The Power of Data: Assessing Primary Care Performance Using Routinely Collected Emergency Department Data

Nicos Savva
Management Science & Operations, London Business School, nsavva@london.edu

Sandra Sülz
Erasmus School of Health Policy & Management, sulz@eshpm.eur.nl

Mark Kinirons, Richard Leach
Guy's and St Thomas' NHS Foundation Trust and Kings College London,
mark.kinirons@gstt.nhs.uk, richard.leach@gstt.nhs.uk

Demand for Emergency Department (ED) care has been rising. At least partly, this increase is driven by the failure to provide timely and high-quality primary care in the community, resulting in patients being forced to use EDs. From a system efficiency perspective, it is important for health authorities to identify Primary Care Practices (PCPs) that place a lower burden than average on EDs so that best practices can be identified and disseminated and PCPs that place a higher than average burden so that they can be supported. This paper develops and validates one such methodology using operational data routinely collected by an ED in England. To do this we focus on the subset of patients *attending* the ED with Ambulatory Care Sensitive (ACS) conditions. These are conditions that are possible and cost effective to treat in a non-ED setting. We use a variance decomposition approach, where we control for observed and unobserved heterogeneity, to show that approximately 37% of the variance in the proportion of ACS attendances is due to systematic differences between PCPs and such differences are large enough to be operationally relevant. Furthermore, PCPs that perform poorly using the variance decomposition approach are more likely to score poorly on patient surveys. We demonstrate that this measure is superior to measures using ACS admissions that have been used in the past literature and conclude with an exploratory study that demonstrates how this measure can be used to quantify the link between operational drivers (e.g., staffing) and PCP performance.

Key words: Primary Care; Emergency Department; Variance decomposition
History: November 11, 2021

1. Introduction

Emergency departments (EDs) across the developed world are reporting a significant increase in demand for their services (see Figure 1 for evidence from England and Berchet (2015) for evidence from around the world). Some of this increase can be attributed to demographic factors, such as the continuing rise in life expectancy which leads to increasing numbers of older people with chronic conditions. However, by some estimates, as many as 40% of ED attendances are deemed

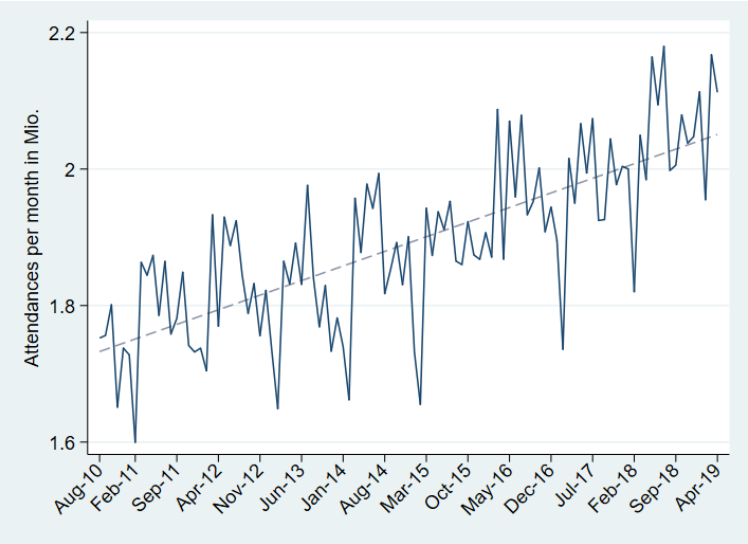


Figure 1 Emergency Department attendances in England, Aug 2010- Apr 2019. Source: Unify2 / SDCS data collections - WSitAE and MSitA

to have been inappropriate and at least partially due to failures in primary care (Ismail et al. 2013). Although many primary care practices (PCPs) provide an excellent range of services in a timely fashion, some are perceived as having unacceptably long waiting times for appointments, few options (if any) after regular office hours, or limited diagnostic resources that force patients to seek care at EDs (Berchet 2015, MacKichan et al. 2017, Bavafa et al. 2021). Treating such patients at the ED is considerably more expensive than providing care in non-emergency settings (Weinick et al. 2010,Galarraga et al. 2015), creates delays and worse outcomes for high-acuity patients (McCarthy et al. 2009, Soltani et al. 2020), and contributes to ED physician/nurse burnout (Watson et al. 2019). From a system efficiency perspective it is therefore important to identify PCPs whose patients place a lower than average burden on ED departments so that best practices can be identified and disseminated, and identify PCPs whose patients place a higher than average burden in order to provide support.

