Institute for Fiscal Studies

22/30

Working paper

George Stoye

The distribution of doctor quality: evidence from cardiologists in England

Economic and Social Research Council

The Distribution of Doctor Quality: Evidence from Cardiologists in England*

George Stoye

August 2022

Abstract

There is widespread and unexplained variation in the outcomes of similar patients across place and providers in all developed health systems. This paper provides new evidence on the role senior doctors play in determining patient outcomes. I exploit within-hospital quasi-random assignment of patients to senior doctors following a heart attack to estimate the effectiveness of individual doctors, and to estimate returns to experience for these doctors. 28% of doctors work in multiple hospitals over a 13 year period, enabling the separate identification of doctor effects from hospital effects or observable patient characteristics. I find that a standard deviation increase in doctor quality reduces mortality rates over the next year by 3.6 percentage points, or 25% of mean mortality. There are relatively modest returns to specific experience, with mortality reductions from a standard deviation increase in the physician's 3-year caseload equivalent to around 6% of a standard deviation in permanent doctor quality. Estimating the effectiveness of each physician when treating patients with specific diagnoses, I analyse potential mortality reductions from reallocating doctors across patients. I find that mortality could be reduced by 8% by reassigning doctors within-hospital to patients on the basis of their comparative ability to treat each patient type. These results suggest that substantial improvements in patient outcomes could be achieved by reallocating existing senior staff resources.

JEL Classification: I10; I14; J24.

Keywords: Doctor quality; Doctor productivity; Health inequalities

*I thank Richard Blundell, Joe Doyle, Eric French, Jon Gruber, Tom Hoe, Attila Lindner, Rory McGee, Carol Propper, Imran Rasul, Marcos Vera-Hernández, Ben Zaranko, seminar participants at IFS, UCL and King's College London, and conference participants at the 2017 IIPF, 2018 RES, 2019 EEA, 2021 ASHEcon, 2021 Essen, 2022 SOLE and 2022 EUHEA conferences for helpful discussions and comments on this paper. I thank NHS Digital and the Office for National Statistics for access to the Hospital Episode Statistics and official mortality statistics under data sharing agreement CON-205762-B8S7B. I gratefully acknowledge financial support from the UK Economic and Social Research Council through the ESRC Centre for the Microeconomic Analysis of Public Policy (ES/T014334/1) and 'The impact of medical labour on variation in patient outcomes: evidence from English public hospitals' research grant (ES/S003118/1). Any errors are my own. Author affiliation: Stoye (Institute for Fiscal Studies, 7 Ridgmount Street, London, WC1E 7AE, UK. george_s@ifs.org.uk).

1 Introduction

Widespread variation in patient outcomes across place and providers has been well documented in many developed countries (Skinner, 2011). An important driver of this variation is likely to be the behaviour and performance of doctors, who play a key role within all healthcare systems - diagnosing patients, and deciding on and administering treatment. A growing literature has documented wide variation in the beliefs and prefences of physicians (Cutler et al., 2019; Chan et al., 2022), their preferred styles of treatment (Currie et al., 2016; Molitor, 2018; Currie and Macleod, 2020) and differences in the resource utilisation of different doctors (Doyle et al., 2010; Van Parys, 2016). Given the essential roles doctors play, these variations in behaviour can literally mean the difference between life and death for patients. Better understanding of the variation in doctor performance is therefore key to providing better and more efficient care.

This paper provides new empirical evidence on the role played by senior doctors in explaining variation in patient outcomes. I exploit two key institutional features of the English National Health Service (NHS) to estimate the distribution of individual doctor fixed effects on patient survival following a heart attack and the returns to experience in treating such patients. Furthermore, I show that the performance of doctors varies when treating patients of different severity, and use these type-specific estimates of doctor quality to show that substantial improvements in patient outcomes could be attained by reallocating doctors to specific patients.

I focus on patients being treated following a heart attack in English public hospitals. This provides an ideal setting to minimise concerns about selection of patients to specific doctors, or vice versa, that could potentially bias estimates of doctor quality (Glance et al., 2008). Heart attack patients require rapid care, with outcomes worsening significantly following delays. Patients in England can seek care in any of the public hospitals in the country, with no available private market for emergency medicine. Upon admission, patients are assigned to an available cardiologist. This therefore provides a setting where within-hospital patient assignment to doctors is essentially random.

A second advantage of the English hospital setting is that doctors are public, salaried employees. Doctors are not paid on the basis of the procedures that they carry out and therefore do not face obvious financial incentives to deviate from what they believe to be medical best practice. Senior doctors in England are routinely contracted to multiple hospitals, while moves across hospitals over time are also common. As a result, more than a quarter of senior doctors are observed treating heart attack patients in multiple hospitals over the 13-year period studied. This provides a setting in which the underlying individual productivity of doctors can be measured separately from the hospitals in which they work, and free of the influence of financial incentives that could mask differences in innate productivity.

My analysis uses administrative hospital data from the census of publicly funded inpatient hospital care in England. These data provide extensive information on patient diagnoses and treatments. Patients are tracked over time to create detailed health histories, and all admissions are tagged to an identifier of the senior doctor with overall responsibility for the patient. The hospital records are linked to official mortality statistics that record mortality for patients up to a year after hospital admission occurs. This enables me to study the treatments provided to, and the long-run mortality outcomes of, all patients in England who received treatment from a cardiologist following a heart attack between April 2005 and March 2018.

The analysis proceeds in two parts. In the first part of my analysis, I set out a simple model for patient survival following a heart attack. I estimate this model by regressing an indicator of patient survival on a rich set of patient characteristics, in addition to a full set of physician and hospital dummy variables, and measures of doctor experience. This is analogous to the two-way fixed effects models commonly used to study worker and firm effects in wage settings, as first proposed by Abowd et al. (1999), with an additional term that captures doctor experience. Under a set of identifying assumptions, this approach recovers estimates of the persistent differences in patient outcomes across doctors, and the returns to specific and general experience in treating cardiology patients.

Identifying these parameters depends on two key conditions. First, that after conditioning on hospital, time of treatment, and observed characteristics, patients are randomly assigned to doctors. The institutional features outlined above suggest that this is highly likely for heart attack patients within the English system. I provide further empirical evidence in support of this assumption by showing there is no clinically meaningful relationship between a wider set of patient characteristics and the average outcomes of their assigned doctor. Second, doctor and hospital fixed effects can only be separately identified within a connected set of doctors and hospitals, formed by doctors treating patients in multiple hospitals. I show that working in multiple hospitals - either contemporaneously or over the course of a career - is common, and identifies a common set that captures the majority of heart attack patients in England.

The results show there is considerable variation in the quality of individual doctors: a standard deviation increase in quality reduces one-year mortality rates by 3.6 percentage points, equivalent to 25% of mean mortality, or 0.1 standard deviations in mortality. This suggests that patient outcomes could be considerably improved by raising average doctor quality.

The estimates also indicate that there are some gains to specialisation among cardiologists, but the improvements in performance with experience are small relative to permanent differences in the performance of these (already experienced) doctors. I find that a standard deviation increase in the number of heart attack patients treated over a three-year period increases doctor quality by the equivalent of 6% of a standard deviation in permanent quality.

I use these estimates to show that average doctor quality varies substantially across different geographic regions. For example, doctors treating patients in London and the South East are estimated to have, on average, fixed effects that are a standard deviation higher than doctors treating patients in the North East. More generally, estimated average doctor quality is higher for patients living in the least deprived and more urban areas of the country, potentially reinforcing existing inequalities in health status across groups living in different areas.

In the second part of my analysis, I extend the model to allow doctors to vary in their effectiveness in treating different patient types. I split patients by whether or not they are experiencing a particular type of heart attack (a ST-elevated myocardial infarction, or STEMI) that requires quicker and more invasive treatment, and estimate fixed effects for each doctor when treating each patient type. Using these estimates, I rank doctors by their comparative advantage in treating STEMI patients and consider two scenarios where STEMI patients are reallocated to the doctors with the greatest comparative advantage. This includes an 'unconstrained' scenario where doctors can be assigned to patients in any hospital, and a 'constrained' scenario where doctors can only be reassigned to patients within the hospitals in which they work. In line with similar past exercises that examine the potential gains from replacing poorly performing teachers (Hanushek, 2009), I also estimate the gains from replacing the worst performing 10% of doctors as a comparison case.

These estimates show that doctors do vary in their ability to treat different patient types, and that there are potentially substantial gains to reallocating patients to more appropriate doctors. The estimates suggest that, in 2017, deaths in the year following heart attack treatment could have been reduced by 21% if doctors were reallocated in the unconstrained scenario and 8% when restricting reallocations to be only within hospital. These effects are actually larger than the 5% reduction from simply replacing the worst performing 10% of doctors. This indicates that even among highly trained doctors, there is variation in their ability to carry out different tasks. Changing shifts patterns to better match patients and doctors therefore has the potential to improve patient outcomes.

My work contributes to two literatures. First, to the literature on variation in care quality across providers. Much of the past work has focused on differences in care quality across hospitals, without explicitly considering the roles played by individual doctors (Gowrisankaran and Town, 1999; Geweke et al., 2003; Doyle et al., 2015; Hull, 2020). An early exception is Epstein et al. (2010) who show significant variation in the effectiveness of obstetricians who specialise in different delivery methods. More recent work has begun to further examine variation in the performance of individual doctors in different settings (Currie and Zhang, 2021; Dahlstrand, 2022; Ginja et al., 2022) and the returns to experience for doctors (Chen, 2021; Lundborg et al., 2021). I contribute to this nascent literature by estimating quality measures for individual doctors across a large, national health system, and use these measures to both document inequalities in access to high quality care, and to show how health outcomes could be improved by reorganising existing staff resources.

Second, I contribute to a literature that examines the individual behaviour of physicians (e.g. Chandra and Staiger, 2007; Epstein and Nicholson, 2010; Currie et al., 2016; Van Parys, 2016; Molitor, 2018; Currie and Macleod, 2020; Doyle, 2020; Chan et al., 2022). This work typically focuses on the impact of individual decisions or behaviours on costs and patient outcomes. My work most closely relates to past work that has exploited movements in doctors across hospitals in order to study the behaviour of individual doctors (Molitor, 2018; Kwok, 2019; Fadlon and Parys, 2020). I extend this work by providing a summary measure of physician productivity in order to demonstrate the wide variation in the persistent performance of these doctors.

My empirical approach also relates to a broader set of papers that use two-way fixed effects models, which are usually identified by changes in firm of employment or place of residence by individuals, in order to identify individual effects separately from broader factors. In particular, this relates to literatures that examine the individual productivity of workers and firms (e.g. Abowd et al., 1999; Card et al., 2013; Bonhomme et al., 2019), the impact of teachers on their students' test scores and later life outcomes (e.g. Aaronson et al., 2007; Mansfield, 2015; Chetty et al., 2014a,b) and the drivers of geographic variation in health care utilisation and outcomes (Finkelstein et al., 2016, 2021).

The rest of this paper is organised as follows. Section 2 describes the institutional features of the English public hospital system and the data used in the analysis. Section 3 sets out a model of the determinants of patient outcomes following a heart attack. Section 4 describes my empirical approach and identifying assumptions. Section 5 presents the baseline results, before discussing a series of robustness checks and alternative specifications. Section 6 examines how average doctor quality varies by region. Section 7 sets out a more flexible model to estimate doctor quality when treating different patient types, and estimates the potential improvements in patient mortality outcomes when reallocating patients across doctors. Section 8 concludes.

2 Background and data

2.1 Emergency health care in England

Emergency health care in England is publicly funded through general taxation, and is available free at the point of use for all residents. Acute care is provided by large, publicly owned hospitals, who are reimbursed by the government for the care they provide, and staffed by public employees.[1] There are no restrictions on where patients can access emergency care, but patients typically attend their nearest hospital. There is no private market for emergency care.[2] As a result, patients who require acute emergency care are all treated by a public hospital.

This paper focuses on patients receiving treatment for an acute myocardial infarction (AMI) or heart attack. Heart attack patients require immediate care, with marked improvements in survival rates among patients receiving treatment within hours of onset (Maxwell, 1999). Emergency treatment for these patients typically follows one of two pathways, with different types of treatment given at different stages. The majority of patients will first be treated in the emergency department (ED) before being admitted to a specialist cardiology department.[3] Patients either arrive independently at the ED, or by ambulance following an emergency call. In 2017/18, 71% of heart attack patients who received ED treatment arrived by ambulance. Upon arrival, patients undergo an initial assessment to establish the severity of their condition, followed by a series of investigations and (if necessary) treatments to stablise their conditions. When required, patients are then admitted as an inpatient, and sent to the cardiology department where they will be assessed and treated by a trained cardiologist.

[1] Payments to hospitals vary according to the care provided. Treatments are assigned to a Healthcare Resource Group (HRG), similar to Diagnosis-related Groups (DRGs) in the US. Each treatment is assigned to a nationally set tariff, with small adjustments made across regions with varying fixed costs and for very long length of stays.

[2] There is a small private market for elective health care in England, with treatment funded out-of-pocket or by private medical insurance.

[3] EDs are known as Accident and Emergency (A&E) departments in England.

Alternatively, some ambulance patients may be admitted directly to the cardiology department, without first attending an ED. These patients will undergo initial tests and treatments in the ambulance, who communicates these results directly to the hospital. In 2017/18, a third of heart attack patients treated by a cardiologist were admitted without attending an ED first.

Patients can receive a number of treatments as a cardiology inpatient depending on the nature of the heart attack and the preferences of the staff treating them. Two types of treatments are commonly used, and sometimes combined. First, intravenously administered thrombolytic drugs can be used to dissolve blood clots that have blocked blood flow through major arteries and veins. This treatment is most effective when provided to patients within 12 hours of the onset of the heart attack (Windecker, 2014). Second, more invasive treatments can also be used to restore blood flow to the heart, with 77% of AMI patients undergoing a procedure when being treated by a cardiologist in 2017/18.[4] The most commonly used procedure for this purpose is angioplasty, where a balloon is used to open blocked arteries, and which can be combined with the insertion of a stent to maintain blood flow (a process known as a percutaenous coronary intervention, or PCI). Where angioplasty is not appropriate, patients may instead undergo more invasive surgery, such as a coronary artery bypass graft (CABG) which diverts blood flow around the blockage in the artery.

2.2 The role of senior doctors

All inpatient care in public hospitals is overseen by a senior doctor, known as a consultant.[5] These doctors are legally responsible for patients, and are in charge of the treatment given to patients. They will manage the overall treatment of patients, either directly treating patients or overseeing decisions made by more junior staff. This will include assessing patient severity, deciding on the course of treatment, and performing individual surgeries. They will also set out plans for secondary prevention of future heart attacks and related conditions, including providing lifestyle advice and prescribing post-discharge medication. This paper focuses on estimating differences in the performance of senior doctors treating heart attack patients, and will combine all of these elements.[6]

[4] As I discuss in Section 7, procedure rates vary considerably across specific diagnoses, with patients experiencing STEMI heart attacks significantly more likely to undergo a surgical procedure.

[5] Consultants are equivalent to 'attending physicians' in the US.

[6] Junior staff are not included in the data, and no indication is provided of who actually performs surgery. While organisation of staff varies across hospitals, consultants will typically share registrars (equivalent to residents in the US) rather than always working with the same doctor on every shift. The estimates of doctor performance will therefore combine managerial ability, underlying medical skill and effort.

