close	Use ISP	ISP Close	Use ISP
Deprivation Quintile														
1 (least deprived)	0.034	0.037	0.069	0.077	0.174	0.116	0.177	0.138	0.342	0.204	0.327	0.213	0.371	0.228
2	0.027	0.040	0.064	0.077	0.168	0.114	0.180	0.137	0.304	0.196	0.317	0.205	0.357	0.222
3	0.023	0.038	0.053	0.072	0.159	0.116	0.152	0.130	0.282	0.183	0.296	0.190	0.324	0.209
4	0.013	0.028	0.045	0.071	0.145	0.104	0.125	0.102	0.250	0.146	0.267	0.159	0.297	0.169
5 (Most Deprived)	0.013	0.031	0.028	0.057	0.116	0.085	0.103	0.091	0.214	0.122	0.228	0.120	0.221	0.127
All	0.023	0.035	0.054	0.072	0.156	0.109	0.153	0.124	0.287	0.177	0.294	0.185	0.326	0.200

The sample of hospitals includes those that treat at least 20 hip replacements in the given financial year. The sample of patients includes those who were treated by these hospitals, and who chose a hospital that was among their closest 10 (or a specialist hospital within 50km). ISP close is defined as having an ISP (conducting at least 20 hip replacements) closer than the nearest NHS hospital. Distances are measured in straight lines from the centroid of the patients LSOA to the full postcode of the hospital.

Table 5: **Mixed Logit Results: Hospital Attributes**

	Distance Choice Set			GP Choice Set		
	Coeff	SE	p-value	Coeff	SE	p-value
Distance						
Mean	-0.0895	0.0079	0.000	-0.0765	0.0065	0.000
SD	0.1187	0.0017	0.000	-0.0669	0.0014	0.000
ISP						
Mean	-1.6935	0.2076	0.000	-1.5723	0.2081	0.000
SD	2.9803	0.0907	0.000	2.8348	0.0899	0.000
Emergency Readmissions						
Mean	-1.3277	0.1249	0.000	-0.9999	0.132	0.000
SD	0.0404	0.0472	0.392	0.0852	0.1662	0.608
Attributes w/out RC						
Early Foundation Trust	0.8323	0.1484	0.000	0.4033	0.1462	0.000
Waiting times (weeks)	-0.0811	0.0143	0.000	-0.0778	0.0146	0.000
Specialist Orthopedic Hosp	1.4939	0.2044	0.000	3.0488	0.1713	0.000

Notes: The sample includes all patients who had an elective hip replacement in financial year 2012/13 and chose one of the ten closest hospitals to the centroid of their LSOA. The model also includes interactions between all hospital characteristics and age, ethnic minority status, underlying health, and prior GP referral patterns.

Table 6: **Mixed Logit Results: Heterogeneity by patient characteristics**

	Distance Choice Set			GP Choice Set		
	Coeff	SE	p-value	Coeff	SE	p-value
ISPs						
ISP x Age	-0.0026	0.0023	0.000	-0.0033	0.0023	0.000
ISP x Ethnic Minority	-1.1849	0.1743	0.000	-1.1681	0.1807	0.000
ISP x Deprivation (richest)	0.4081	0.2157	0.059	0.1036	0.2151	0.631
ISP x Deprivation (poorest half)	-0.1297	0.1047	0.216	-0.224	0.1046	0.032
ISP x Prev Em Admit	-1.1102	0.0701	0.000	-1.0189	0.069	0.000
ISP x Prev El Admit	-0.5097	0.0519	0.000	-0.5734	0.0523	0.000
ISP x CI of 1	-0.7007	0.0657	0.000	-0.7381	0.0656	0.000
ISP x CI of 2+	-1.4143	0.1067	0.000	-1.4627	0.106	0.000
ISP x GP HHI	-1.3018	0.1505	0.000	-0.5086	0.149	0.000
ISP x GP refs to ISPs	3.3969	0.3381	0.000	3.1574	0.3277	0.000
Specialist Hospitals						
Spec Hosp x Ethnic Minority	0.4665	0.1291	0.000	0.2468	0.1128	0.029
Spec Hosp x Deprivation (richest half)	0.7834	0.2369	0.025	0.7891	0.1943	0.000
Spec Hosp x Deprivation (poorest half)	0.2446	0.1092	0.215	0.0346	0.0901	0.701
Spec Hosp x Pre Em Admit	-0.0682	0.0549	0.119	-0.2104	0.0462	0.000

Notes: The sample includes all patients who had an elective hip replacement in financial year 2012/13 and chose one of the ten closest hospitals to the centroid of their LSOA. The model also includes interactions all hospital characteristics in Table 5 and age, ethnic minority status, underlying health, and prior GP referral patterns. Random coefficients are estimated at the patient level.