Despite its systemic importance, assessing PCP performance in many healthcare systems is difficult due to lack of data and service fragmentation – PCP data is maintained locally using incompatible information systems, with complex and non-transparent data ownership structures, making direct comparisons between PCPs difficult (Gentil et al. 2017, Clarke et al. 2019). In addition, PCP data does not typically include information on quoted waiting time for appointments and/or the number of patients that tried but failed to get an appointment. As a result, PCP performance is routinely assessed by means of patient surveys (Cowling et al. 2015). Surveys are by construction subjective, they can be slow and expensive to administer, and although they may provide valuable insights on patient perceptions of PCP's performance, they do cannot measure the

impact of failures in primary care on ED demand. This paper develops and validates one alternative methodology based on variance decomposition methods using operational data routinely collected in the course of providing emergency care in one major National Healthcare Service (NHS) hospital based in England. The methodology allows a health authority, such as NHS England, to identify PCPs that perform better or worse than average in terms of their patients use of ED care and, in addition, makes it possible to (at least) partially estimate the ED cost attributed to variation in PCP performance.

The setting. Healthcare in England is provided by the NHS and is funded through taxes. Residents can access primary, emergency, and specialist care free of charge. We will first describe the primary and emergency care setting and then the patient decision-making process.

• Primary care: is provided by General Practices (or PCPs as we will refer to them for the rest of the paper), which are financially independent entities contracted by the NHS to offer a comprehensive range of primary care services (e.g., consultations, prevention, screening, immunization, some diagnostic services, and minor treatments) and serve as gatekeepers for diagnostic and other specialist services typically offered in hospital settings. They are reimbursed on the basis of a risk-adjusted capitation system augmented by pay-for-performance elements, e.g., the patient experience survey, (Roland and Guthrie 2016). To access care, UK residents need to register with a PCP. Until 2015 patients had to register with a PCP in whose catchment area they resided. This has been relaxed since then with patients being allowed to choose PCPs more freely (NHS 2016). Patients typically access PCPs by appointment during their usual operating hours. These vary but for most practices they are between 08:00 and 18:30, Monday to Friday. Out-of-hours access to primary care varies and may include telephone service, out-of-hours clinics, or home visits (National Audit Office 2014).

• Emergency Care: is provided in hospital EDs such as the one we study. EDs are typically open 24/7 and offer a wide range of services and are accessible without an appointment. Since 2005 the government has set a target for EDs to assess, diagnose, treat, and admit or discharge patients visiting EDs within four hours of arrival (Weber et al. 2012), a target that was met for 95% of patients until 2015 and for over 85% of patients after that.

• Patient's decision making process: Within the NHS patients are free to seek care at the PCP or the hospital ED. The decision to visit the PCP or ED is complex and influenced by a number of patient-specific and PCP-related factors (see review by Huntley et al. (2014)). For instance, the interviews conducted by MacKichan et al. (2017) suggest that factors that contributed to patients choosing to attend EDs as opposed to PCP were the intricate and difficult to navigate PCP appointment systems, PCP appointment availability, the lack of access to PCP care out of hours, and patient perceptions about the better level of care at the ED compared to the PCP.

Our approach focuses on routinely collected ED data over a period of five years. The aim is to identify PCPs whose patients make disproportionably more use of the ED resources, especially when this use is clinically unwarranted. To do so we focus on the small number of PCP practices that contribute the most to the number of attendances for the ED (the top 84 PCPs that generate 75% of demand). Not surprisingly, these PCPs are located near the hospital (average distance of 3.8km). We assess PCP performance based on the number of patients *attending* the ED with ambulatory care sensitive (ACS) conditions – these are conditions where effective community care and case management is known to help prevent the need for hospital admission (Busby et al. 2015, Blunt 2013) and, in addition, treating ACS conditions in an ED setting can cost twice as much as in a non-ED setting (Galarraga et al. 2015). The premise is that all things being equal, patients with ACS conditions registered with PCPs that offer better care (e.g., timely access, effective case management) will exhibit fewer ED attendances. We further develop a methodology that directly controls for observed heterogeneity based on exogenous factors (e.g., differences in the scale, patient case mix, patient socioeconomic status) and unobserved heterogeneity (e.g., patient preferences) and use a random-effect model (Wooldridge 2010) to decompose the residual variance in ACS attendances to a component that is PCP specific (between PCP variation) and a component that is random (within PCP variation).