All inpatients are assigned to a responsible consultant upon admission. In the case of a heart attack, patients will be assigned to the cardiologist consultant who is working at the time of admission.[7] Hospitals schedule, for every shift, a consultant to be available to carry out emergency treatment if a heart attack patient arrives. The identity of the consultant is scheduled in advance but is not publicly available, and so would be unknown to patients ahead of time. Patients do not receive a choice about which consultant they are assigned to (conditional on attending a particular hospital at a particular time), while consultants cannot refuse to treat specific patients. As a result, conditional on hospital and the time attended, patients are essentially randomly assigned to consultants.

Consultants are all experienced doctors. At a minimum, cardiologist consultants will have undergone at least five years of medical school, two years (or more) of basic training, and a five-year cardiology residency (which can be extended if consultants also undertake research during this period).

An important feature of the English National Health System (NHS) is that doctors regularly work in multiple hospitals. Doctors are contracted by each (publicly owned) hospital organisation separately.[8] However, many consultants hold positions at multiple hospitals simultaneously, or move between hospitals over time. As a result, doctors are observed treating patients at multiple hospitals. This enables me to separately study the impact of doctors and hospitals on patient outcomes, as set out in detail below.

2.3 Data

The primary source of data for the analysis comes from the inpatient Hospital Episode Statistics (HES). These data cover all public hospital admissions between April 1997 and March 2018. Observations are recorded at the episode level, with an episode recording the period of care under the responsibility of a single consultant. The data record detailed information about the patient and the care they receive, including age, sex, local area of residence, admission and discharge dates, a primary diagnosis and up to 19 secondary diagnoses, an extensive list of procedure codes, and a hospital identifier. All patients are tracked by a pseudonymised patient identifier, which can be used to create detailed histories of past hospital treatment and

[7] Consultants may also be 'on call' for emergencies, with specific consultants available to attend the hospital in the case of a heart attack patient arriving and requiring treatment.

[8] Hospitals in the same area are grouped into 'Trusts', with shared management. In most cases, trusts have only one or two major acute hospitals, with smaller hospitals providing specialist non-acute care. Throughout the text, I refer to trusts as 'hospitals'.

diagnoses.

From March 2003, all patients are also linked to a consultant identifier. This enables the assignment of patients to consultants, and also allows me to derive histories of consultant activity. Doctor experience is observed imperfectly: patient records between March 2003 and April 2018 can be used to track activity of all consultants during this period, but does not provide information on experience in other roles or prior to this date. The data do not provide any information on less senior doctors or nurses who also treat the patient.

Mortality outcomes are recorded by the UK Office for National Statistics (ONS), and linked to individual patients through anonymised identifiers based on patient National Insurance (Social Security) numbers. These data include the date of death for all individuals who died in the UK, or UK citizens who die abroad, between April 2005 and March 2019. I use these data to create indicators of whether a heart attack patient died within 30 days and a year of initial treatment.

Table 1 reports summary statistics for the 510,249 patients included in the analysis, who were treated in public hospitals between April 2005 and March 2018.[9] This shows the mean and standard deviation for a range of patient characteristics and outcomes. Mean patient age was 67.8 years. 68% of patients were male, and 81% were white. In the previous year, patients had received a mean of £1,661 of inpatient treatment, had an average of 0.5 emergency admissions, and spent 2.1 days in hospital. 0.7% of patients had received treatment for a stroke since 1997. The average Charlson index score (a commonly used measure of patient comorbidity) was 1.9. 40.9% of patients were treated for a STEMI.

The central focus of this paper is to understand how individual senior doctors affect the mortality outcomes of their patients, and how these outcomes could be achieved if existing staff resources were reallocated. Table 1 shows that mortality is a common outcome following a heart attack. 6.6% of patients died within 30 days of admission. Over the course of a year, the mortality rate more than doubled, to 14.1%.

[9] I discuss in detail the construction of this sample in Section 4.2.

Table 1: Descriptive statistics of AMI patients

	Mean (1)	S.D. (2)
Demographics		
Age	67.8	13.6
Male	0.68	0.47
White	0.81	0.39
Health and past utilisation		
STEMI	0.409	0.492
Charlson index	1.91	1.28
Past stroke	0.007	0.081
Treatment cost (t-1)	1,661	8,155
No. em. admissions (t-1)	0.49	4.80
Days in hospital (t-1)	2.11	10.75
Mortality outcomes		
30-day death rate	0.066	0.248
1-year death rate	0.141	0.348
Number of patients	510,249	

Notes: (1) A STEMI is identified by ICD-10 codes I210, I211, I212, I213, I220 and I221; (2) Past stroke records whether the patient has been admitted to an English hospital since April 1997 for stroke treatment; (3) Treatment costs include inpatient treatment costs in the past year, and are reported in 2018 GBP; (4) No. em. admissions is a count of emergency admissions in the past year; (5) Days in hospitals counts the number of total days spent in hospital in the past year.

3 Health production function

After patient i suffers a heart attack in period t, they attend hospital k and are treated by doctor j. Survival up to one year after treatment (Y_{ijkt}) is determined as follows:

$$Y_{ijkt} = \beta X_{it} + \mu_{jt} + \psi_{kt} + \epsilon_{ijkt} \tag{1}$$

X_{it} captures observable individual determinants of patient survival, including patient demographic characteristics and health history. μ_{jt} is the impact of being treated by doctor j in period t. ψ_{kt} is the impact of being treated in hospital k in period t. ϵ_{ijkt} captures any remaining factors that are unobserved and affect patient survival.

The impact of being treated by a specific doctor in each period (μ_{jt}) can be decomposed into three terms:

$$\mu_{jt} = \mu_j + \delta Exp_{jt} + \nu_{jt} \tag{2}$$

where μ_j is a doctor fixed effect. It is interpreted as the permanent quality of doctor j, as measured by their impact on the survival probability of their patients. The paramter captures the effect of assigning overall responsibility for a patient's care to a particular senior doctor. As noted above, senior doctors play several roles in planning and providing patient care. I do not attempt to separate the extent to which this parameter captures underlying medical skill, managerial ability or physician effort.

Exp_{jt} measures the experience of the doctor in each year. δ therefore represents an experience profile in treating heart attack patients that is common across all doctors. ν_{jt} captures any year-specific shocks in doctor performance. I assume that this shock has a mean of zero. This means that doctor performance only systematically varies across years on the basis of experience.

This model has two further important restrictions. First, the effects of hospitals and doctors are additively separable. This is a strong assumption: it rules out that doctors perform better or worse in particular hospitals. This would be violated, for example, if hospitals differ in the technology available for treating heart attack patients, and doctors vary in their ability to use different technologies. I discuss this assumption further in Section 5.2 and present empirical evidence that suggests that match effects between doctors and hospitals are limited.

Second, the model also restricts doctor quality to be fixed across all patients. A more flexible model would allow heterogeneity in the ability of each doctor to treat patients with different conditions or characteristics. In Section 7, I explicitly relax this assumption to examine variation in doctor quality across different patient types (based on their specific diagnosis). I then use this to examine potential gains in patient survival from reallocating doctors to treat patients of specific types.

4 Empirical strategy

4.1 Baseline estimation

The main focus of this paper is to produce empirical estimates of permanent quality for each doctor (μ_j) and the common experience profile across doctors (δ). To do this, I estimate the following specification:

$$Y_{ijkt} = \beta X_{it} + \mu_j + \delta Exp_{jt} + \psi_k + \xi_{ijkt} \tag{3}$$

Y_{ijkt} is a binary variable that takes the value of one if patient i survives for one year after receiving treatment, and zero otherwise.[10] X_{it} captures a rich set of patient demographic and health characteristics. This includes age dummay variables, sex (also interacted with age), ethnicity, dummy variables for each element of Charlson comorbidity index, the cost of any hospital treatment in the previous year, an indicator of whether the patient has previously had a stroke, a set of fixed effects capturing the primary and one other diagnosis recorded at the time of admission (ICD-10 codes), indicators of the day of the week, month and year when the patient is admitted, and an interaction between month and year of admission.[11]

μ_j and ψ_k are doctor and hospital fixed effects respectively. The inclusion of a hospital fixed effect will control for permanent differences in the casemix of patients treated in each hospital, the supply side of hospitals (e.g. management and equipment), and other elements of care quality. Patients are free to choose which hospital they attend, and although they often attend their nearest hospital, there may be selection into hospitals based on the characteristics of patients who live in the local area. I therefore view the inclusion of hospital fixed effects as an important control, but do not make any causal claims about the impact of receiving treatment at a particular hospital on patient mortality outcomes.[12]

Experience (as measured by the number of years worked as a cardiologist) is imperfectly measured in the data. The senior doctors included in the data will also all be relatively experienced already. As a result, I focus on two measures of activity undertaken by these doctors in the baseline specification as measures of specific and general experience. First, I include the number of heart attack patients treated by doctor j in the previous three years. This measures specific experience in treating heart attack patients in a recent period. Second, I include the number of all other patients treated in the previous three years. This measures recent general experience.

The error term (ξ_{ijkt}) captures any further factors that influence patient mortality outcomes. As discussed below, the validity of the estimates of doctor quality will rely on a lack of correlation between this error term and the identity of the doctor treating the patient. As a result, it is

[10] I also repeat this using 30-day survival.

[11] For the period after 2009, when complete ED data are available, I also include indicators of whether the patient arrived by ambulance, the exact time of arrival at the ED and discharge to the cardiology department, and the number and type of investigations and treatments. The inclusion of these variables has a negligible impact on the estimates of μ_j. Results available upon request.

[12] Note that in Equation (2.3) I do not allow the effect of hospitals to vary over time (as specified in Equation (1)). As I discuss below, this allows me to identify fixed effects for a greater number of doctors. In Section 5.2, I show for a smaller set of doctors that the results change very little when including time-varying hospital effects.

helpful to consider a further decomposition of this error term into three parts:

$$\xi_{ijkt} = \nu_{jt} + \lambda_{kt} + \upsilon_{ijkt} \tag{4}$$

As in equation (2), ν_{jt} represents year-specific shocks in doctor performance that could arise due to a range of factors, including their own health and personal life, or the influence of past performance that is not fully captured by the experience measures included in equation (3). λ_{kt} represents year-specific shocks in hospital performance. These include changes in wider management practices or in the equipment available for staff (e.g. if a hospital installs a catheterisation lab). Finally, υ_{ijkt} includes any remaining individual survival factors, including luck, and any measurement error.

4.2 Identifying assumptions

Estimation of permanent doctor quality relies on two key assumptions. The first concerns the within-hospital assignment of patients to doctors:

Assumption 1 (Quasi-random conditional assignment of patients to doctors)**.** *Conditional on hospital and time of treatment, and patient observables, patients are as good as randomly assigned to senior doctors*

$$Cov(\mu_j, \xi_{ijkt}) = 0 \tag{5}$$

This assumption rules out both that patients choose to be treated by a particular doctor, and that doctors choose or are systematically allocated to patients with particular (unobserved) characteristics. Satisfying this assumption has often been problematic in the past when attempting to measure doctor performance, and has been a common criticism of attempts by policymakers to introduce report cards for hospitals or individual doctors (Marshall et al., 2000; Dranove et al., 2003).

Importantly, this assumption is about within-hospital assignment. Patients may choose to attend specific hospitals. Indeed, patient mix is likely to vary across hospitals, with some hospitals treating sicker or more deprived populations than others. This assumption does not rule out patients systematically sorting into particular hospitals, but only that doctors working at a particular hospital are not assigned to particular patients within the pool of patients attending the hospital. I examine geographic patterns of doctor-patient sorting in Section 6.

Two institutional factors make this assumption likely to hold in practice. First, heart attack patients require rapid treatment, with outcomes deteriorating as time passes (Maxwell, 1999). Treatment is therefore unlikely to be routinely delayed in order to find alternative physicians (either at the request of the patient or the doctor). Second, as noted in Section 2.2, patients are assigned to senior doctors who are either physically present in the hospital at the time of arrival, or who are specifically 'on call' for emergency cases. Scheduling of these doctors are not publicly available, and patients would be unlikely to choose a particular time and hospital to seek heart attack treatment from even if the identity of doctors was known in advance.

In Section 5.2 I provide empirical evidence in support of this assumption by examining the relationship between augmented estimates of doctor quality and patient characteristics. I show that there is no meaningful relationship between these observables and the estimated quality of the assigned doctor, suggesting that the assumption holds in practice.

In addition to the assumption on patient assignment, the doctor and hospital effects set out in equation (3) are only identified within a set of hospitals that are connected by physicians who worked at multiple hospitals (Abowd and Kramarz, 1999). This gives rise to the second identifying assumption:

Assumption 2 (Connected set of doctors and hospitals)**.** *There is a connected set of doctors and hospitals, where doctors treat patients in multiple hospitals, such that all doctors in this set can be identified separately from the hospitals in which they work.*

This assumption allows me to compare doctors both within and across different hospitals, and is equivalent to having doctors who treat more than a minimum number of patients in at least two hospitals over the 13 years considered. Intuitively, this allows me to compare the performance of doctors within each hospital, and then use doctors who work in two hospitals as a benchmark for comparing doctors in the two connected hospitals. Expanding this across the whole national system allows me to compare the performance of doctors working in any hospital which is linked to the others through a doctor working in more than one hospital.

Hospitals are linked both by doctors who transfer across jobs over time, and by doctors who work for multiple hospitals at the same time. The institutional features of the English public hospital system are important in providing these links between hospitals, with many doctors working in multiple hospitals at the same time. The size of the connected set, and the number of patients and doctors observed within it, depend on the restrictions placed upon the number of patients that 'linking' doctors must treat in each hospital to be included within this set. Doctors who treat very few patients in a specific hospital would have noisy outcomes, and would not provide good comparators with other doctors in this hospital.

In the baseline analysis, I create this connected set in the following way. First, in order to minimise selection of patients to consultants, I include only patients treated by a cardiologist (rather than a consultant from another specialty) and exclude patients who had previously received treatment for a heart attack in an English public hospital since 1997.[13] Second, I excluded all patients treated by cardiologists who treated fewer than 10 heart attack patients in a single hospital over the 13 year period. Finally, I define the connected set to include all remaining patients in hospitals where at least one cardiologist treated more than 10 heart attack patients in at least one other hospital between April 2005 and March 2018.[14] This provides a final analysis sample of 510,249 patients treated by 1,587 consultants across 141 hospitals. This includes all cardiologists treating more than 10 patients in these hospitals: the 437 'linking' consultants who treated at least ten patients in two hospitals during this period of time (who accounted for the treatment of 153,181 patients, or 30% of all patients) and the 1,150 who are observed only working in a single hospital.

Appendix Figure A.1 shows the distribution of 'linking' consultants across hospitals. This shows that there is variation across hospitals in the number of consultants that link the hospital to the rest of the set: some hospitals are linked by a single consultant, while others are linked by more than 20 consultants. Similarly, Appendix Figure A.2 shows that there is wide variation in the number of patients treated by doctors in their 'secondary' hospital (e.g. where they treat the fewest patients). The modal caseload in the secondary hospital is between 10 and 20 patients, but some consultants treat hundreds of patients in multiple hospitals over 13 years.

Table 2 shows summary statistics for the 1,587 cardiologists included in the analysis. On average, these consultants treated 322 heart attack patients over a 13 year period, 41% (132) of which were STEMI patients. Consultants treated heart attack patients for an average of 7.2 years, and had an average caseload of 50 heart attack patients in each year.