Table 7: **Mixed Logit Results: Alternative Choice Set Definition**

	Coeff	SE	p-value
Attributes with RC			
Distance			
Mean	-0.137	0.011	0.000
SD	0.133	0.003	0.000
ISP			
Mean	-2.981	0.265	0.000
SD	3.745	0.127	0.000
Emergency Readmissions			
Mean	-0.911	0.142	0.000
SD	-0.019	0.062	0.377
Attributes w/out RC			
Early Foundation Trust	0.849	0.167	0.000
Waiting times (weeks)	-0.119	0.016	0.000
Specialist Orthopedic Hosp	2.000	0.232	0.000
ISPs			
ISP x Age	-0.003	0.003	0.142
ISP x Ethnic Minority	-1.904	0.222	0.000
ISP x Deprivation (richest)	0.364	0.418	0.802
ISP x Deprivation (poorest half)	-0.9292	0.4351	0.016
ISP x Prev Em Admit	-1.336	0.088	0.000
ISP x Prev El Admit	-0.610	0.064	0.000
ISP x CI of 1	-0.864	0.081	0.000
ISP x CI of 2+	-1.698	0.133	0.000
ISP x GP HHI	-0.836	0.182	0.000
ISP x GP refs to ISPs	7.621	0.478	0.000
Spec Hosp x Ethnic Minority	0.220	0.139	0.057
Spec Hosp x Deprivation (richest)	0.230	0.415	0.290
Spec Hosp x Deprivation (poorest half)	0.190	0.418	0.325
Spec Hosp x Pre Em Admit	-0.056	0.061	0.179

Notes: Choice set definition - up to the 30 nearest hospitals within 30km relative to the centroid of the patient's LSOA. The sample includes all patients who had an elective hip replacement in financial year 2012/13. Random coefficients are estimated at the patient level.

Table 8: **Simulation Results: Expected ISP Volumes by Deprivation quintile and ethnicity**

	Data			Simulations				
	Patients	Share ISP	Model	a) Health	b) Ethnicity	c) Prefs	d) GP effects	e) DCS
Deprivation								
1 (least)	15,176	0.222	0.264	0.299	0.298	0.298	0.305	0.422
2	15,106	0.217	0.260	0.293	0.293	0.295	0.301	0.420
3	13,830	0.205	0.251	0.284	0.284	0.287	0.293	0.416
4	10,634	0.162	0.227	0.265	0.264	0.269	0.276	0.402
5 (most)	7,949	0.119	0.207	0.252	0.251	0.257	0.263	0.387
Ethnicity								
Majority	60,573	0.198	0.250	0.284		0.287	0.293	0.414
Minority	2,122	0.075	0.154	0.242		0.244	0.229	0.372

The first two columns give the total number of hip replacements and the share of hip replacements conducted by ISPs by the local area deprivation quintile of the patient's lower super output area. The third column gives the mean predicted probability that a patient chooses an ISP using the GP choice set model. These predicted probabilities are calculated by summing the predicted probabilities for ISP alternatives for each patient. The final five columns consider successive, cumulative simulations. The fourth column gives all patients the mean underlying health, column 5 equalises ethnicity, column 6 equalises preferences by removing interaction between deprivation and hospital attributes (top panel) and ethnic minority status and hospital attributes (bottom panel).

Table 9: **Percentage of Patients in each deprivation quintile, by NHS hospital and ISP type**

	NHS Hospital	ISP		
Deprivation quintile		Private Hospital	Phase 1 ISTC	Phase 2 ISTC
1 (richest)	23.15	28.55	26.87	26.95
2	23.69	27.19	30.55	23.6
3	21.75	22.21	23.31	22.78
4	17.72	14.75	13.79	14.57
5 (poorest)	13.7	7.3	5.49	12.11
% of patients	100	100	100	100
Patients	50709	9220	1712	1462

Sample includes hip replacement patients that choose a hospital within their closest 10 providers conducting more than 50 hip replacements. Deprivation quintile given by the Index of Multiple Deprivaion (IMD) rank of the patient's LSOA.