We find that about 37% of the variation in ACS attendances can be attributed to systematic differences between PCPs and that these differences are not only statistically significant but their magnitude is large enough to be operationally relevant. For example, if all PCPs that performed worse than the 75th percentile were to improve their performance to the 75th percentile, then approximately 10.5% of ACS attendances could be avoided. This would translate to a UK-wide reduction in ED attendances of 480,000 patients annually saving the NHS approximately £76 million per annum. Although it is important to emphasize that not all PCPs that score poorly based on this measure are necessarily providing a poor level of service (e.g., there may be residual unobserved heterogeneity that the model specified in this paper does not control for), it is nevertheless useful to use this decomposition to create a cardinal measure of the pressure that individual PCPs place on ED resources – PCPs that generate a systematically higher (lower) number of ACS attendances than average can be thought of as placing a higher (lower) burden on ED resources than average.

We then demonstrate that the measure constructed using the variance decomposition method proposed by this work has validity. We do so by using the results of the annual patient survey conducted by the NHS. More specifically, we show that the between PCP variation is correlated with patients' answers – PCPs that rank poorly in the proposed measure are those where a larger proportion of patients report they were unable to get an appointment in a timely fashion and are less

likely to recommend their PCP to others. Importantly, the within PCP variation (i.e., the random error) is not correlated with patient survey responses. This is consistent with the interpretation that the within PCP variation is indeed due to random fluctuations. Furthermore, we repeat the analysis but instead of focusing on ACS conditions we focus on conditions that are less likely to be influenced by PCP performance such as road accidents, myocardial infractions (AMI), and strokes. As expected, for these conditions we find much smaller systematic variability between PCPs, this sytematic variability is negatively correlated with the one constructed using ACS conditions, and is not correlated with patient survey responses. We also demonstrate that our approach, which is based on ACS attendances at the ED, is superior in quantifying PCP performance compared to approaches that rely on the more readily available data on ACS admissions that have been proposed by past literature (see literature review). A number of additional robustness tests confirm that the measures constructed are not particularly sensitive to modeling choices.

We then proceed to demonstrate the practical usefulness of this measure by conducting an exploratory study on one of the determinants of PCP performance – staffing levels. More specifically, we show that PCPs that have relatively large ratios of patients per full-time physician employee are more likely to score poorly and more so if these PCPs are relatively small. If a relatively small PCP (i.e., a PCP with 4,441 registered patients which is equivalent to 1 standard deviation below the mean) was to reduce the patient-to-staff ratio by 1,000 then this would reduce the number of ED attendances by patients with ACS conditions by approximately 5.5%. Similarly, if two such small PCPs were to merge to form a PCP of average size, this would reduce the number of ACS attendances by approximately 0.4% without requiring any additional staff.

Collectively, this work demonstrates how one can utilize routinely collected operational data downstream (at the ED level) to infer performance at multiple upstream locations (the PCPs). This methodology should aid health authorities to overcome data limitations that exist at the more upstream level and assess performance in a cost effective and objective manner (i.e., that does not rely on patient surveys). We also demonstrate how this methodology can be used to establish that one PCP operational measure, the number of patients per full-time physician employee, affects ED attendances. Further work could utilize this methodology along with data from all EDs operating nationally to identify whether there is evidence of regional variation in the quality of primary care across the UK that needs to be addressed. Furthermore, if one has access to more detailed PCP data, this methodology can be used to identify clinical pathways and operational practices that allow some PCPs to perform better than others and focus on disseminating such practices. Furthermore, to the extent that this methodology enables health authorities to better estimate the cost savings at the ED level associated with interventions at the PCP level, it may aid the development of business cases to fund such interventions. Beyond the NHS, where this study takes

place, this methodology would be useful to any public payer such as the Centers of Medicare and Medicaid in the US or private insurances providers that are responsible for reimbursing primary and hospital-based care. More radically, an appropriately applied version of this methodology could be used as part of the reimbursement formula for PCPs, with PCPs that perform better (or worse) than average in the proposed measure subjected to a financial reward (penalty). Such financial incentives that rely on relative benchmarking could provide the impetus for PCPs to internalize the cost burden associated with providing care at the ED to patients who would be better looked after in primary care and, importantly, it can be implemented with already existing data (see Shleifer (1985), Savva et al. (2019)).