[13]Patients who are treated by non-cardiologists may be different to patients treated by cardiologists and so I drop these patients over concerns about potential selection. This removes 189,727 heart attack patients. Similarly, patients who have previously been treated for a heart attack may be more likely to request or be transferred to their previous cardiologist. I therefore drop 19,303 patients who had previously received heart attack treatment. Patients by definition only appear in the sample once as a result.

[14]These two sample restrictions remove 9,766 patients from the final analysis. Using an alternative minimum caseload to define this set does not alter the results substantially. Results are available upon request.

Table 2: Descriptive statistics of cardiologists

	Cardiology	
	Mean	S.D.
Volume	321.52	311.85
Volume (STEMI only)	131.6	164.83
Annual Volume	48.99	34.69
30-day death rate	0.065	0.033
1-year death rate	0.143	0.051
Multiple hospital (entire period)	0.275	0.447
Multiple hospital (same month)	0.137	0.344
Number of hospitals	1.38	0.97
Number of years in the data	7.21	4.11
Number of consultants	1,587	

Notes: (1) Includes all consultants in the connected set who treated a minimum of 10 AMI patients between April 2005 and March 2018; (2) Multiple hospital takes the value of one if a consultant is observed treating at least ten patients in more than one hospital over the specified period.

Table 2 also shows that many doctors are observed treating patients in multiple hospitals. Over the 13 year period, 27.5% of consultants are observed treating at least one patient in multiple hospitals. On average, consultants treated patients at 1.4 hospitals over the entire period. Doctors were also regularly working in multiple hospitals at the same time, with 13.7% of doctors treated patients in multiple hospitals in the same month. These working patterns provide the important variation that enable the separate identification of doctors and hospitals within the connected set.[15]

4.3 Empirical implementation issues

A common concern in literatures that conduct similar exercises for other workers (most notably teachers) is that the variation in worker fixed effects may be overestimated when such estimates are based on small samples for individual workers (Card et al., 2013). Intuitively, doctors who are observed treating very few patients may be lucky or unlucky in their outcomes. In this case, these doctors will either have very large positive or negative estimated fixed effects but these estimates will simply reflect statistical noise as opposed to being a good signal of the true ability of these doctors.

[15] Appendix Table A.1 shows summary statistics separately for 'linking' and 'non-linking' consultants. 'Linking' consultants were, on average, observed for an additional year in the data. They also treated more AMI patients over the period (an average of 350 compared to 310 patients) and slightly more each year (an average of 53 patients, compared to 48 patients for all other cardiologists). However, there were no statistically significant differences between the groups in the percentage of STEMI patients, or in 30-day and 365-day mortality dates.

To address the impact of statistical noise on the estimates of each doctor fixed effects, I implement an Empirical Bayes shrinkage estimator. This has been commonly used in the teacher value-added literature to adjust estimates for low class sizes (Kane and Staiger, 2008; Kane et al., 2008; Chetty et al., 2014a). I shrink estimates according to a shrinkage factor, λ_j, that measures the proportion of the variation in the average doctor residual that is due to signal variance:

$$\lambda_j = \frac{\sigma_u^2}{\sigma_u^2 + (\sigma_e^2/n_j)} \tag{6}$$

σ_u^2 is the between-doctor variance in mortality outcomes, σ_e^2 is the within-doctor variance, and n_j is the total number of heart attack patients treated by doctor j over the entire period. Estimates are shrunk towards zero if doctors treat few patients, or if a larger share of the overall variation is attributed to within-doctor variation.

In my baseline estimates, I estimate these parameters as sample analogues from my fixed effects regression. Alternative methods have been used in the teacher value-added regression, including a two-step method that uses the covariance within-teachers across different class years in order to estimate the reliability of the estimates (Kane and Staiger, 2008; Chetty et al., 2014a; Bitler et al., 2021). I show in Section 5.2 that while this method reduces the magnitude of the variation in doctor fixed effects, the qualitative conclusions are unchanged.

5 Results

I first report and discuss the estimates of permanent doctor quality and returns to experience from estimating equation (3). I then present evidence that the key identifying assumption of within-hospital random assignment of patients to doctors holds, and discuss a series of robustness checks and alternative specifications that explore other threats to identification.

5.1 Estimates of doctor quality and returns to experience

Table 3 shows the estimates of doctor quality ($\hat{\mu_j}$) obtained by estimating equation (3). The dependent variable in the first two columns is 30-day survival following a heart attack, with raw estimates reported in the first column and the shrunken estimates in the second column. Columns three and four repeat this for 365-day survival as the dependent variable. In both cases, I report the standard deviation, the variance, and the 10th, 25th, 50th, 75th and 90th percentiles. Bootstrapped standard errors are reported for each point of the distribution.

The unadjusted results indicate that a standard deviation improvement in doctor quality is related to a 3.7 percentage point reduction in 30-day mortality. This is equivalent to 56% of the mean 30-day mortality rate of 6.6%, or 15% of a standard deviation. The adjusted estimates are smaller, with a standard deviation improvement in doctor quality equal to a 2.5 percentage point reduction in 30-day mortality. This is equivalent to 38% of the mean mortality rate, or 0.1 standard deviations. In both cases, the estimates are statistically significant at the 1% level.

For 365-day survival, the unadjusted results indicate that a standard deviation improvement in doctor quality is related to a 5.2 percentage point reduction in mortality, or 47% of the mean one-year mortality rate of 14.1%. The adjusted estimates indicate that a standard deviation improvement in doctor quality is equal to a 3.6 percentage point reduction in one-year mortality. This is equivalent to 25% of the mean mortality rate, or 0.1 standard deviations.

Table 3: Estimated doctor fixed effects, 30-day and 365-day survival rates

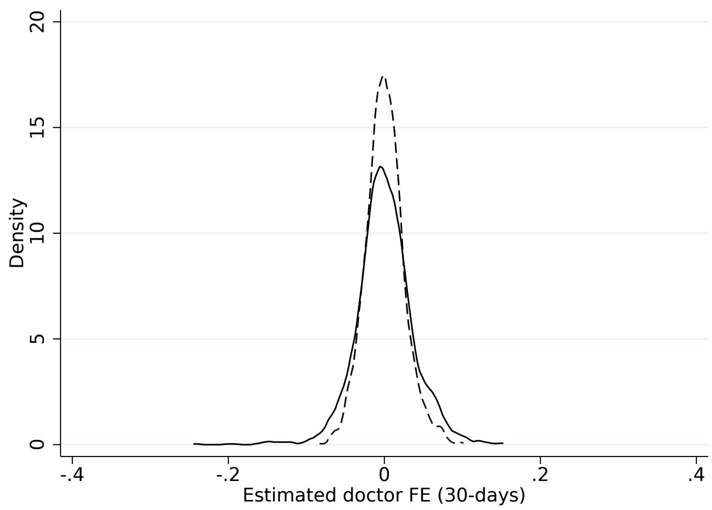
	Doctor fixed effects			
	30-day survival		365 day survival	
	Unadjusted (1)	Adjusted (2)	Unadjusted (3)	Adjusted (4)
Std Deviation	0.037***	0.025***	0.052***	0.036***
	(0.002)	(0.003)	(0.003)	(0.004)
Variance	0.001***	0.001***	0.003***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
10th percentile	-0.040***	-0.030***	-0.055***	-0.040***
	(0.003)	(0.003)	(0.004)	(0.004)
25th percentile	-0.020***	-0.015***	-0.025***	-0.018***
	(0.002)	(0.002)	(0.002)	(0.002)
50th percentile	0.000	0.000	0.003	0.002
	(0.001)	(0.001)	(0.002)	(0.002)
75th percentile	0.021***	0.015***	0.031***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)
90th percentile	0.044***	0.031***	0.062***	0.045***
	(0.003)	(0.003)	(0.004)	(0.004)
Number of patients	510,249	510,249	510,249	510,249
Number of doctors	1,587	1,587	1,587	1,587
Number of hospitals	141	141	141	141

Notes: (1) Controls include age and sex dummies (and an interaction between all of these dummies), ethnicity, dummies of each component of the charlson comorbidity index, the cost of any hospital treatment in the previous year, indicators of whether the patient has experienced a stroke in the past year, fixed effects capturing the first two recorded diagnoses at the time of admission, indicators of the day of the week, month and year when the patient is admitted, and an interaction between month and year of admission; (2) Bootstrapped standard errors clustered at the doctor level (199 repetitions).

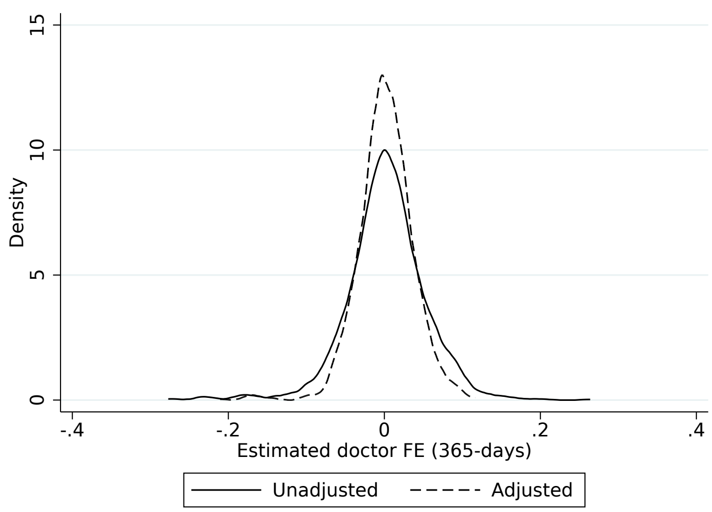
Figure 1 shows the distribution of unadjusted and adjusted doctor fixed effects for 30-day (Panel A) and one-year (Panel B) patient survival. This demonstrates the impact on the distribution when the shrinkage procedure is applied, with both the left and right tails of the distribution becoming less pronounced when adjusting the estimates for statistical noise. It also displays the large differences in the performance of doctors at either end of the distribution. For example, using the adjusted estimates, moving between a doctor at the 10th percentile to one at the 90th percentile reduces mortality by 8.5 percentage point, or around 60% of the mean mortality rate.

Figure 1: The distribution of the raw and adjusted estimated of doctor fixed effects, by survival length

(a) 30-day survival



(b) 365-day survival



Notes: (1) Doctor fixed effects from same regression as described in Table 3; (2) Adjusted fixed effects are adjusted using the procedure set out in Equation (6).

Comparing the estimated impacts over 30 days and a year suggest that the impact of being treated by a more effective doctor is persistent over time. In absolute terms, a standard deviation increase in doctor quality reduces mortality by a larger amount over the course of a year than it does over a month. As a share of mortality, this effect is reduced as mean mortality rises over a longer period. Appendix Figure A.3 plots the estimated 30-day impact on survival against the estimated 365-day impact for each doctor in the sample. There is a strong positive correlation (0.7). This indicates that doctors who perform well in the short-term are also likely to be those with good long-term outcomes.

Table 4 shows the estimates of the returns to specific and general experience from the same fixed estimates regressions as above. Column one shows the results when using 30-day survival as the dependent variable and column two shows results for 365-day survival. The count of AMI and other patients are standardised to have a mean of zero and a standard deviation of one.

Table 4: Estimated impact of recent experience on one-year survival rates

	Patient survival	
	30-day (1)	365-day (2)
Experience (last 3 years)		
AMI patients	0.0021***	0.0019*
	(0.0008)	(0.0010)
Other patients	-0.0033	0.0015
	(0.0021)	(0.0029)
Observations	510,249	510,249
Number of doctors	1,587	1,587
Number of hospitals	141	141
R-squared	0.158	0.227

Notes: (1) Estimates taken from same regression as described in Table 3; (2) Variables record the number of AMI (heart attack) and other (non-heart attack) patients treated over the past three years, and are standardised to have mean zero and a standard deviation of one; (3) Standard errors are clustered at the doctor level.

The results suggest that there is a small return to specialisation. Treating more heart attack (AMI) patients over the previous three years is associated with a small decrease in mortality over both a 30 and 365 day period. For example, a standard deviation increase in the number of AMI patients treated over the previous three years (or 293 patients) reduces mortality by 0.21 percentage points over a year. This is equivalent to an improvement of just under 6% of a standard deviation in (adjusted) doctor quality. The coefficient on the volume of other patients treated over the same period is statistically insignificant for both 30- and 365-day survival.

These results show that there are some returns to specialised experience when treating heart attack patients, but these are relatively small when compared to the permanent differences in the performance across doctors.

5.2 Robustness checks

The above results rely on the identifying assumptions outlined in Section 4.2. I now further discuss these assumptions and present evidence that they are likely to hold in practice. I first examine the empirical evidence around within-hospital quasi-random assignment of patients to doctors. I then explore other threats to identification, including the presence of hospital-specific trends in performance and potential match effects between hospitals and doctors. Finally, I set out alternative approaches to address sampling error arising from small samples.

5.2.1 Evidence of quasi-random assignment

As discussed in Section 4.2, identification of the doctor fixed effects relies on an assumption of within-hospital random assignment of patients to doctors. This seems highly plausible given the nature of heart attack treatment and the set-up of English hospitals. In addition, while this assumption cannot be tested directly, I can provide strongly suggestive evidence in support of it by examining the relationship between patient characteristics and the estimated quality of the doctor that patients are assigned to. If conditional assignment is truly random, then there should be no relationship between estimated quality and patient characteristics after taking into account the hospital and time of treatment.

I begin by examining the relationship between average doctor performance and the patient characteristics included in the initial estimates of doctor quality (i.e. the characteristics included in X_{it} in equation (3)). This includes age, sex, ethnicity, Charlson score, hospital treatment costs in the previous year, and an indicator of whether a patient was treated for a stroke in the previous year. As the assumption is that assignment is random conditional on the hospital and time of admission, I first regress each characteristic on a full set of hospital dummy variables, admission day-of-the-week, and an interaction between admission month and year.[16]

Importantly, given that these variables are all included in X_{it} in equation (3), they will by

[16]The controls for hospital and time of admission are the same as in Equation (3). The assumption is that assignment is random conditional on these factors: one would expect a correlation between patient characteristics and estimated doctor quality when these controls are removed if populations with particular characteristics have access to better or worse quality doctors. I explore this further in Section 6.

definition be uncorrelated with the estimates of doctor quality ($\hat{\mu}$) presented in Section 5.1. I therefore estimate an alternative measure of average doctor performance where these patient characteristics are omitted. Specifically, I estimate:

$$Y_{ijkt} = \gamma_j + \psi_k + \tau_t + u_{ijkt} \tag{7}$$

where Y_{ijkt} is an indicator that takes the value of one if the patient survives for 30 days after being treated by doctor j in hospital k at time t. ψ_k is a hospital fixed effect, and τ_t includes day of the week indicators and an interaction between indicators for each month and financial year in the sample period. $\hat{\gamma_j}$ therefore captures an estimate of the average outcome of patients treated by doctor j without taking into account any observed differences in patient characteristics. I use this measure to examine the relationship between average doctor performance and the characteristics of patients assigned to them.

I first show the evidence visually. Appendix Figure A.4 plots each of the residualised characteristics against the alternative estimates of doctor quality from Equation (7). This shows no obvious patterns that would suggest that certain types of patients are systematically assigned to doctors with better or worse estimated quality.