Table 10: **Percentage of Patients in each deprivation quintile, by NHS hospital and ISP type**

HRG Groups	Total Patients	Tariff, £($)	Share prev em admits	Share ISP
HB12C: Major Hip Procedures for non Trauma Category 1 without CC	47,280	5,382 (6,728)	0.170	0.222
HB12A: Major Hip Procedures for non Trauma Category 1 with Major CC	3,100	8,305 (10,381)	0.345	0.072
HB12B: Major Hip Procedures for non Trauma Category 1 with CC	3,187	6,021 (7,526)	0.340	0.091
HB11C: Major Hip Procedures for non Trauma Category 2 without CC	1,491	6,579 (8,224)	0.370	0.190
Other	7,654	-	0.324	0.089
Total	62,712		0.211	0.191

Source: NHS England 2012-13 tariff information spreadsheet

A GP Choice Set Determinants

There are at least three potential mechanisms that may mean that patients are not presented with all the hospitals present in the standard distance choice set.

First, GPs may act as a patient surrogate, i.e. as an altruistic agent who presents patients only with the highest ranked alternatives. A GP might therefore exclude hospitals that are far away and of low quality. In a full information setting, in principle the GP could choose on behalf of the patient, and a mandate to offer choice would be unnecessary.

Second, information on providers is often costly to acquire and to disseminate. The costs of information acquisition mean that the patient is likely to defer to the GP in terms of choice alternatives to consider, but also imply that GPs may not acquire knowledge about all providers. This is supported by results from GP surveys which indicate that GPs rely on "soft" knowledge from previous experience and referrals, rather than comparing clinical indicators (Dixon *et al.*, 2010). Incomplete information on the part of the GP may be particularly relevant for the inclusion of ISPs, as the providers are new and GPs will have less information based on previous referrals. The cost of communicating and disseminating information about choice options to patients is costly both to GPs themselves and for patients, where large choice sets may complicate the choice problem (see, for example, Kamenica (2008) on the tyranny of choice and choice overload). As a result, GPs may limit the number of hospital alternatives they present to patients to a small number, either because (i) GPs do not have an incentive to acquire information about further hospitals or (ii) some hospitals that the GP does have information about are withheld [41]. The resulting narrow choice set may exclude hospital alternatives that patients would rank highly if they had perfect information. This pre-selection is potentially efficient, conditional the costs of information acquisition and dissemination, because it saves patients the cost of collecting the necessary information themselves. The question is then whether there is a way of reducing these information costs to overcome the market friction. Efficiency also hinges on the incentives of GP and patient being aligned.

Finally, if GPs face incentives that are not aligned with those of the patient, then such pre-selection on the part of the GP may be distortive. It comprises situations in which GPs are uninformed about, or unresponsive to, evaluation criteria relevant to patients; and situations in which GPs face idiosyncratic incentives that patients are unaware of. For

[41] These assumptions are consistent with evidence (Dixon *et al.*, 2010; Monitor, 2015) that, the choice mandate notwithstanding, the majority of patients gets to discuss no more than five options with the GP and that GPs feel that they operate under resource constraints that do not permit them to discuss more options while seeing the same number of patients. Such resource constraints suggest that GPs decide on a relatively tightly delineated, standardized set of alternatives that they discuss with their patients

example, the contracts granted to Wave 1 ISTCs, which compensate providers for a fixed number of procedures, irrespective of whether those procedures were conducted, provided GPs with an incentive to refer to those providers; patients would not know or care about the underlying financial arrangements.

A.1 Determinant of inclusion in GP choice sets

The GP choice sets used to estimate the GP choice set model are determined by referral patterns over the previous three years (2009/10 - 2011/12). We include hospitals where GP practices referred more than 0.5% of patients. The mean number of referrals was 420 per GP practice. We then add any hospital that a GP referred a hip replacement patient to in that year, if these are no already included in the sample.