2. Literature review

2.1. Empirical work in operations management

This paper contributes to the relatively recent empirical literature in operations management that uses observational data to rigorously measure and assess performance in the context of service systems. For example, Tan and Netessine (2014) use data collected by a restaurant software system to establish that the workload of restaurant servers affects meal duration and sales; Staats and Gino (2012) and Xu et al. (2020) use bank transaction data to examine the impact of repetition on worker productivity and how workload affects error rates, respectively; Ibanez and Toffel (2020) use data on the schedule of restaurant inspections to establish that the order of inspections affects outcomes; Wang and Zhou (2018) use supermarket checkout data to establish that clerks working in dedicated queues work faster compared to servers working in pooled queues, a result also established in the context of ED service times by Song et al. (2015). More related to our work is the rich empirical literature that exploits data generated in a healthcare setting (as in the Song et al. (2015) example). More specifically, a number of studies have established that ward- or provider-level workload affects patient service rates (KC and Terwiesch 2009, Berry Jaeker and Tucker 2017), admission decisions (KC and Terwiesch 2017), intensity of services provided (Freeman et al. 2017), hospital reimbursement (Powell et al. 2012), nurse absenteeism (Green et al. 2013), and patient mortality rate (Kuntz et al. 2015), amongst others. Beyond the impact of workload, work has exploited radiology data to examine the effect of discretionary task ordering on worker productivity (Ibanez et al. 2018), admissions data to establish that delays in ICU admission affect patient length of stay (Kim et al. 2015) and that assigning patients to beds designated for different services increase patient length of stay (Song et al. 2020), ED service-time data to establish that performance feedback affects worker productivity (Song et al. 2018) and that queue length and waiting times affect patient decisions to leave without being seen (Batt and Terwiesch 2015). As in this paper, the data exploited by these studies was routinely collected in the process of providing a service. In

contrast to this paper, the work cited above uses the available data to study aspects of performance at the level of the unit in which the data was collected (e.g., ED data to assess aspects of ED performance).

More recently, studies have taken a more system-wide perspective to examine whether data collected at one level can tell us anything about performance further downstream. Examples include Freeman et al. (2020), who examine how ED workload affects hospital admission decisions, Soltani et al. (2020) who examine how ED workload affects the efficiency with which hospital inpatient treatment is provided, Song et al. (2021) who analyze how the duration of home healthcare visits affects hospital readmissions, and Bavafa et al. (2021) who study lower than average PCP availability affects ED visits and readmissions. Similar to these studies, our work also takes a system perspective but instead of using data collected at one unit to assess the performance of units further *downstream*, our work focuses on units that are further *upstream*. These are units that act as a precursor to the unit in which the data is collected. Furthermore, our goal is more descriptive, we aim to provide a validated measure of performance for the upstream units (the PCPs) based on the richer data available downstream (the EDs) and demonstrate how this measure can be used to identify features of upstream units that are associated with better performance. Furthermore, such work can be used to assess the potential cost implications of upstream (PCP-level) interventions to the unit downstream (the ED).

2.2. Healthcare Literature

Our paper also contributes to the literature on health economics and health service research that examines primary care performance. This field has received considerable interest as indicated by the number of systematic reviews that seek to synthesize how primary care performance is affected by financial incentives (Scott et al. 2011, Gibson et al. 2013), patient demographics (Huntley et al. 2014), organizational and workforce aspects (van Loenen et al. 2014, Gibson et al. 2013, Huntley et al. 2014), and accessibility (Rosano et al. 2013). The range of systematic reviews underlines the variety of research foci. Studies do not only differ with respect to their objective but also on the type of data and methodologies that are deployed to assess primary care performance. Examples range from qualitative interview studies (Maisey et al. 2008, McDonald and Roland 2009), patient surveys (Cowling et al. 2016, Schoen et al. 2004, Campbell et al. 2009), routinely collected quality indicators at the PCP level (Brown and Lilford 2006, Downing et al. 2007, Hong et al. 2010, Dusheiko et al. 2011), and to hospital admission data (Harrison et al. 2014, Barker et al. 2017, Dusheiko et al. 2006, Vuik et al. 2017, Lavoie et al. 2019, Busby et al. 2017, Dusheiko et al. 2011). It can be debated as to how far this data is equally useful or informative for performance assessment. Interviews and surveys generate insights into perceptions about performance but cannot quantify performance objectively.

Routinely collected quality indicators at the PCP level, which are primarily used for reimbursement purposes, can distort the assessment due to upcoding and gaming behaviour (Jürges and Köberlein 2015, Bastani et al. 2019). Likewise, hospital admission data may not be representative as it can be affected by "great variation at the hospital level and physician level in the ED regarding the decision to admit or discharge..." Galarraga et al. (2015, p.176). In contrast to these papers, our study assesses primary care performance based on ED attendances. Since ED attendances reflect the decisions and preferences of patients, they are an objective measure that has the advantage of being exogenous to the hospital. Use if ED attendances is limited in extant literature, perhaps due to lack of data availability. One notable exception is Dolton and Pathania (2016), which used ED attendances to show that a piloted policy change, which extended PCP services during the weekend, reduced demand for emergency care. Our work complements this analysis by showing that ED attendances can be used to construct measures that allow for meaningful comparisons between PCPs.