I also regress each of the residualised patient characteristics on assigned doctor performance, with all variables standardised to have a mean of zero and standard deviation for ease of interpretation.[17] Column one of Table 5 shows the results. The coefficients are all very small in magnitude, and are not statistically significantly different from zero in the majority of the cases. The only exception is patient age, where there is a statistically significant correlation. However, this is again small in magnitude: the estimates suggest that a one standard deviation increase in (residualised) patient age is associated with a 0.005 standard deviation decrease in the average 30-day survival rate of the assigned doctor.

I also examine a further set of patient characteristics that are not included in the initial controls variables in equation (3). These are socio-economic characteristics based on the local area in which the patient lives, and measures of past hospital use. This information is unlikely to be known to the hospital at the time of admission, and so unlikely to be used to sort patients to doctors. However, if assignment is (conditionally) random then we would expect there to be no significant differences in the types of patients assigned to doctors of different quality.

[17] Boostrapped standard errors are clustered at the physician level.

Table 5: The relationship between residualised mean doctor outcomes and patient characteristics

	Residualised physician average 30-day survival			
	(1)	(2)	(3)	(4)
Original control variables				
Age	-0.0050***		-0.0049***	
	(0.0010)		(0.0010)	
Charlson index	-0.0003		-0.0004	
	(0.0010)		(0.0010)	
White	0.0020		0.0020*	
	(0.0012)		(0.0012)	
Male	0.0009		0.0009	
	(0.0009)		(0.0009)	
Hospital treatment cost (t-1)	0.0001		0.0010	
	(0.0008)		(0.0036)	
Stroke (t-1)	0.0001		0.0002	
	(0.0008)		(0.0009)	
Additional patient characteristics				
Local area deprivation		0.0015	0.0016	
		(0.0024)	(0.0024)	
Working age benefits		-0.0009	-0.0014	
		(0.0022)	(0.0022)	
House sales		-0.0002	-0.0001	
		(0.0010)	(0.0010)	
No. of emergency admissions (t-1)		0.0000	-0.0008	
		(0.0008)	(0.0034)	
No. of elective admissions (t-1)		0.0008	0.0010	
		(0.0009)	(0.0011)	
Days in hospital (t-1)		-0.0019**	-0.0018*	
		(0.0010)	(0.0010)	
Combined patient characteristics				
Predicted 30-day survival				0.0159***
				(0.0015)
Observations	510,249	510,249	510,249	510,249

Notes: (1) All variables residualised by regressing on hospital dummies and time variables, and standardised to mean zero and standard deviation one; (2) Local area deprivation is the LSOA's Index of Multiple Deprivation score measured in 2010; (3) Working age benefits is the number of working age benefit recipients living in the LSOA in November 2010; (4) Houses sales is the the number of house sales in the Middle Super Output Area in 2010; (5) All variables denoted 't-1' are measured in the previous financial year; (6) Predicted survival regresses a survival indicator on all of the variables included in the table, in addition to a complete set of interaction between age and sex dummies, and diagnoses dummies; (7) Bootstrapped standard errors clustered at the doctor level (199 repetitions).

The socio-economic characteristics include local area deprivation, the number of local working age benefit claimants and the annual number of local house sales.[18] Measures of past hospital treatment include total days spent in hospital, and the number of emergency and elective hospital admissions, in the previous financial year.

Appendix Figure A.5 shows that there is no clear correlation between any of the residualised characteristics and the average outcomes of assigned doctors. Column two in Table 5 shows the results from the regression analysis. Again, all coefficients are very small in magnitude and are generally statistically insignificant. The only statistically significant coefficient is for the number of days spent in hospital in the previous year, with sicker patients (those with a greater number of days spent in hospital) assigned to physicians with worse average outcomes. However, the magnitude is again very small: a one standard deviation increase in past length of stay is associated with a 0.002 standard deviation reduction in the average physician outcome.

Column three repeats the analysis when including all of the residualised characteristics together. The coefficients are largely unchanged, with the majority of coefficients not statistically significantly different from zero. Again, in the rare case that a correlation is statistically significant, the coefficient remains small in magnitude.

Column four presents results for predicted 30-day survival, based on the linear combination of the individual characteristics included in column three.[19] The results show that there is a positive and statistically significant relationship between predicted survival and the average survival outcomes of the assigned doctor. However, the magnitude is again small: a one standard deviation in predicted 30-day survival is associated with a 0.02 increase in the average doctor fixed effect. This is further illustrated in Appendix Figure A.6 which plots the mean standardised predicted survival and augmented estimates of doctor quality for the ventiles of doctor quality. While there is a small positive gradient in predicted survival as estimated quality improves, this is tiny compared to the gradient in actual outcomes.

One potential threat to conditional randomisation is that patients have pre-existing relationships with a cardiologist, who may then choose to be assigned to a patient in an emergency setting. In particular, cardiologists who have previously treated a given patient may have additional information about a patient that is unobserved in the data.[20] In Appendix Table A.2,

[18] Residence is located at the Lower Layer Super Output area (LSOA). This provides a small area around where the patient lives. In 2011, there were 32,482 LSOAs in England, with an average population of 1,500 people.

[19] I estimate predicted survival by regressing an indicator of 30-day survival on all the characteristics included in X_{it} and Z_{it}, and predicting survival for each patient using these estimated coefficients.

[20] 13.2% of patients are recorded as receiving inpatient treatment from the same cardiologist at some point

I drop all patients who have previously been treated (for any reason) by the assigned cardiologist to check whether excluding these previous matches alters the estimates. The results show that this makes very little difference: for year-long survival, a standard deviation in estimated quality is slightly larger at 3.9 percentage points, but this is not statistically significant from the baseline results. This indicates that selective matching between patients and doctors on the basis of past interactions does little to alter the results.

Taken together, these results suggest that the assumption of conditional random assignment of patients to doctors is highly plausible in this setting. For the vast majority of characteristics studied, there is no statistically significant relationship with the estimated quality of the physician that they are assigned to, and where the tests fail they are very small in magnitude.[21] Selective sorting between doctors and patients is therefore unlikely to substantially bias the estimates of doctor quality in this setting.

5.2.2 Exploring further threats to identifcation

The evidence presented above suggests that patients and doctors are randomly matched within hospitals. However, two main threats to identification remain. First, the estimates may be biased if doctors systematically move away from hospitals that are becoming worse in quality. The estimates of μ are identified by comparing the performance of doctors within the same hospital over the entire 13-year period. If doctors leave a hospital that is becoming worse over time, or is taking on an increasingly severe mix of patients, their observed performance will improve relative to colleagues at the hospital at a later period of time. In the baseline estimates, hospital performance and patient mix is assumed to be constant over time, with deviations captured by λ_{kt} in the error term. Staff movements could therefore lead to a correlation between observed doctor performance and the unobserved deviations in hospital performance over time.

One way to examine how problematic hospital-specific performance is for the estimates is to directly include hospital-year dummy variables in the estimation. Identification for doctor performance then arises from comparing the performance of doctors working in the same hospital in the same year, and by a connected set of doctors who work in multiple hospitals simultaneously. This makes the conditions for the connected set more restrictive, and limits the sample to a

prior to their heart attack admission date.

[21] Some small amount of sorting on the basis of observables (such as age) remains consistent with assumption 1, where assignment of patients to doctor is random conditional on hospital, time and patient observables. Indeed, hospitals might be expected to sort on the basis of easily observable characteristics but given limitations imposed by the availability of staff at any given time we would expect this correlation to be small in magnitude.

smaller number of doctors and hospitals. For this reason, I do not include hospital-year effects in the baseline estimates but instead present this as a robustness check. When implementing these restrictions, the connected set is reduced to include 1,430 doctors treating 273,558 patients across 134 hospitals.

Table A.3 in Appendix A shows the results of this exercise. In column 1, I repeat the baseline analysis for the smaller connected set, where I include time-invariant hospital fixed effects. This shows that the distribution of doctor quality is somewhat more widespread for this sample than in the baseline: a one standard deviation improvement in (adjusted) doctor quality when using the adjusted estimates leads to a reduction in one-year mortality of 4.3 percentage points. Column 2 repeats this estimate with time-varying hospital-year effects included. The estimated standard deviation in quality is slightly larger, at 4.6 percentage points. However, the difference between the estimates is not particularly large in magnitude and is not statistically significantly different from zero. This suggests that hospital-specific deviations in outcomes over time do not have a meaningful impact on the estimates of permanent differences in doctor quality.

The second remaining threat to identification relates to potential match effects between doctors and hospitals. In the baseline model, doctors and hospitals are assumed to have additively separable impacts on patient survival. This rules out a scenario where doctors are more effective when working in particular hospitals, either because they are better suited to the available technology in each hospital or because they work better with other staff at specific locations. This assumption has been routinely criticised within the larger literature that examines worker and firm effects in wages and productivity (Eeckhout and Kircher, 2011; Woodcock, 2015). If doctors systematically choose to work in hospitals where they are better matched then this would bias the estimates of their persistent quality in any setting.

To explore whether an assumption of additive separability is plausible, I consider two pieces of evidence. Following Card et al. (2013), I first re-estimate the model with a fully saturated model that includes separate dummy variables for each hospital-doctor pair instead of separate doctor and hospital fixed effects. If the match effects are quantitatively important, then this model should provide a much better statistical fit of the data, as measured by an increase in the estimated R-squared. Carrying out this exercise suggests that there is essentially no difference between the two models, with the R-squared rising from 0.217 in the baseline estimates to 0.218 in the model with hospital-doctor pairs included.

An alternative way of testing whether there is a match effect between doctors and hospitals is to examine the difference in patient outcomes across hospitals for doctors who are observed treating patients in multiple locations. In the absence of specific physician-hospital effects, the difference in mortality outcomes for patients treated by doctor j in two separate hospitals should be explained by differences in the characteristics of patients treated in each hospital, and the estimated hospital effect. I therefore test for evidence of match effects by examining variation in the differences between predicted and actual survival for each doctor across different hospitals.

I do this in the following way. First, I estimate a fixed effect for each hospital by regressing one-year survival on patient characteristics, and a full set of doctor and hospital fixed effects. Second, I predict survival for each patient (absent hospital or doctor fixed effects) and take the mean predicted survival rate for each doctor in every hospital that they work in. This is then adjusted by the estimated hospital fixed effect to calculate the hospital-specific predicted survival rate of patients treated by every doctor in each hospital they work in. Taking the difference between the observed and predicted survival rates for each doctor-hospital pair provides an indication of doctor performance in each hospital that they work in.

While performance is unlikely to perfectly correlate due to differences in sample sizes (and associated measurement error) across hospitals, doctors who perform well in one hospital would be expected to do similarly well in another if match effects are small. I therefore separately sort doctors on the basis of the differences between predicted and actual survival in the two hospitals where they treat the most patients. I then compare these doctor rankings across hospitals. In the absence of match effects, doctors should perform similarly across the different hospitals that they work in.[22]

Table 6 shows that this is indeed the case. The table plots the quartile of doctors working in the hospital where they treat the most patients ('Hospital 1') against the quartile of the same doctor when working in the hospital where they treat the second most number of patients ('Hospital 2'). Quartile 1 contains the 25% of doctors with the worst outcomes (after accounting for the characteristics of their patient and the hospital) while quartile 4 contains the 25% of doctors with the best outcomes. Each cell reports the fraction of doctors in a particular quartile in hospital 1 who appears in each quintile of performance in hospital 2. If performance at the

[22]This is similar in spirit to the event studies used by Card et al. (2013) that study how wages change as workers move from high to low wage firms. In this case, many doctors are simultaneously working in two hospitals and therefore do not 'move'. However, the difference in their performance across the two hospitals should be explained by the different patient mix that they treat in each hospital and any differences in the overall quality of the hospitals.

two hospitals were unrelated, one would expect 25% of each quartile in the lower ranked hospital to appear in each quartile of the highest ranked hospital. This is clearly not the case here: 47.7% of doctors who perform best in hospital 1 also appeared in the top quartile of hospital 2, while only 11.0% appeared in the lowest performing quartile. Similarly, 50.0% of those in the bottom quartile in hospital 1 were also found in the bottom quartile of hospital 2, while only 15.5% appeared in the top quartile.

Taken together, this evidence suggests that match effects between physicians and doctors are unlikely to be large in practice.

Table 6: Within-consultant correlation of differences between actual and predicted survival across hospitals

	Quartile in hospital 2			
Quartile in hospital 1	1	2	3	4
1	50.0	22.7	11.8	15.5
	(3.9)	(3.7)	(3.3)	(3.1)
2	24.8	32.1	27.5	15.6
	(3.4)	(3.9)	(3.4)	(3.4)
3	14.7	33.9	30.3	21.1
	(3.3)	(3.8)	(4.0)	(3.5)
4	11.0	11.0	30.03	47.7
	(2.8)	(3.3)	(3.7)	(4.2)

Notes: (1) Sample includes only doctors working in at least two hospitals; (2) Hospital 1 is the hospital where the consultant treats the most patients, and hospital 2 is the hospital where the consultant treats the second most patients; (3) Doctors are assigned to quartiles in each hospital on the basis of the difference between actual survival rates and predicted survival rates; (4) Patient survival is predicted by regressing one-year survival on age and sex dummies (and interactions between these), race, dummies for each component of the charlson index, treatment costs in the previous financial year, whether the patient suffered a stroke in the previous year, a set of dummy variables for the first two recorded diagnoses, day of the week, and an interaction between month and year of admission, plus a hospital fixed effect (estimated from a separate regression, controlling both for the set of controls listed in (4), doctor fixed effects and a full set of hospital fixed effects); (5) Bootstrapped standard errors clustered at the doctor level (199 repetitions).

5.2.3 Addressing small samples

As noted in Section 4.3, variation in the estimated doctor fixed effects may be overestimated if these estimates are based on small samples for individual doctors. If sample sizes for each doctor are too small, then individual estimates of quality will be unreliable due to statistical imprecision. This has been a common criticism of the practical use of estimates of teacher value-added to evaluate the performance of teachers (McCaffrey et al., 2009; Schochet and Chiang, 2013). I therefore now examine in detail whether my estimates of doctor quality are likely to be driven by statistical noise.

I examine this in two ways. First, I examine how the (adjusted) distribution of results change when I progressively restrict the sample to include only doctors who treat a large number of patients. Columns 1-3 of Table 7 show the results of this exercise. The first column shows the baseline results, when all doctors treating a minimum of 10 patients within the connected set are included. In this case, a standard deviation improvement in doctor quality is associated with a 3.6 percentage point reduction in mortality. In column 2, I increase this minimum number of patients to 50. This reduces the number of doctors and hospitals in the connected set (1,181 doctors across 122 hospitals). It does reduce the estimated magnitude of a standard deviation of doctor quality to 2.4 percentage points, although this difference is not statistically significantly different from the baseline estimate. The 90-10 ratio also falls from 8.5 to 6.1 percentage points due to shorter tails of the distribution. In column 3, I increase this further to a minimum number of 100 patients. This further reduces the connected set (796 doctors across 89 hospitals) but there is little additional impact on the estimated distribution: under the more restrictive set, a one standard deviation increase in doctor quality is associated with a 2.6 percentage point reduction in one-year mortality. Together, these results suggest that statistical noise is unlikely to be driving my qualitative conclusions.