This section estimates a logit model for whether each of the 313 hospitals that conducted at least 20 procedures were included in GP choice set. Results are presented in Table A1. Column 1 shows the odds that an NHS hospital is in a GP choice set given the characteristics of the hospital. As expected, the odds of inclusion decline with distance, increase in hospital quality and decline in waiting times. This fits with the model of GPs acting as altruistic agents for their patients. However, the role of quality is relatively weak. This is consistent with responses to a survey of Providers in 2008 and 2009, where it was perceived that GP referral patterns paid little attention to quality (Dixon *et al.*, 2010). Specialist Orthopedic hospitals are 54 times more likely to be included in GP choice sets, holding other characteristics constant, and hospitals located in the same Primary Care Trust (PCT) are 10 times more likely to be included. NHS acute hospitals are separate from PCTs, however this within PCT effect may be explained by differences in information on GPs wishing to ensure that hospitals where their patients receive emergency treatment continue to receive a stream of funding.

Coumn 2 of Table A1 shows the factors that determine whether an ISP is included in a GP choice set are similar to those for an NHS hospital. Factors of importance to patients, such as distance, waiting times, and clinical quality do affect referrals, although the effect of quality is relatively small. The type of ISP is a very strong determinant. The odds of including an ISP are 6.4 times higher for Wave 1 ISTCs and 3.3 times higher for Wave 2 ISTCs, relative to hospitals that also treat private patients. As all ISPs place similar restrictions on the types of patients that are eligible for ISP treatment, this must operate either through differences in information on incentives to refer. In particular, Wave 1 ISTCs received payments for contracted procedures whether or not undertaken. Primary Care Trusts therefore had an incentive to ask GPs to refer. In common with NHS hospitals, ISPs are more likely to be

included in choice sets if they are located in the same PCT, although the magnitude of this difference is lower (with odds 3 times higher). There are no clear incentives for GPs or Primary Care Trusts for favouring ISPs within the same PCT, suggesting that this is likely to reflect imperfect information rather than incentives.

Columns 3 and 4 add the characteristics of the local area including the share of the MSOA population that is non white, the share over 75 and local area deprivation. These coefficients capture whether these characteristics affect the average number of alternatives of each provider type that are included. Local area deprivation does not have an impact on the size of the choice set for either NHS or ISP hospitals. The probability of an alternative being included, conditional on hospital characteristics, is increasing in the share of the population over 75 and decreasing in population density. For both types of provider, the probability any given provider is included is decreasing in the non-white share of the population. The magnitude of this effect is larger for ISPs than for NHS hospitals. This is consistent with results of the Dixon *et al.* (2010) survey, where GPs noted that language difficulties may limit the extent to which ethnic minority patients can participate in choice.

Columns 5 and 6 include characteristics of nearest NHS Trust and the local health economy. This is because the incentives of GPs and patients to invest in learning about new providers is likely to depend on the quality and quantity of NHS hospitals in the local area. For both NHS and ISP providers, the probability that a hospital is included in a GP's choice set is decreasing in the number of NHS hospitals within 15km. This effect is greater for ISPs than NHS hospitals. For both types of providers, the odds of choice set inclusion is increasing in the waiting times of the nearest NHS hospital and lower in areas where the nearest NHS hospital is of higher quality (as measured by early Foundation Trust status).

Finally, columns 7 and 8 include characteristics of the GP practice. For both types of provider, the odds that a hospital is included in the choice set is increasing in the practice list size. This effect is stronger for ISPs than NHS hospitals. Part of this effect is likely to come from the higher number of overall referrals and therefore the greater chance that we will observe a referral in the data. For NHS hospitals, the odds of inclusion does not depend on GP practice size. However, GP practices with more than 2 GPs are more likely to include any given ISP in their choice set than single GP practices. For both provider types, GP practices with a higher fraction of GPs from outside the EEA and a higher fraction of GPs under 40 are less likely to refer to any given provider. This is consistent with these types of GPs having less knowledge and experience of the local health economy, and being less efficient in acquiring information. However, it should also be noted that there may be unobservable characteristics of practice lists that are correlated the GP characteristics.

Taken together, the estimates in Table A1 are consistent with GP's acting as altruistic

agents for their patients. In particular, whether a hospital is included in a GP's choice set depends on hospital characteristics and the composition of the practice list. There is also strong evidence that whether a hospital, and particularly, an ISP is included in the choice set depends on the existing health economy. This may be efficient given the costs of acquiring information and providing choices to patients. However, it is also important to note that whether a hospital is included is also influenced by factors other than patient health, in particular whether the hospital is in the same PCT and the particular type of ISP contract. This illustrates that GP's do respond to incentives or pressures.