Our work is also connected to the growing literature on provider profiling. This literature focuses on developing methodology aiming to rigorously assess and visualize performance differences between providers (e.g. Jones and Spiegelhalter 2011, Spiegelhalter 2005, Racz and Sedransk 2010, Paddock 2014) with applications in the context of hospitals (Paddock et al. 2015), medical specialists (Adams et al. 2010), and PCPs (Thomas et al. 2004). Such applications typically rely on performance data collected from the same unit or financial (claims) data centrally collected at the purchaser level. In contrast, our work seeks to conduct upstream provider profiling using downstream data.

3. Empirical analysis

The goal of this work is to use routinely-collected operational hospital data, and more specifically patient attendances at hospital EDs, to identify PCP practices whose patients are placing a disproportionate pressure on ED resources, especially when these attendances are not medically warranted. For this reason, we focus on the subset of patients that present to the ED with ACS conditions. These are conditions where effective community care and case management can help to prevent (but not completely eliminate) the need for ED attendance or hospital admission (Busby et al. 2015, Blunt 2013). More specifically, ACS conditions can be categorized into acute, chronic, and vaccine-preventable conditions (see Table 1 for a complete list of these conditions). Examples of acute conditions include ear, nose, and throat infections or urinary tract infections. These conditions can be (and most often are) treated effectively in an ambulatory-care setting provided the patient has timely access to care (e.g., a PCP appointment). Examples of chronic conditions include asthma, chronic obstructive pulmonary disease, and congestive heart failure. These conditions require monitoring in the community to ensure that the condition remains stable and the

patient adheres to guidelines and medications. Examples of vaccine-preventable conditions include influenza and pneumonia. Their prevalence can be reduced or completely eliminated if an adequate vaccination scheme is in place. All things being equal, we expect patients with ACS conditions registered with PCP practices that offer timely access to appointments, effective management of patients' chronic conditions, and administer vaccinations as per national guidelines to exhibit fewer ED attendances compared to patients registered with less effective PCP practices.

Table 1 List of ACS conditions

Acute	Chronic	Vaccine-preventable
Cellulitis	Angina	Influenza
Dehydration	Asthma	Pneumonia
Dental conditions	Chronic obstructive pulmonary disease	Tuberculosis
Ear, nose, and throat infections	Congestive heart failure	Other vaccine-preventable conditions
Gangrene	Convulsions and epilepsy	
Gastroenteritis	Diabetic complications	
Nutritional deficiencies	Hypertension	
Pelvic inflammatory disease	Iron deficiency	
Perforated or bleeding ulcer		
Urinary tract infections or pyelonephritis		

3.1. Data

The empirical analysis focuses on a dataset of approximately 700,000 patient ED attendances taking place at a large teaching hospital in a major metropolitan area in England over 5 years (2013–2017). In 2017, the ED had approximately 400 patient attendances per day, 23.3% of which were admitted to the hospital. The data includes the date/time of arrival, patient demographics (age, gender), an identifier for the PCP the patient is registered with, information on presenting complaints and diagnosis, and patient disposal (e.g., whether the patient was admitted to the hospital or discharged home). This dataset is augmented with information from nine publicly available sources that contain information on PCPs (e.g., location, payments), NHS hospitals (location), and information on UK geolocations and deprivation scores at each location, outlined in Table 2.

3.2. Data pre-processing

As the hospital we study is located in a major metropolitan area, it attracts a substantial number of commuters and tourists that are registered with PCP practices outside of the hospital's primary catchment area or, in the cases of overseas tourists, patients who are not registered with any PCP practice at all. More specifically, 74.6% of the ED attendances are from patients registered with 7,692 different PCPs scattered across the UK. (To put this number in perspective, the total number of PCPs in the whole of England and Wales in the duration of the study is approximately 15,000.) The rest of the patients were either overseas visitors who were not registered with the NHS (7.7%),