Second, I implement an alternative shrinkage procedure. Bitler et al. (2021) show that standard shrinkage techniques can fail to fully account for statistical noise, as demonstrated by showing non-zero estimated effects of teachers on outcomes that they cannot feasibly impact (i.e. height). They find that a zero result is only recovered when using the covariance in effects from multiple years of classroom data to estimate the signal-to-noise ratio. This follows an approach by Kane and Staiger (2008) that can be implemented only for teachers who are observed working in multiple years or classrooms. In this case, the covariance between teacher residuals across years or classrooms can be used to estimate the signal component.

Such an approach is possible to implement in my setting, with doctors observed treating patients over multiple years. I therefore compute estimates using the method of Kane and Staiger (2008). This uses a two-step approach: first, I regress patient survival on my full set of controls and hospital fixed effects. This yields a residual for each doctor in each year, which I then regress on a full set of doctor fixed effects. I take from this estimates of σ_u and σ_e, which I use to estimate the shrinkage estimator in Equation (6) by applying these to the doctor fixed effects from the same regression.

Column 4 in Table 7 shows the results. Using this method does decrease the variation in doctor fixed effects: a one standard deviation increase in doctor quality reduces one-year mortality by 1.6 percentage points, or 13% of mean mortality. This is just under half the size of the baseline estimates. The tails of the distribution are also less extreme: moving from the 10th to the 90th percentile of the estimated distribution would reduce mortality by 4 percentage points as compared to 8.5 percentage points in the initial estimates. While smaller than the initial estimates, this suggests there remains substantial variation in the permanent differences in patient outcomes achieved by individual doctors.

Table 7: Estimated distribution of doctor quality under alternative empirical implementations

	Doctor fixed effects			
	Minimum number of patients per consultant			
	10 1-step (1)	50 1-step (2)	100 1-step (3)	10 2-step (4)
Std Deviation	0.036***	0.024***	0.026***	0.016***
	(0.004)	(0.003)	(0.003)	(0.001)
Variance	0.001***	0.001***	0.001***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
10th percentile	-0.040***	-0.030***	-0.031***	-0.020***
	(0.004)	(0.003)	(0.004)	(0.001)
25th percentile	-0.018***	-0.015***	-0.016***	-0.009***
	(0.002)	(0.002)	(0.003)	(0.001)
50th percentile	0.002	0.001	0.001	0.001
	(0.002)	(0.001)	(0.002)	(0.000)
75th percentile	0.023***	0.016***	0.018***	0.011***
	(0.002)	(0.002)	(0.003)	(0.001)
90th percentile	0.045***	0.031***	0.032***	0.021***
	(0.004)	(0.003)	(0.005)	(0.001)
Number of patients	510,249	451,603	329,276	510,249
Number of doctors	1,587	1,181	796	1,587
Number of hospitals	141	122	89	141

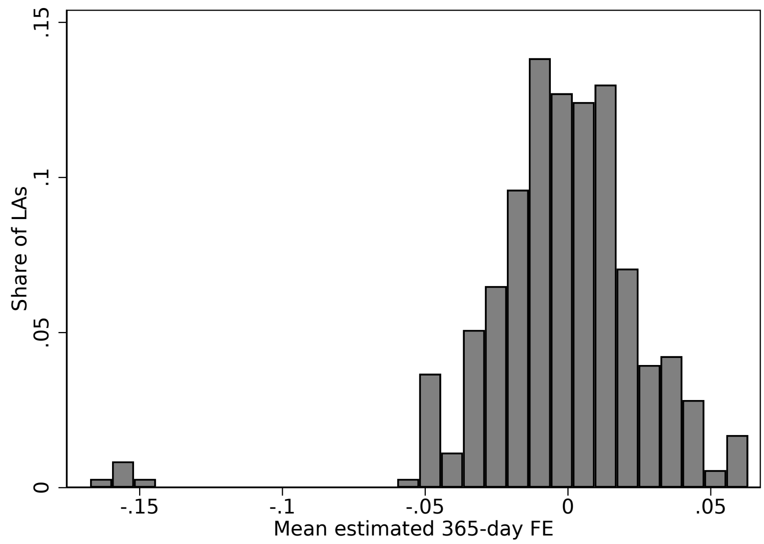
Notes: (1) All specifications include the same controls as the results in Table 3; (2) Column 4 uses the 2-step procedure outlined in Section 5.2.3; (3) Bootstrapped standard errors clustered at the doctor level (199 repetitions).

6 Geographic variation in doctor quality

The results so far show that cardiologists vary substantially in their effectiveness in treating heart attack patients. The evidence suggests that patients are randomly assigned to patients within hospital, but it does not rule out sorting of patients to doctors of different quality across hospitals. As a result, there may be geographic inequalities in access to the same quality of care. In this section, I therefore use my estimates to examine how doctor quality varies across different areas in England. Importantly, these estimates come from within-hospital comparisons of the performance of different doctors: variation in patient mix across areas should therefore have no impact on the estimates of average doctor quality across these areas.

I first examine variation in mean doctor quality across 354 local government areas in England.[23] Figure 2 shows the distribution of the mean estimated doctor fixed effects on 365-day survival across these areas. This reveals that there is substantial variation in the average quality of care received by heart attack patients in different areas of the country, with a standard deviation equal to 3.0 percentage points. Residents in a small number of areas received treatment from cardiologists with very large negative estimated effects.

Figure 2: Distribution of average estimated doctor quality across local areas, 2005 - 2017



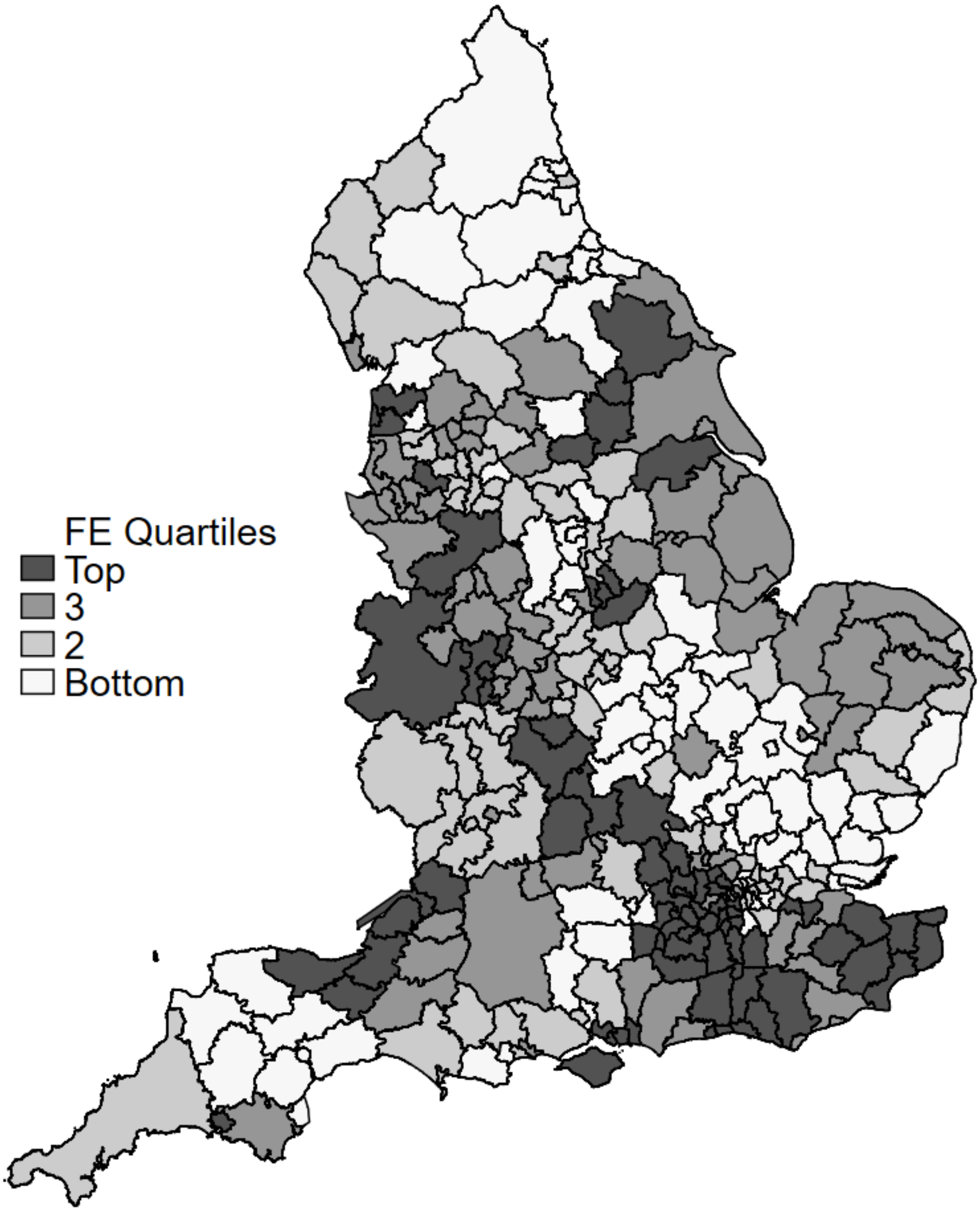
Notes: (1) Estimated doctor fixed effects on patient survival over 365 days (Table 3, column 2) averaged across all patients living in each 2011 local authority district between April 2005 and March 2017; (2) Areas (City of London and Isles of Scilly) with small sample sizes (fewer than 150 patients) are excluded.

Figure 3 maps the geographic distribution of these estimated effects. Areas are sorted into quartiles, with darker colours indicating areas in which average doctor quality is estimated to be higher. The figure shows that, in general, the most effective doctors treated patients in a

[23] Average effects are for the period between 2005 and 2017, and are defined using the area in which the patient lives but not necessarily where they are treated. I use 'Local Authority Districts' defined by their 2011 borders. These are administrative areas in England that are responsible for running many local services (e.g. garbage collection, long-term care, libraries etc) but do not organise health services, which are provided and funded centrally by the National Health Service.

band of areas running diagonally across the country, from the north west to the south east (including London). The doctors that were estimated to be least effective generally treated patients in the north and eastern parts of the country.

Figure 3: Average estimated doctor quality by local area, 2005 - 2017



Notes: (1) Estimated doctor fixed effects on patient survival over 365 days (Table 3, column 2) averaged across all patients treated in each 2011 local authority district between April 2005 and March 2017; (2) Areas (City of London and Isles of Scilly) with small sample sizes (fewer than 150 patients) are excluded.

To summarise this pattern, I regress, at the patient level, the estimated doctor fixed effect on a set of nine regional dummies indicating the area in which the patient lives.[24] Appendix

[24]Areas are UK 'Government Office Regions'. Bootstrapped standard errors clustered at the doctor level.

Table A.4 shows the results, reporting the difference in average estimated doctor quality in each region relative to London. The average doctor fixed effect in the North East was 1.0 standard deviations lower than in London. This is equivalent to a difference of 3.6 percentage points in survival over a year, or just under a quarter of the mean 365-day mortality rate. For the East of England, this difference was 0.6 standard deviations (2.1 percentage points). No other coefficients when studying one-year mortality were statistically significant.

These patterns mirror the general patterns of deprivation in England. This raises the prospect that geographic sorting of doctors results in differential access on the basis of quality across socioeconomic groups. The patterns also highlight differences in the quality of doctors treating patients in rural and urban areas. There is a clear concentration of better doctors treating patients in major cities (such as London, Birmingham and Manchester). In contrast, the areas in the bottom quartiles are typically rural areas.

To further explore these patterns, I examine the correlation of average doctor quality with a summary measure of deprivation and rurality at a finer local area level. I assign 32,448 small areas into seven groups based on their 2001 Census Area Classification.[25] These categories include: countryside, professional city life, fringe urban, white collar urban, multicultural, disadvantaged urban and miscellaneous urban areas. Averaging the estimated doctor fixed effects for all patients living in each of these areas, I regress these averages on a set of dummy indicators for these groups.[26]

Table 8 reports the results, with all differences in average doctor quality in the area reported relative to "countryside areas". The results indicate that average estimated doctor quality is substantially higher in less deprived, urban areas: residents in 'professional city life' areas were treated by doctors who were rated 0.3 standard deviations higher in quality, residents in urban fringe areas were treated by doctors who were rated 0.2 standard deviations higher, and residents living in miscellaneous urban areas (including suburban areas) were treated by doctors who were rated 0.10 standard deviations higher, than patients living in rural areas. In contrast, there

[25] I use 2001 Local Layer Super Output Areas (LSOAs), as used in the analysis in Table 5. Areas are grouped using K-means clustering to represent the main dimensions of the census: demographic structure, household structure, socio-economic character, employment and industry. For a detailed methodology, see Office for National Statistics (2007).

[26] Appendix Table A.5 shows the results when correlating these averages with other measures of deprivation and rurality. Column one shows a negative deprivation gradient using the 30-day mortality scores, with patients living in the least deprived quartile of areas treated, on average, by doctors who were rated 0.2 standard deviations higher than those treating patients in the most deprived quartile. The coefficients in column two show the same pattern using the 365-day mortality scores but these differences are no longer statistically significant. The coefficients in column three and four suggest that average estimated quality is about 0.1 standard deviations lower in rural areas but these differences are not statistically significant.

were no statistically significant differences with more deprived urban areas, and the estimated coefficients are close to zero.

Table 8: The relationship between estimated doctor quality and area type

	Average doctor fixed effects
Professional city	0.302**
	(0.153)
Fringe urban	0.171**
	(0.082)
White collar urban	0.074
	(0.061)
Multicultural	0.203
	(0.149)
Disadvantaged urban	0.021
	(0.095)
Miscellaneous urban	0.124*
	(0.070)
Observations	32,240

Notes: (1) The outcome is the mean estimated doctor fixed effects on patient survival over 365 days for all patients living in the LSOA between 2005 and 2018; (2) Area type assigned based on the LSOA of patient residence, the omitted category is "countryside areas"; (3) Bootstrapped standard errors clustered at the doctor level (199 repetitions).

Taken together, these results indicate that there are meaningful differences in the quality of care that patients living in different parts of the country receive. This has the potential to compound differences in the levels or availability of any care – for example, differences in waiting times (so called 'postcode lotteries') are well documented within England.[27] If people in more deprived communities also receive worse quality care this has the potential to exacerbate pre-existing health inequalities across socioeconomic groups. This is significantly at odds with the stated equity aims of the NHS, and suggests that policymakers should carefully consider how it can ensure that patients living in more deprived areas can gain access to high quality doctors.

7 Reallocating doctors to reduce mortality

As noted in Section 3, the baseline model restricts doctors to be equally effective at treating all patients, regardless of the underlying characteristics of patients. However, doctors may vary in their ability to treat different patients based on the severity of their condition, or their wider characteristics. In particular, there are multiple ways to treat heart attack patients, with some

[27]See, for example, https://www.theguardian.com/society/2021/sep/01/huge-discrepancy-nhs-england-waiting-times-common-procedures

treatment types more suitable for patients with particular symptoms. Previous work has shown that doctors often prefer to treat patients for heart attacks in a particular way, regardless of the suitability of such treatment in the case of that specific patient (Chandra and Staiger, 2007; Currie et al., 2016). More generally, some doctors may be more effective when treating different patient types. As a result, patient outcomes could potentially be improved if patients were reassigned to doctors who perform best when treating patients of their particular type.

To examine this further, I modify the model to allow doctors to vary in the quality of care that they provide to patients on the basis of their exact diagnosis. I then examine the potential reductions in patient mortality that could be achieved by reallocating the existing set of doctors across patients of different types under different scenarios.