A.2 GP Level Random Coefficients

The estimated models using the distance and GP choice set definitions have both assumed that random coefficients operate at the patient level. Given the likely role of the GP in forming choice sets and offering advice, it is possible that unobserved variation in preferences is not attributable to the patient, but instead to the GP. We therefore re-estimate the GP choice set model with random coefficients at the GP level. For computational reasons we do so by drawing a random sample of 1000 GP practices and include all their hip replacement patients in the estimation. The model amounts to re-interpreting the choice outcomes as those that a GP might arrive at when deciding on behalf of each of the GP's. This acts as a robustness test to our preferred GP choice set specification, to ensure that our main results are not driven by the relative importance of GPs and patients in the decision-making process.

Parameter estimates from the GP level model are presented in Table A2. In terms of valuations of hospital attribute, introducing GP-level random coefficients has the greatest impact on the mean and standard deviation parameters for an ISP provider and emergency readmissions. The negative parameter estimate for ISP use is smaller in absolute value than for the patient level model, in absolute terms and relative to the distance parameter. However, the smaller sample size means that the mean valuation is not statistically significant. The extent of the unobserved variation is also smaller and remains statistically significant. This suggests that GPs experience less heterogeneity in valuations of ISPs. For emergency readmissions, the estimated parameter remains negative, but as with the ISP indicator, is not statistically significant. However, the random coefficient goes from very small and not statistically significant in the patient level models to sizeable and statistically significant when random coefficients are estimated at the GP level. Estimates of the remaining coefficients, notably on the various interactions, are broadly similar to those of our preferred specifications but typically closer to zero.

Differences between the patient-level and GP-level models are relatively small. However, interpreting those differences is not straightforward. The GP model can be viewed as a version of a choice model that blends the patient's and GP's contributions to the choice outcome. Take for example, the estimated random coefficient for 30 day readmissions, where the two models differ most. The statistically significant random coefficient in the GP-level model could arise from the GP observing patient characteristics that the econometrician does not observe and that lead the GP to choose a hospital for the patient that excels along other dimensions relative to quality.

We include the results from this model because they demonstrate the robustness of our headline results to modelling assumptions. We caution, however, against attempts to directly interpret the results from this model.