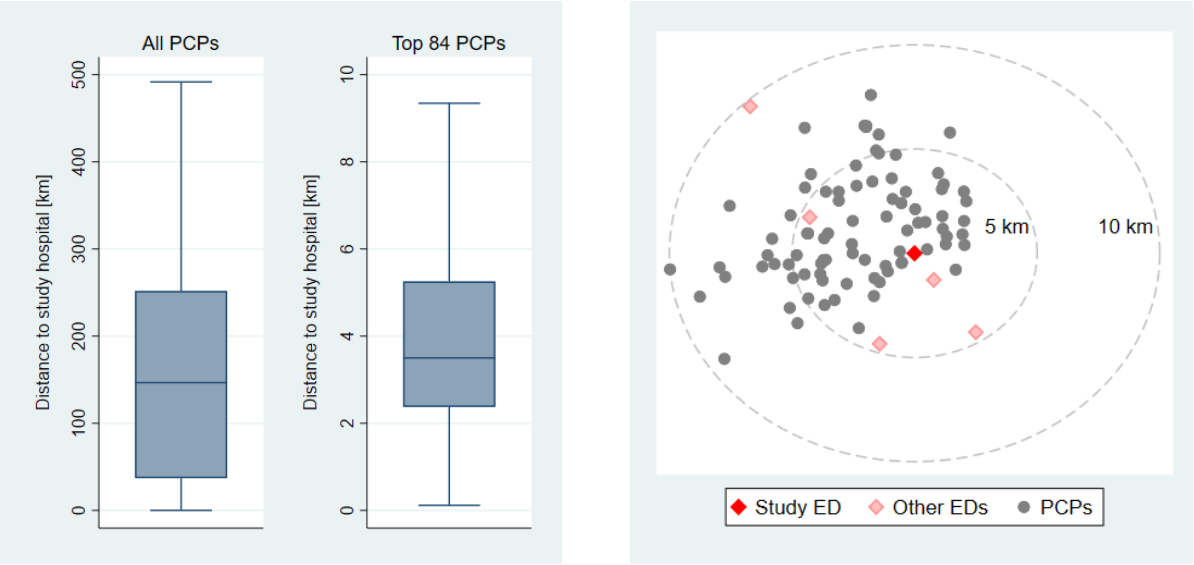
Table 2 Publicly available data bases used in this study

Publisher, Data set	Extracted Information	Link
NHS Digital, Epraccur	Name and addresses of PCPs and when they ended operating (time of closure)	https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data
NHS Digital, Ebranches	PCP branches and when they started operating	https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data
NHS Digital, Patients registered at a GP Practice	Number and demographics of patients registered at PCP	https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice
NHS Digital, Payments to general practice	PCP patient case-mix	https://digital.nhs.uk/data-and-information/publications/statistical/nhs-payments-to-general-practice
NHS Digital, Etrust	Location of hospital sites	https://digital.nhs.uk/services/organisation-data-service/data-downloads/other-nhs-organisations
NHS Digital, General and Personal Medical Services	FTE physicians working at PCPs	https://webarchive.nationalarchives.gov.uk/20180328140045/http://digital.nhs.uk/catalogue/PUB20503
NHS England, GP Patient Survey	Patient perception of PCPs	https://gp-patient.co.uk/about
Free Map Tools, UK Postcode database	Latitude and longitudinate coordinates of UK postcodes	https://www.freemaptools.com/contact.htm
Ministry of Housing, Communities & Local Government	English indices of deprivation at postcode level	https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015

Table 3 Patient characteristics depending on PCP status

	PCPs assigned	PCPs not assigned	PCPs not applicable
% of ED attendances	74.6%	17.7%	7.7%
Age: Mean (SD)	36.9 (23.1)	34.5 (20.3)	34.7 (17.3)
Male	50.5%	52.3%	59.9%
ACS condition	17.7%	17.0%	16.6%
Ambulance arrival	25.4%	24.7%	26.4%

or the PCP could not be determined upon ED arrival (17.7%). Since the goal of this study is to compare PCP performance, we can safely exclude data from overseas visitors. We also exclude patients whose PCP practice could not be determined. This is a relatively small subset which, as indicated by Table 3, appears to be similar on the basis of observable characteristics to those patients that are included in the sample.



(a) Distance to study hospital (b) Geographic location of top 84 PCPs.

Figure 2 Interquartile range of distance and geographic location of top 84 PCPs.

To preserve anonymity of the ED and PCPs, the geographic coordinates have been transformed.

This leaves 523,069 ED attendances from 7,692 PCP practices of which 92,402 are attributable to ACS conditions. The average and median distances of these PCPs from the hospital are 149.3km and 146.6km (the interquartile range appears in Figure 2a), which indicates that at least some of these PCPs are very far away from the hospital. As indicated by Figure 3, the majority of the ED attendances come from a small number of PCPs. For the purposes of this study we focus on the 84 PCPs that are collectively responsible for 75% of the ACS attendances – the average and median distances for these PCPs are 3.8km and 3.5km (see also Figure 2a). Indeed, Figure 2b shows the locations of these PCPs in relation to the hospital ED. The figure also shows the locations of other hospital EDs. As the figure indicates, all 84 PCPs are within 10km of the hospital. In the Appendix, as a robustness check, we change the sample to include the top 60 and the top 136 PCPs that collectively account for 70% and 80% of the patient ACS attendances, respectively (see Appendix, §3.1).