7.1 Varying doctor quality across patient types

To examine whether doctors vary in their effectiveness when treating patients with different needs, I estimate the quality of each doctor when treating patients of two different types. Specifically, I split patients on the basis of whether or not they are experiencing a particular type of heart attack: a ST-elevation myocardial infarction (STEMI). STEMI patients have higher mortality rates in the short-run, require immediate treatment, and are generally treated in different ways.[28] STEMI patients are much more likely to undergo a procedure rather instead of being treated by medical management alone: for example, in 2017 only 7.9% of STEMI patients received no recorded intervention, compared to 32.6% of non-STEMI patients. More specifically, STEMI patients are much more likely to be treated using percutaneous coronary intervention (PCI): in 2017, 77.0% STEMI patients underwent a PCI compared to 26.3% of non-STEMI patients. Doctors may therefore vary in their ability in providing different types of treatment, while the suitability of this treatment may vary across patients.

Using this distinction, I modify Equation (3) to estimate the following specification:

$$Y_{ijkt}^{s} = \beta X_{it} + \mu_j^s + \delta.Exp_{jt}^{s} + \psi_k + \epsilon_{ijkt}, \quad \forall s = [N, S] \tag{8}$$

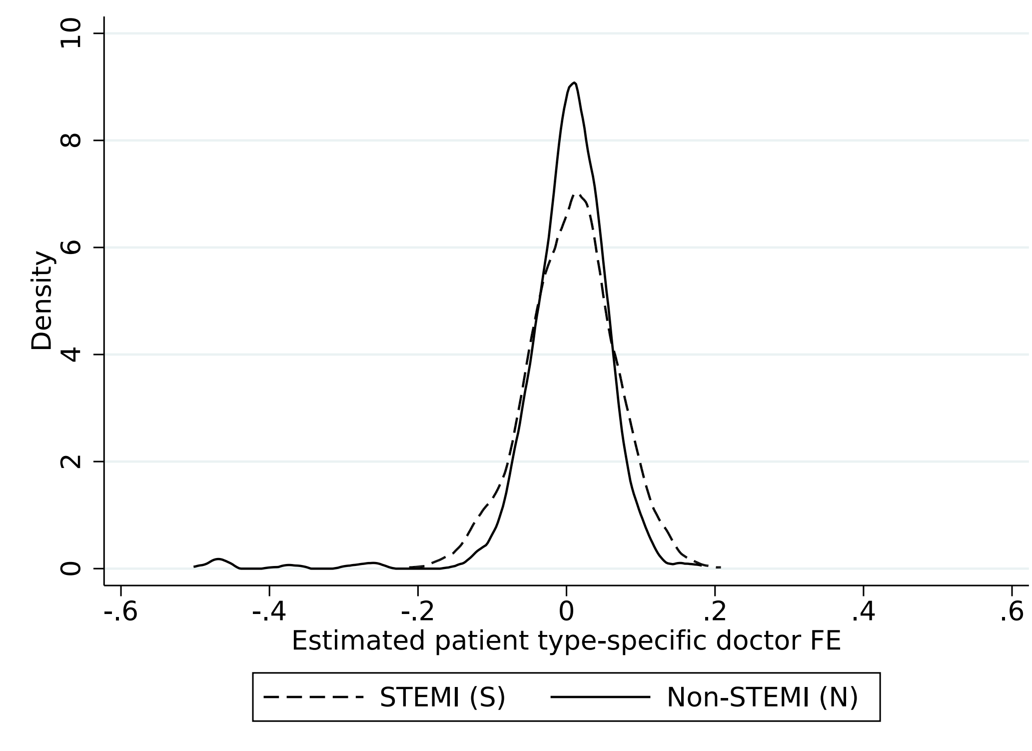
Y_{ijkt}^{s} is a binary variable that measures survival over 365 days for a patient of type s. μ_j^s now measures the permanent difference in patient outcomes for doctor j when treating a patient

[28] 30-day mortality in 2017 was 8.1% for STEMI patients compared to 4.5% for non-STEMI patients, while 365-day mortality was roughly equal among the two groups (12.7% for STEMI patients vs 12.4% for non-STEMI patients).

of type s, and δ captures the experience profile from treating patients of type s . All other elements of the specification are the same as in Equation (3).

This yields two estimated fixed effects for each doctor (μ_j^N, μ_j^S).[29] Figure 4 shows the distribution of the estimated effects for each patient type. The figure shows that there is slightly greater variation in the performance of doctors when treating non-STEMI patients: improving quality by one standard deviation when non-STEMI patients is equivalent to a reduction in mortality of 6.7 percentage points compared to a reduction in mortality of 6.0 percentage points when increasing doctor quality by one standard deviation when treating STEMI patients.

Figure 4: Distribution of estimated doctor fixed effects, by patient type



Notes: (1) Estimates from Equation (8) are adjusted using the one-step method used for the baseline results shown in Table 3.

Appendix Figure A.7 shows there is a weak, positive correlation (0.13) between the two estimated fixed effects for each doctor. On its face, this is a somewhat surprising result, and one might expect the most effective doctors to achieve the best results for all patients.[30] However,

[29] 75 doctors never treat STEMI patients, while 17 patients never treat non-STEMI patients. To simplify the reallocation exercise below I assign a fixed effect of zero for these doctors when treating each type of patients. However, this makes no practical difference to the results as all STEMI patients are ultimately reallocated to doctors with a positive fixed effect when treating STEMI patients.

[30] There is some evidence of this: Appendix Table A.6 shows that doctors ranked in the top quintile when treating non-STEMI patients are also more likely to be ranked in the top quintile, with 33.2% of doctors ranked in the top quintile of the non-STEMI distribution appearing in the highest quintile of the STEMI distribution

as noted above, the tasks carried out by doctors in treating each patient type are quite different, with more intensive interventions (and in particular the use of PCI) used to treat STEMI patients. Chandra and Staiger (2007) showed that among US cardiologists, some doctors were much better at performing PCI, while others were better at medical management. A similar pattern of skills in this setting could therefore explain the differences in performance when treating patients where these treatment types are most appropriate.[31]

7.2 Reallocating doctors across patient types

I now consider different scenarios where doctors can be matched to specific patient types on the basis of their comparative advantage in treating each patient type. Intuitively, matching doctors who are comparatively better at treating STEMI patients to such patients could substantially reduce mortality among heart attack patients.

The estimated comparative advantage of doctor j in treating a STEMI patient rather than a non-STEMI patient is simply the difference between the estimated fixed effects (i.e. $\hat{\mu}_j^S - \hat{\mu}_j^N$). I use the estimates from Equation (8) to estimate the comparative advantage for each doctor. I then rank doctors according to this difference, with those with the greatest comparative advantage in treating STEMI patients ranked higher. These rankings enable me to match doctors with particular patients under different constraints.

I consider three scenarios. First, I consider the potential reductions in mortality if the worst performing 10% of doctors were replaced with a doctor that had zero impact. This is similar to previous experiments in the teacher value-added literature (Hanushek, 2009; Chetty et al., 2014b). This provides a baseline case of mortality reductions that could be reduced by simply replacing the worst performing doctors with better performing ones.

However, simply replacing doctors would be hard in practice, with long training periods required. I therefore examine two further scenarios where I reallocate existing doctors across patients. This includes an 'unconstrained' and 'constrained' exercise. In the 'unconstrained' reallocation exercise, I hold the caseload of each doctor constant and reallocate doctors towards patients of particular types regardless of the hospital in which these doctors work. Specifically, I allocate STEMI patients to doctors with the highest marginal benefit, allocating patients first to the doctor with the highest comparative advantage in treating STEMI patients, and then to

(compared to 25% if this was drawn at random).

[31] This is also consistent with Epstein et al. (2010), who find that the estimated effectiveness of obstetricians when using two different delivery methods is only weakly positively correlated.

the next highest ranked doctor until all STEMI patients are allocated. All remaining patients are assigned to unallocated doctors. This scenario is unconstrained in the sense that I do not place limits on the location where doctors can treated patients: STEMI patients can be located in many hospitals in which the doctor does not work in practice. This therefore represents an upper bound for reductions in mortality that could be achieved by reallocating existing staff, but would not be feasible to implement in reality.

I also examine a third scenario. In this 'constrained' reallocation exercise, doctors can now only be reassigned to patients within a hospital that they currently work in.[32] This is a more feasible scenario in the sense that doctors would be better able to access these patients.[33] In this case, doctors are ranked on their comparative advantage within-hospital. Patients are then reassigned to doctors in the same way within hospital.

Under the first scenario where I simply replace the worst performing doctors, the number of deaths within a year for patients treated between April 2017 and March 2018 would be reduced by 310, a 4.8% reduction from the 6,921 deaths observed in the data. This would reduce the mortality rate from 12.5% to 12.0%. Such a change is equivalent to improving the average doctor quality by 0.1 standard deviations as reported in the baseline results (Table 3).

In the 'unconstrained' scenario, matching doctors to patients reduces the number of deaths by 1,467 in 2017-18, or a reduction of 21.2% on the observed number of deaths. This reduces the mortality rate from 12.5% to 9.9%. This is equivalent to improving average doctor quality by 0.7 standard deviations, and represents a substantial improvement in patient outcomes.

The 'constrained' scenario also produces substantial improvements in patient outcomes compared to observed mortality rates, albeit smaller than in the scenario when doctors can be reallocated to patients anywhere. The estimates indicate that allocating doctors to patients within-hospital could reduce deaths in 2017-18 by 518 patients (7.5%). This is equivalent to a reduction in the one-year mortality rate from 12.5% to 11.6%. The same improvement could be made by raising the average doctor quality by 0.3 standard deviations.

Taken together, the results suggest that substantial improvements in patient survival rates could be achieved by reallocating doctors towards patients to which they are better matched with. In practice, these gains may be hard to fully achieve given the time constraints inherent

[32]For example, if a doctor treats 10 patients in hospital A and 15 patients in hospital B in the observed data in 2017, they will have the same caseloads in the simulation.

[33]I do not constrain the number of days, or the time between patients. In practice doctors would not be available at all times. A further restriction would therefore be to reassign doctors to particular days or shifts where a greater mix of patients are of a particular type.

in treating heart attack patients. However, these exercises do suggest that some gains could be made by rescheduling the shifts of senior doctors or moving doctors across hospitals to better match doctors and patients.

8 Conclusion

Variation in patient outcomes across places and providers is increasingly becoming a concern for policymakers around the world. These differences in the quantity and quality of care provided to different patients risk exacerbating already existing health inequalities. Medical staff - and doctors in particular - are likely to play a key role in driving this variation. Understanding this role in greater detail is therefore an important step towards addressing these inequalities.

In this paper, I exploit features of the English National Health Service to estimate the quality of individual doctors. My results indicate that there is substantial variation in the quality of individual doctors when treating heart attack patients: a standard deviation improvement in doctor quality reduces mortality by 3.6 percentage points, or by a quarter of mean mortality rates, over a year.

Using these estimates, I show that average doctor quality varies substantially across regions. For example, patients in London received treatment from cardiologists with estimated average fixed effects almost a standard deviation higher than patients in the North East. More generally, the results suggest that the most effective doctors are treating patients in more affluent, urban areas. This raises the potential of large inequalities on the basis of quality of care even within a universal health system with explicit equity objectives. Policymakers should therefore carefully explore ways to better monitor these differences in quality, and design incentives to ensure that effective doctors are willing to work across the country.

I then extend this model to show that doctors vary in their quality when treating patients of different types. Dividing patients into those with and without a specific type of heart attack, I estimate the comparative advantage of each doctor in treating each type. I use these estimates to study several scenarios where doctors are replaced or reassigned across patients. This exercise reveals that there are large potential gains for patients in reallocating doctors: reallocating doctors across patients within the hospitals that they work in to better match with the most appropriate patients is estimated to reduce mortality by 8%.

These findings have a number of implications. First, the results suggest that patient outcomes could be improved by reallocating existing staff resources to treat certain patients. Even within a narrow field of medicine, doctors require different skills in treating different patients. Matching doctors to the appropriate patients therefore offers significant efficiency gains. Such gains are consistent with evidence of potential improvements from better matching of doctors with specific patient types in primary care (Dahlstrand, 2022). However, reallocating these doctors in this setting may face significant practical challenges. In this work, I have abstracted from the other tasks that senior doctors are doing: reallocating doctors across hospitals or shifts may take them away from other tasks that they perform well. Similarly, even the exercise that considers reallocating doctors within hospitals does not take into account the difficulties of scheduling doctors to be available at the times when particular patients enter the hospital.[34] Future work could consider whether patient types vary systemically across different times or days of the week, with doctors reassigned to shifts where the greatest number of appropriate patient types are likely to attend, or consider interventions that reassign doctors across hospitals to better match with the hospital's patient mix.

Second, the results show doctors vary considerably in the results they achieve, even when they are already very experienced and highly qualified. This variation is likely to be even greater among more junior staff in earlier parts of their career. Future work should concentrate on studying why these differences emerge. Access to information on the earlier careers of these doctors, including histories of where and who they have worked with, is essential in understanding the formation of skills over the course of an entire career. Previous work suggests that peer effects are important in the formation of doctor habits (Molitor, 2018; Doyle and Staiger, 2021), and any future extension of this work should consider the scope for peer effects in this setting. Such work would help to answer the question of what makes a 'good' doctor, and could ultimately lead to improvements in the way that doctors are recruited and trained.

[34] In practice, hospitals would ideally schedule two cardiologists to be available at any given time, with one having a comparative advantage in treating STEMI patients and the other a comparative advantage in treating non-STEMI patients.

References

Aaronson, D., L. Barrow, and W. Sander (2007). Teachers and student achievement in the chicago public high schools. *Journal of Labor Economics 25*(1), 95–135.

Abowd, J. M. and F. Kramarz (1999). The analysis of labor markets using matched employer-employee data. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3, pp. 749–804. Amsterdam: North-Holland.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Bitler, M., S. Corcoran, T. Domina, and E. Penner (2021). Teacher effects on student achievement and height: A cautionary tale. *Journal of Research on Educational Effectiveness 14*, 900–924.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer employee data. *Econometrica 87*(3), 699–738.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *Quarterly Journal of Economics 128*(3).

Chan, D., M. Gentzkow, and C. Yu (2022). Selection with variation in diagnostic skill: Evidence from radiologists. *Quarterly Journal of Economics 137*, 729–783.

Chandra, A. and D. O. Staiger (2007). Productivity spillovers in healthcare: Evidence from the treatment of heart attacks. *Journal of Political Economy 115*, 103–140.

Chen, Y. (2021). Team-specific human capital and team performance: Evidence from doctors. *American Economic Review 111*(12), 923–62.

Chetty, R., J. Friedman, and J. Rockoff (2014a). Measuring the impacts of teachers i: Evaluation bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

Chetty, R., J. Friedman, and J. Rockoff (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 2633–2679.

Currie, J. and W. B. Macleod (2020). Understanding doctor decision making: The case of depression treatment. *Econometrica 88*(3), 847–878.

Currie, J., W. B. Macleod, and J. V. Parys (2016). Practice style and patient health outcomes: The case of heart attacks. *Journal of Health Economics 47*, 64–80.

Currie, J. and J. Zhang (2021). Doing more with less: Predicting primary care provider effectiveness. *NBER Working Paper 28929*.

Cutler, D., J. S. Skinner, A. D. Stern, and D. Wennberg (2019). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *American Economic Journal Economic Policy 11*(1), 192–221.