Table A1: **The odds that a hospital is included in a GP choice set**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Hosp chars		+ loc area		+GP practice		+ local healthcare	
	NHS	ISP	NHS	ISP	NHS	ISP	NHS	ISP
Log Distance	0.0529***	0.0489***	0.0499***	0.0332***	0.0398***	0.0236***	0.0397***	0.0231***
	(0.00108)	(0.00135)	(0.00108)	(0.00114)	(0.000957)	(0.000919)	(0.000954)	(0.000904)
Waiting Times 2012	0.999**	1.003***	0.999*	0.994***	1.000	0.992***	1.000	0.992***
	(0.000363)	(0.000539)	(0.000361)	(0.000708)	(0.000360)	(0.000752)	(0.000360)	(0.000751)
Same PCT	10.03***	3.374***	8.755***	3.292***	6.901***	2.563***	6.880***	2.587***
	(0.445)	(0.167)	(0.382)	(0.170)	(0.283)	(0.127)	(0.282)	(0.129)
ISP type (relative to private hospitals)								
Phase 1 ISTC		6.397***		7.583***		8.061***		8.149***
		(0.301)		(0.402)		(0.438)		(0.444)
Phase 2 ISTC		3.347***		3.212***		3.033***		3.079***
		(0.183)		(0.211)		(0.210)		(0.215)
Early Foundation Trust	0.993		0.990		1.006		1.006	
	(0.0265)		(0.0263)		(0.0269)		(0.0269)	
Readmission Ratio	0.749***	0.911***	0.748***	1.011	0.755***	0.971	0.755***	0.968
	(0.0160)	(0.0205)	(0.0159)	(0.0315)	(0.0162)	(0.0331)	(0.0162)	(0.0330)
Specialist Orthopedic Hospital	53.69***		56.46***	183.9***	65.86***	297.5***	66.04***	310.8***
	(1.720)		(1.875)	(14.63)	(2.333)	(25.85)	(2.344)	(27.02)
Local Area Deprivation			0.940	1.081	1.013	1.085	1.039	1.190**
			(0.0534)	(0.0946)	(0.0538)	(0.0928)	(0.0552)	(0.102)
% pop non white			0.548***	0.159***	0.765***	0.242***	0.771***	0.249***
			(0.0513)	(0.0357)	(0.0650)	(0.0463)	(0.0657)	(0.0477)
% pop over 75			1.031***	1.092***	1.026***	1.089***	1.025***	1.090***
			(0.0103)	(0.0161)	(0.00935)	(0.0143)	(0.00933)	(0.0131)
Population Density			0.993***	0.987***	0.995***	0.994***	0.995***	0.993***
			(0.000633)	(0.00139)	(0.000606)	(0.00126)	(0.000611)	(0.00126)
No of Trusts ¡15km (relative to 0)								
1					0.555***	0.404***	0.556***	0.406***
					(0.0209)	(0.0215)	(0.0209)	(0.0215)
2					0.287***	0.171***	0.290***	0.176***
					(0.0126)	(0.0116)	(0.0127)	(0.0118)
3					0.212***	0.116***	0.216***	0.125***
					(0.0108)	(0.00979)	(0.0111)	(0.0105)
4					0.169***	0.112***	0.173***	0.122***
					(0.0104)	(0.0117)	(0.0107)	(0.0128)
5+					0.136***	0.0689***	0.139***	0.0742***
					(0.00830)	(0.00741)	(0.00843)	(0.00801)
Nearest Trust Characteristics								
Waiting Times 2012					1.000	1.002***	1.000	1.002***
					(0.000190)	(0.000374)	(0.000190)	(0.000370)
Early Foundation Trust					0.766***	0.733***	0.770***	0.740***
					(0.0250)	(0.0363)	(0.0250)	(0.0358)
GP Practice characteristics								
Practice list size ('000s)							1.016***	1.045***
							(0.00490)	(0.00739)
Number of GPs in the practice (relative to 0)								
2-3							0.982	1.259***
							(0.0446)	(0.105)
4-6							0.906*	1.313***
							(0.0484)	(0.124)
7+							0.945	1.226*
							(0.0643)	(0.142)
% GPs non EEA							0.909**	0.874*
							(0.0358)	(0.0616)
% GPs under 40							0.855***	0.811**
							(0.0470)	(0.0720)
Observations	1,356,285	713,463	1,356,285	713,463	1,356,285	713,463	1,356,285	713,463
Pseudo R2	0.617	0.650	0.621	0.699	0.633	0.717	0.634	0.719

Notes: *** denotes significance at 1%, ** at 5%, and * at 10% level. Observations are at the GP-hospital level. The dependent variable takes the value 1 if the hospital is in the GP's choice set. Estimates are odds ratios. Standard errors are clustered at the GP level and are robust to the presence of heteroskedasticity. All specifications include government office region fixed effects.

Table A2: **Mixed Logit Results: GP Practice level random coefficients**

	Coeff	SE	p-value
Attributes with RC			
Distance			
Mean	-0.101	0.017	0.000
SD	0.083	0.005	0.000
ISP			
Mean	-0.832	0.591	0.149
SD	1.076	0.088	0.000
Emergency Readmissions			
Mean	-0.415	0.449	0.000
SD	1.547	0.091	0.000
Attributes w/out RC			
Early Foundation Trust	-0.219	0.378	0.563
Waiting times (weeks)	-0.066	0.039	0.085
Specialist Orthopedic Hosp	2.620	0.446	0.000
ISPs			
ISP x Age	0.009	0.004	0.033
ISP x Ethnic Minority	-0.441	0.296	0.137
ISP x Deprivation (richest)	-0.368	0.401	0.359
ISP x Deprivation (poorest half)	-0.251	0.203	0.215
ISP x Prev Em Admit	-0.714	0.115	0.000
ISP x Prev El Admit	-0.169	0.088	0.0427
ISP x CI of 1	-0.364	0.110	0.000
ISP x CI of 2+	-1.070	0.182	0.000
ISP x GP HHI	-0.486	0.505	0.337
ISP x GP refs to ISPs	0.471	1.927	0.807
Spec Hosp x Ethnic Minority	0.370	0.305	0.225
Spec Hosp x Deprivation (richest)	0.653	0.509	0.199
Spec Hosp x Deprivation (poorest half)	0.364	0.257	0.156
Spec Hosp x Pre Em Admit	-0.034	0.118	0.775

Notes: The sample includes all hip replacements of a random sample of 1000 GP practices and who chose one of the ten closes hospitals to the centroid of their LSOA. Random coefficients are estimated at the GP level.