PCPs are operating in an environment characterized by occasional mergers, expansions and closures. Since these expansion and closing activities disrupt PCP operations and can distort performance, we exclude these from the analysis. We can identify three PCP closures reported in the Epraccur dataset. For each closure we exclude the final operating year to ensure that we have full-year observations (closures occur throughout year). As a robustness check, we also exclude all three practices that closed during the study period from the entire analysis (see Appendix §3.2). To the best of our knowledge there is no publicly available dataset that tracks PCP mergers. But there is information about PCP expansion. If PCPs start operating a new branch, this information

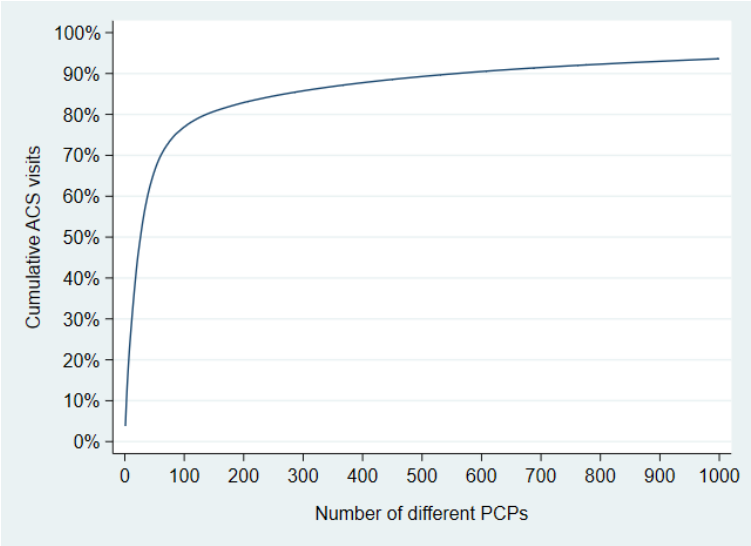


Figure 3 Proportion of ACS attendances accounted for by different PCPs.

is reported in the dataset ebranches. Such a new branch can be the result of a merger because typically, if PCPs merge, one site will become the main site and the other sites will become branches. There are two PCPs with new branches starting to operate during our study period. We exclude these PCPs from our analysis from the first year in which the new branch started operating. As a robustness check, we also exclude these two practices from the entire analysis (see Appendix §3.2).

3.3. Measuring PCP performance: A variance decomposition approach

For every PCP g we denote the count ACS attendances in period t by:

$$P_{gt} = \sum_{i=1}^{N} X_{i,g,ACS,t},$$

where $X_{i,g,c,t}$ is equal to 1 if patient $i \in \{1, ..., N\}$ is registered with practice $g \in \{1, ..., G\}$ in period t and presents to the ED with a condition c that belongs to the set of ACS conditions and 0 otherwise. For the primary analysis we will choose the duration of the period t to be a year. This ensures a sufficiently large number of observations per PCP (the average annual number of ACS ED attendances for the 84 PCPs in the sample is 165 with a standard deviation of 150), and ensures that the measure constructed is less subject to seasonal variation.

Naturally, the number of ACS attendances P_{gt} will vary and this variation may be attributed to (i) stochastic fluctuation within PCPs over time, (ii) variation between PCPs that can be attributed to observed heterogeneity (for example differences in the size of the practice, patient case mix, distance to the focal hospital and other hospitals in the area) or unobservable heterogeneity that has nothing to do with PCP quality of care (for example, differences in patient preferences over EDs of different hospitals), and (iii) variation between practices that is due to persistent differences in

PCP quality of care. Of primary interest is the latter and in what follows we develop an econometric specification that aims to isolate this source of variability as far as possible.

To (at least partially) control for unobserved heterogeneity that is orthogonal to PCP performance we take advantage of the fact that in addition to patient attendances with ACS conditions we also have information on patient attendances with non-ACS conditions. These conditions, which include chest pain, suspected stroke, and trauma, require ED attendance and less dependent on PCP quality of care compared compared to ACS conditions. If there exists heterogeneity (not related to quality of care) that makes patients from PCP i more likely to visit the ED of the study hospital compared to patients from PCP j (e.g., because patients in PCP i prefer the study hospital while patients in PCP j prefer a different hospital), then, to the extent that these differences are common for patients with ACS and non-ACS conditions, both ACS and non-ACS attendances will be higher for PCP i compared to PCP j. Therefore, we propose to measure PCP performance by normalizing the number of ACS attendances by dividing it with all patient attendances originating from PCP g in period t:

$$A_{gt} = \frac{P_{gt}}{\sum_{c' \in C} \sum_{i=1}^{N} X_{i,g,c',t}}, \tag{1}$$

where $X_{i,g,c,t}$ is defined as above and c' includes patients with ACS and non-ACS conditions. The proportion of ACS attendances A_{gt} constitutes the unit of analysis for the rest of this work and defines a (potentially unbalanced) panel of observations. Since the measure A_{gt} exhibits substantial dispersion, we will log-transform this variable.