Dahlstrand, A. (2022). Defying distance? the provision of services in the digital age. Mimeo.

Doyle, J. J. (2020). Physician characteristics and patient survival: Evidence from physician availability. *NBER Working Paper* (27458).

Doyle, J. J., S. M. Ewer, and T. H. Wagner (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of Health Economics 29*(6), 866–882.

Doyle, J. J., J. Graves, J. Gruber, and S. Kleiner (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy 123*(1), 170–214.

Doyle, J. J. and B. Staiger (2021). Physician group influences on treatment intensity and health: Evidence from physician switchers. NBER Working Paper 29613.

Dranove, D., K. Daniel, M. McCellan, and M. Satterthwaite (2003). Is more information better? The effects of "Report Card" on health care providers. *Journal of Political Economy 111*(2), 555–588.

Eeckhout, J. and P. Kircher (2011). Identifying sorting: In theory. *Review of Economic Studies 78*(3).

Epstein, A. J., J. D. Ketcham, and S. Nicholson (2010). Specialization and matching in professional services firms. *The RAND Journal of Economics 41*(4 (Winter 2010)), 811–834.

Epstein, A. J. and S. Nicholson (2010). The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics 28*, 1126–1140.

Fadlon, I. and J. V. Parys (2020). Primary care physician practice styles and patient care: Evidence from physician exits in medicare. *Journal of Health Economics 41*.

Finkelstein, A., M. Gentzkow, and H. Williams (2016). Sources of geographic variation in health care: Evidence from patient migration. *Quarterly Journal of Economics 131*(4), 1681–1726.

Finkelstein, A., M. Gentzkow, and H. Williams (2021, August). Place-based drivers of mortality: Evidence from migration. *American Economic Review 111*(8), 2697–2735.

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

Ginja, R., J. Riise, B. Willage, and A. Willen (2022). Does your doctor matter? doctor quality and patient outcomes. CESifo Working Paper No 9788.

Glance, L., A. Dick, D. Mukamel, Y. Li, and T. Osler (2008). Are high-quality cardiac surgeons less likely to operate on high-risk patients compared to low-quality surgeons? evidence from new york state. *Health Services Research 43*(1), 300–312.

Gowrisankaran, G. and R. Town (1999). Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics 18*(6), 747–767.

Hanushek, E. (2009). Teacher deselection. In D. Goldhaber and J. Hannaway (Eds.), *Creating a New Teaching Profession*, pp. 165–180. Washington, DC: Urban Institute Press.

Hull, P. (2020, January). Estimating hospital quality with quasi-experimental data. Working Paper.

Kane, T., J. Rockoff, and D. Staiger (2008). What does certification tell us about teacher effectiveness? evidence from new york city. *Economics of Education Review 27*, 615–631.

Kane, T. and D. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *NBER Working Paper 14607.*

Kwok, J. (2019, July). How do primary care physicians influence healthcare? evidence on practice styles and switching costs from medicare. Working Paper.

Lundborg, P., S. James, B. Lagerqvist, and J. Vikstrom (2021, September). Learning-by-doing and productivity growth among high-skilled workers: Evidence from the treatment of heart attacks. IZA DP No. 14744.

Mansfield, R. K. (2015). Teacher quality and student inequality. *Journal of Labor Economics 33*(3), 751–788.

Marshall, M., P. Shekelle, S. Leatherman, and R. Brook (2000). The public release of performance data: what do we expect to gain? a review of the evidence. *Journal of the American Medical Association 281*, 2098–2105.

Maxwell, S. (1999). Emergency management of acute myocardial infarction. *Br J Clin Pharmacol 48*(3), 284–298.

McCaffrey, D. F., T. R. Sass, J. R. Lockwood, and K. Mihaly (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy 4*, 572–606.

Molitor, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy 10*(1), 326–356.

Office for National Statistics (2007). Area classification of super output areas and datazones. project final report. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20160128205034/http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/ns-area-classifications/index/methodology-and-variables/super-output-areas/super-output-areas.html (last accessed 19 May 2020).

Schochet, P. Z. and H. S. Chiang (2013). What are error rates for classifying teacher and school performance using value-added models? *Journal of Educational and Behavioral Statistics 38*, 142–171.

Skinner, J. (2011). Causes and consequences of regional variations in health care. In *Handbook of Health Economics*, Volume 2, Chapter 2, pp. 45–93. Elsevier.

Van Parys, J. (2016). Variation in physician practice styles withinand across emergency departments. *PLoS ONE 11*(8).

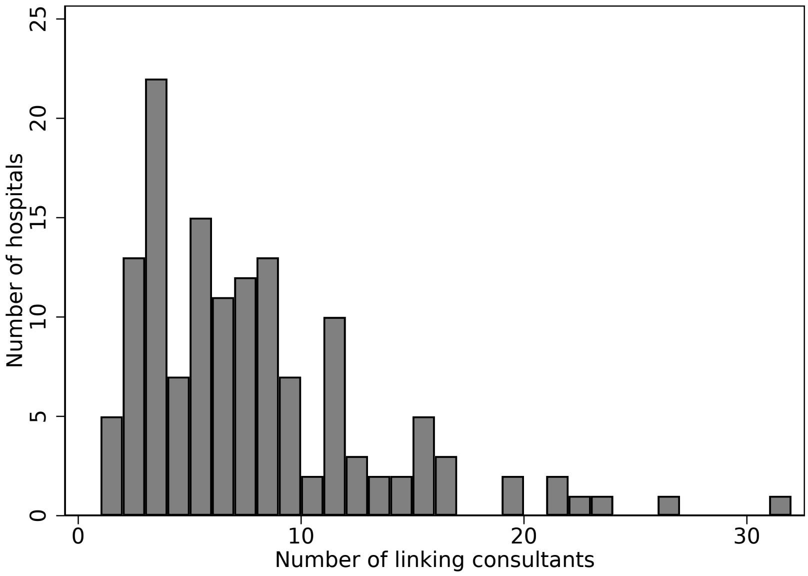
Windecker, S. (2014). 2014 esc/eacts guidelines on myocardial revascularization: The task force on myocardial revascularization of the european society of cardiology (esc) and the european association for cardio-thoracic surgery (eacts) developed with the special contribution of the european association of percutaneous cardiovascular interventions (eapci). *European Heart Journal 35*(37), 2541–2619.

Woodcock, S. (2015). Match effects. *Research in Economics 69*(1).

Appendices

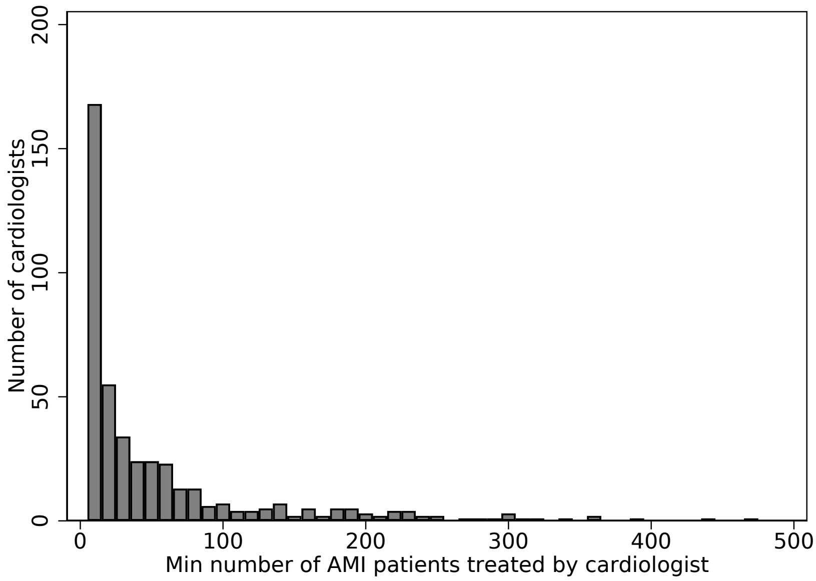
Appendix A Additional figures and tables

Figure A.1: The distribution of linking consultants across hospitals



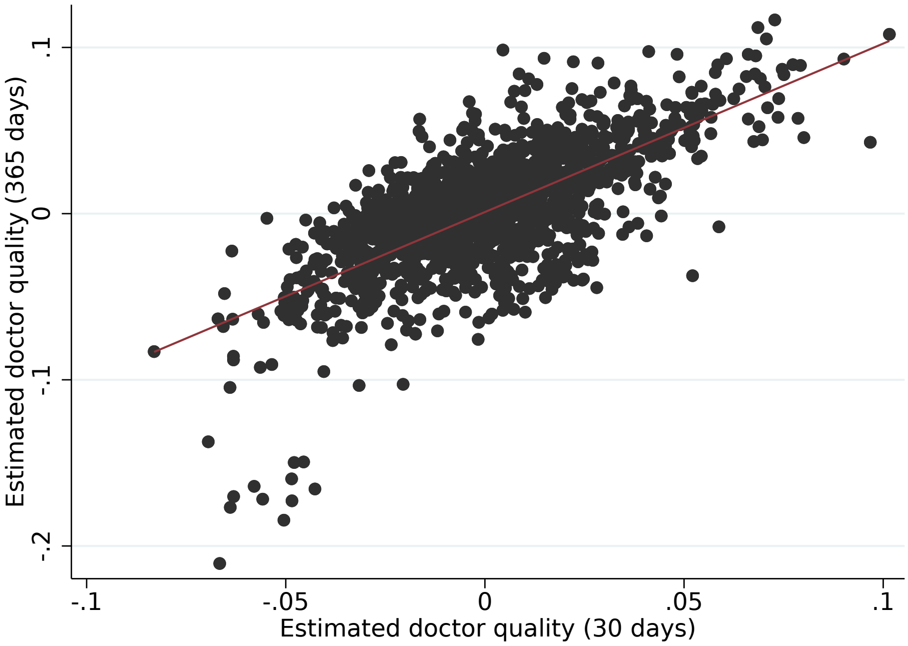
Notes: (1) Observations are at the hospital level; (2) Number of linking cardiologists indicate the number of consultants working at the hospital who treated at least 10 patients in another hospital within the connected set between April 2005 and March 2018.

Figure A.2: Patient distribution of linking consultants in 'secondary' hospitals



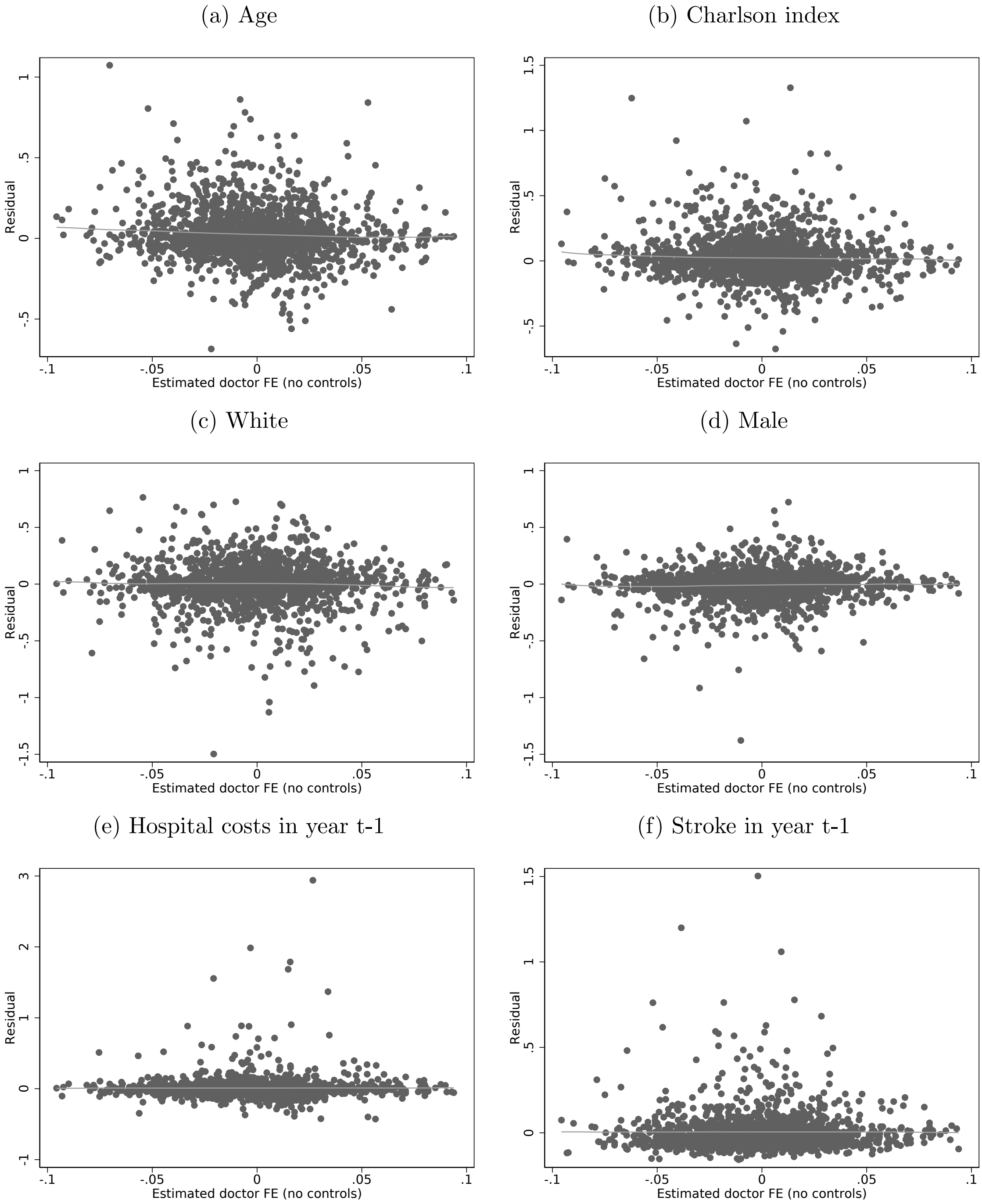
Notes: (1) Observations at the consultant level; (2) The sample includes all linking cardologists who treat patients in multiple hospitals; (3) 'Secondary Hospitals' are defined as the hospital in which the consultant treats the fewest patients over the period.

Figure A.3: The correlation between estimated doctor impacts over 30 and 365 days



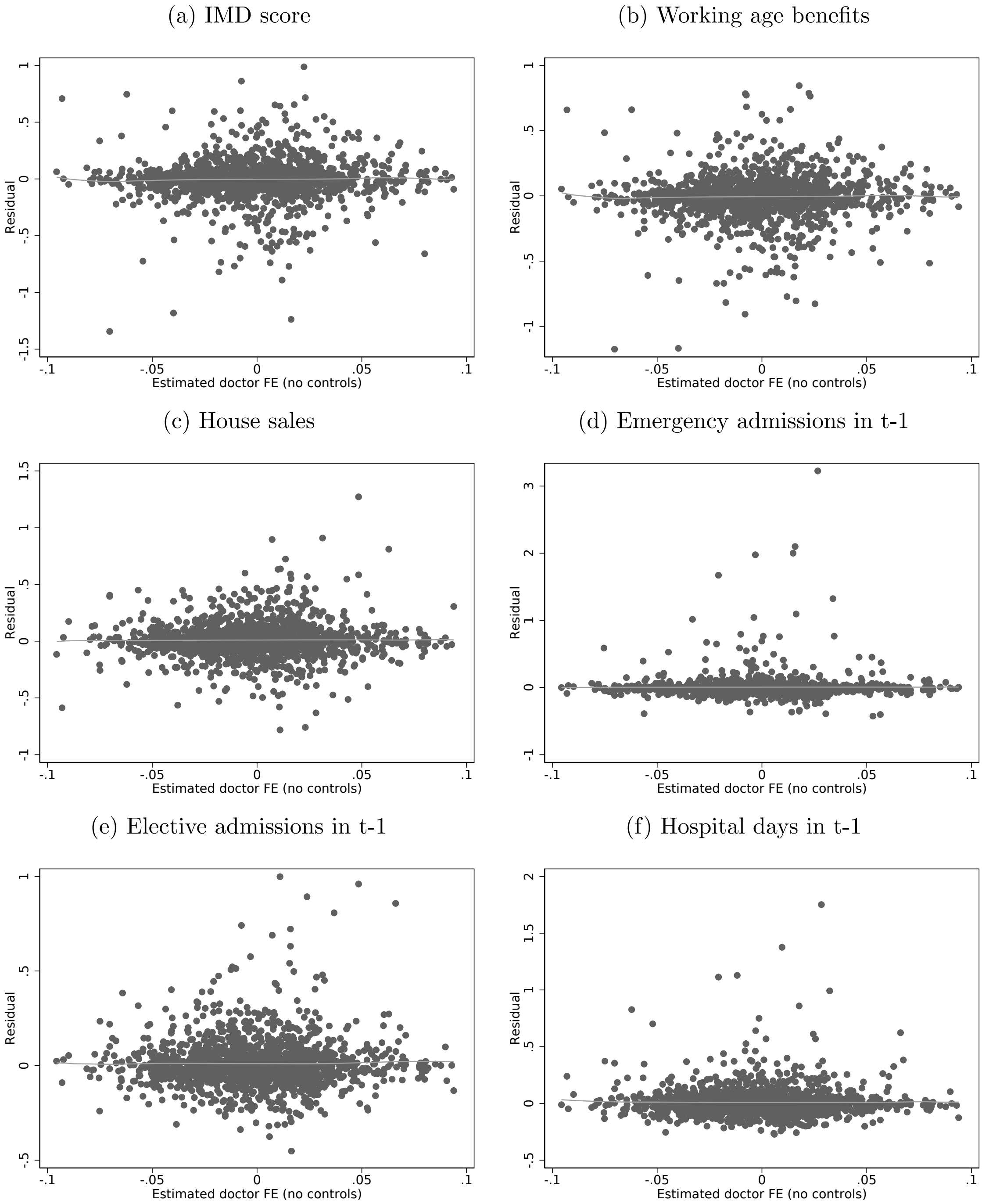
Notes: (1) Doctor fixed effects from same regression as described in Table 3

Figure A.4: Correlation between estimated doctor effects and mean patient characteristics



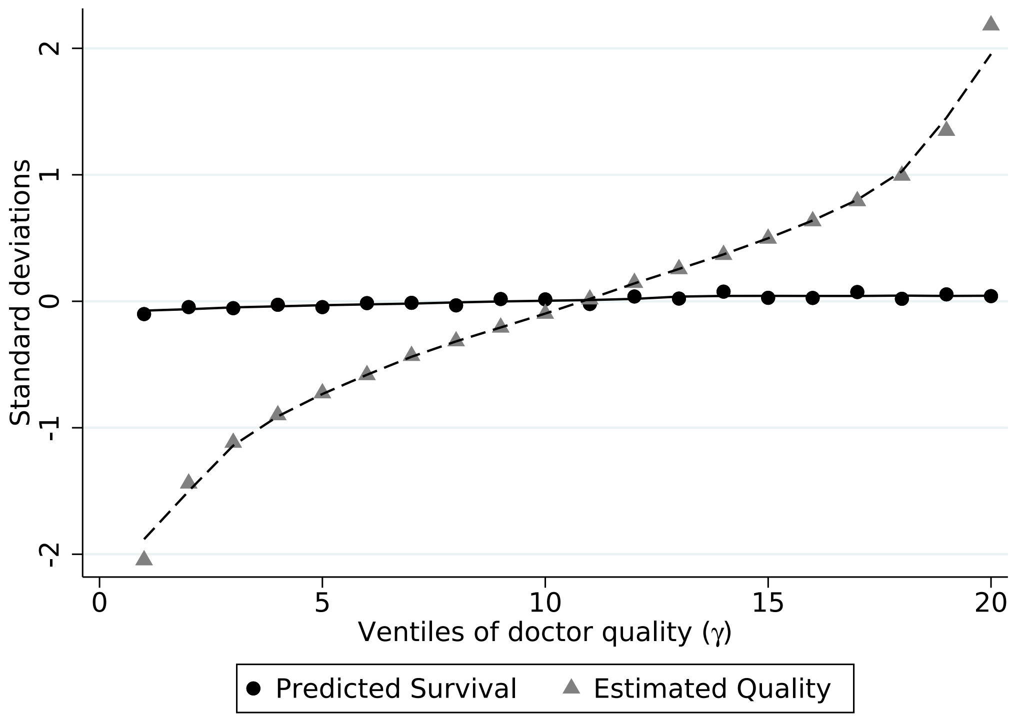
Notes: (1) All outcomes are the mean standardised residuals (at the doctor level) from an initial regression of the outcome of interest on hospital dummies, day of the week, and an interaction between month and financial year; (2) Panel A uses patient age, Panel B uses the Charlson index, Panel C uses an indicator equal to one if the patient is recorded as white, Panel D uses an indicator equal to one if the patient is male, Panel E uses total hospital cost in the previous year, and Panel F uses an indicator equal to one if the patient received treatment for a stroke in the previous year.

Figure A.5: Correlation between estimated doctor effects and additional mean patient characteristics



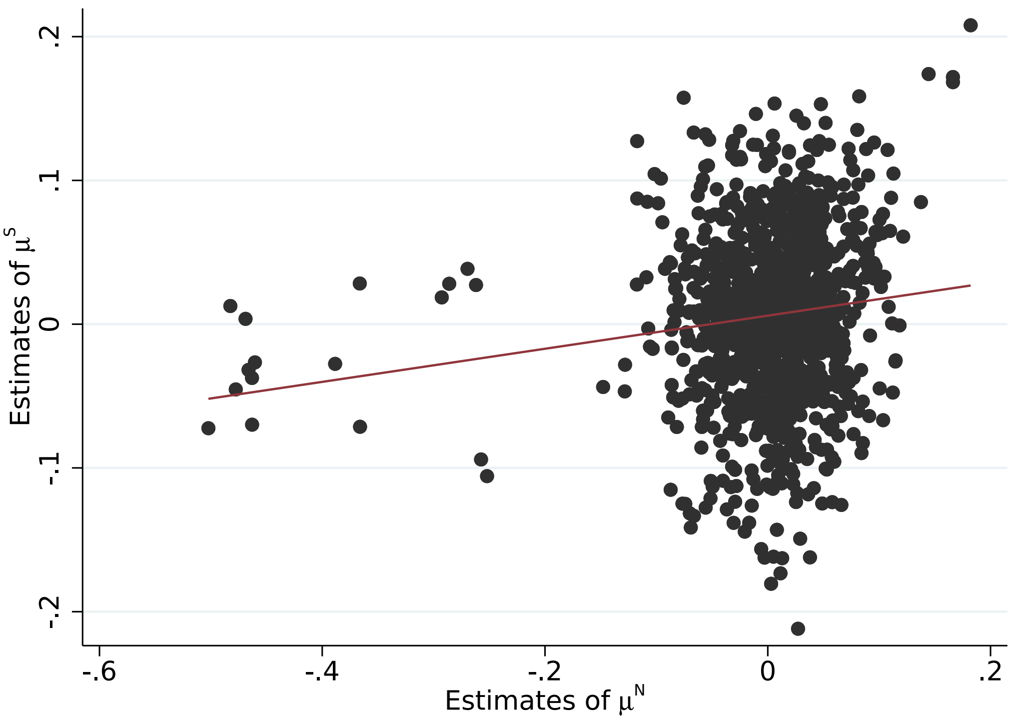
Notes: (1) All outcomes are the mean standardised residuals (at the doctor level) from an initial regression of the outcome of interest on hospital dummies, day of the week, and an interaction between month and financial year; (2) Panel A uses the Index of Multiple Deprivation (2004) from the local Lower Super Output Area (LSOA), Panel B uses the number of working age people receiving means-tested benefits within the LSOA in November 2010, Panel C captures the number of house sales in the Middle Super Output Area in 2010, Panel D and Panel E uses the number of emergency and elective admission, respectively, to hospital in the previous year, and Panel F uses the total number of days spent in hospital in the previous year.

Figure A.6: Standardised mean predicted and observed survival, by doctor quality ventile



Notes: (1) Predicted survival regresses an indicator of 30-day survival on age and sex dummies (and an interaction between the two sets of dummies), ethnicity dummies, dummy variables for each component of the charlson index, hospital costs in the previous financial year, an indicator of whether the patient has previously been treated for a stroke, and dummy variables for the primary and one secondary diagnosis; (2) Estimated quality shows the 'alternative' measure of quality described in Section 5.2, without patient characteristics included as controls;(3) Both outcomes standardised to have mean zero and standard deviation one; (4) Ventiles on the basis of the alternative quality measure for each doctor.

Figure A.7: Distribution of estimated doctor fixed effects, by patient type



Notes: (1) Estimates from Equation (8) are adjusted using the one-step method used for the baseline results.

Table A.1: Descriptive statistics of cardiologists, by linking status

	Linking		Non-linking	
	Mean	S.D.	Mean	S.D.
Volume	350.53	298.55	310.49	316.19
Volume (STEMI only)	140.42	153.14	128.21	169.00
Annual Volume	52.59	31.76	47.62	35.66
30-day death rate	0.064	0.026	0.065	0.036
1-year death rate	0.144	0.042	0.142	0.054
Number of hospitals	2.39	1.41	1.00	0.00
Number of years in the data	7.96	3.75	6.92	4.21
Number of consultants	437		1,150	

Notes: (1) Includes all consultants in the connected set who treated a minimum of 10 AMI patients between April 2005 and March 2018; (2) 'Linking' consultants are those observed treating at least ten patients in more than one hospital over the specified period.

Table A.2: The estimated distribution of doctor quality excluding previous patient-cardiologist interactions

	Doctor fixed effects	
	30 days (1)	365 days (2)
Std Deviation	0.029***	0.039***
	(0.002)	(0.004)
Variance	0.001***	0.002***
	(0.000)	(0.000)
10th percentile	-0.036***	-0.041***
	(0.003)	(0.004)
25th percentile	-0.019***	-0.019***
	(0.002)	(0.002)
50th percentile	0.000	0.002
	(0.001)	(0.002)
75th percentile	0.018***	0.027***
	(0.002)	(0.002)
90th percentile	0.037***	0.050***
	(0.003)	(0.006)
Number of patients	422,991	422,991
Number of doctors	1,575	1,575
Number of hospitals	135	135

Notes: (1) Sample excludes all patients who are recorded as having received treatment for an AMI in any year, and any patients who has previously received inpatient treatment from the treating cardiologist; (2) Controls include age and sex dummies (and an interaction between all of these dummies), ethnicity, dummies of each component of the charlson comorbidity index, the cost of any hospital treatment in the previous year, indicators of whether the patient has experienced a stroke in the past year, fixed effects capturing the first two recorded diagnoses at the time of admission, indicators of the day of the week, month and year when the patient is admitted, and an interaction between month and year of admission; (3) Bootstrapped standard errors clustered at the doctor level (199 repetitions).

Table A.3: The estimated distribution of doctor quality under alternative hospital controls

	Doctor fixed effects (365 day)	
	(1)	(2)
Std Deviation	0.043***	0.046***
	(0.011)	(0.011)
Variance	0.002	0.002
	(0.002)	(0.002)
10th percentile	-0.046***	-0.047***
	(0.006)	(0.006)
25th percentile	-0.023***	-0.023***
	(0.003)	(0.003)
50th percentile	0.001	0.001
	(0.002)	(0.002)
75th percentile	0.022***	0.022***
	(0.003)	(0.003)
90th percentile	0.041***	0.043***
	(0.005)	(0.006)
Hospital FE	Yes	No
Hospital-year FE	No	Yes
Number of patients	273,558	273,558
Number of doctors	1,430	1,430
Number of hospitals	134	134

Notes: (1) Column one repeats the baseline analysis for the smaller connected set (connected by within-year 'movers') and column two includes time-varying hospital trends instead of hospital fixed effects; (2) All others controls are the same as those used in the baseline analysis, see Table 3 for information; (3) Bootstrapped standard errors are clustered at the doctor level (199 repetitions).

Table A.4: Estimated average doctor quality by region, 2005 - 2017

	Standardised doctor effect	
	30 days	365 days
	(1)	(2)
North East	-0.489	-0.999**
	(0.437)	(0.414)
North West	0.076	-0.139
	(0.263)	(0.277)
Yorks and Humber	0.651**	-0.215
	(0.306)	(0.283)
East Midlands	-0.234	-0.410
	(0.280)	(0.299)
West Midlands	0.350	0.004
	(0.273)	(0.280)
East of England	-0.215	-0.585**
	(0.227)	(0.238)
South East	0.675***	0.337
	(0.259)	(0.272)
South West	-0.023	-0.340
	(0.284)	(0.319)
Observations	510,249	510,249

Notes: (1) Regional dummies capture the area of residence of the patient (regardless of which hospital they are treated by); (2) London is the omitted category; (3) Bootstrapped standard errors are clustered at the doctor level (199 repetitions).

Table A.5: The relationship between estimated doctor quality and area characteristics

	Average doctor fixed effects			
	30-day	365-day	30-day	365-day
	(1)	(2)	(3)	(4)
Deprivation quartile 2	0.065	0.036		
	(0.050)	(0.046)		
Deprivation quartile 3	0.128**	0.073		
	(0.057)	(0.057)		
Deprivation quartile 4	0.187**	0.117		
	(0.073)	(0.073)		
Rural			-0.058	-0.127
			(0.073)	(0.081)
Observations	32,240	32,240	32,240	32,240

Notes: (1) The outcome in all cases is the mean estimated doctor fixed effects on patient survival over 365 days for all patients living in the LSOA between 2005 and 2018; (2) Area characteristics assigned based on the 2001 LSOA of patient residence; (3) Deprivation based on Office of National Statistics Index of Multiple Deprivation measured in 2010; (4) Bootstrapped standard errors clustered at the doctor level (199 repetitions).

Table A.6: Correlation of doctor rankings when treating STEMI and non-STEMI patients

	Non-STEMI quartile			
STEMI quartile	1	2	3	4
1	25.5	23.0	30.2	19.8
	(4.0)	(2.8)	(3.0)	(3.7)
2	25.3	33.6	17.7	22.0
	(3.0)	(2.7)	(2.6)	(2.8)
3	24.9	23.8	27.1	22.7
	(2.9)	(2.5)	(2.4)	(2.8)
4	20.2	19.1	25.6	33.2
	(3.7)	(2.9)	(2.8)	(3.5)

Notes: (1) Doctors are assigned to quartiles on the basis of their type-specific estimated fixed effects (adjusted estimates using the one-step method); (2) Bootstrapped standard errors are clustered at the doctor level (199 repetitions).