To identify the variation in the proportion of ACS attendances A_{gt} that can be attributed to systematic differences between PCPs, we estimate the following model:

$$ln(A_{gt}) = \alpha_0 + \alpha_C C_{gt} + u_g + \epsilon_{gt}. \tag{2}$$

This specification controls for any variation in the proportion of ACS attendances that is due to observed heterogeneity by controlling for a vector of (possibly time-varying) PCP-characteristics C_{gt} on which we expand in §3.4. The residual error is decomposed into two parts ($u_g + \epsilon_{gt}$): i) u_g denotes the part of the variance that is explained by systematic differences between PCPs (also referred to as between-PCP variation); and ii) $\epsilon_{gt} \sim N(0, \sigma^2)$ denotes the idiosyncratic error term (also referred to as within-PCP variation). We make the usual strict exogeneity assumption, which can be expressed in terms of conditional expectations as $E(\epsilon_{gt}|C_{g1}, C_{g2}, ..., C_{gT}, u_g) = 0$ for all $t = 1, ..., T$, see Chapter 10 Wooldridge (2010)).

One possible choice in estimating the model of equation 2 and, in particular, the between-PCP variation u_g is to use a fixed-effect specification. Such specification has the advantage of allowing u_g to be arbitrarily correlated with the control variables C_{gt}. However, since we only

have 5 years of data (i.e., at most 5 data points per PCP) such an estimation would yield noisy estimates of PCP performance. Furthermore, the fixed-effect specification does not allow to control for any systematic differences across PCPs that remain unchanged over time (e.g., whether the study ED is the closest to the PCP or not) as these would be collinear with the PCP fixed effect. Therefore, the estimated PCP fixed effect would be confounded by any such time-invariant differences that are not related to PCP performance. To overcome both of these limitations we will used a random-effect model. More specifically, we make the additional parametric assumption that any heterogeneity between PCPs is drawn from Normal distribution ($u_g \sim N(0, \tau^2)$). This specification only requires one additional parameter to be estimated (τ^2) and allows to control for time-invariant heterogeneity between PCPs. However, these advantages come at a cost of requiring the additional identification assumption that the variables C_{gt} are orthogonal to the PCP random effect (i.e., $E(u_g|C_{g1}, C_{g2}, ..., C_{gT}) = E(u_g) = 0$ see Chapter 10 Wooldridge (2010)). We note that this assumption may be violated if, for example, larger PCPs are more likely to make investments in resources that allow them to treat patients more effectively. In this case the PCP effect u_g would be correlated with the control for scale $Scale_{gt}$. Therefore, after we estimate the model we assess whether there is any evidence that this assumption is violated. In addition, as a robustness test, in Appendix 2.3 we also estimate the fixed-effect model which allows for arbitrary correlations between the control variables and the PCP effect.

Under these assumptions, the total variance of the random part of the model is equal to $\tau^2 + \sigma^2$ and the proportion of the variance that is explained by systematic differences between PCPs, often referred to as intra-class correlation (ICC), is given by $ICC = \tau^2/(\tau^2 + \sigma^2)$. A higher value of ICC suggests that more variability is due to systematic differences at the PCP-level, while a lower value suggests that most of the variability is due to random fluctuations within PCPs.

We calculate robust standard errors clustered at the PCP level. This modelling choice allows for errors to be heteroskedastic across PCP and correlated within PCP.

3.4. Accounting for observable PCP heterogeneity

By defining the dependent variable as the proportion of ACS to total ED attendances (A_{gt}) we control for the portion of PCP heterogeneity that affects ACS and non-ACS patients equally. In this section we discuss how we further control for observed and potentially time-varying heterogeneity, which may be important in cases where heterogeneity might affect ACS patients differently to non-ACS patients.

Scale and case-mix differences. We control for differences in scale and patient demographics between PCPs with the following variables: $Scale_{gt}$ which represents the number of registered patients (in 1,000), $Female_{gt}$ which indicates the proportion of female patients, and $Elderly_{gt}$

which denotes the proportion of patients aged 75 years. For most practices this information is reported biannually (January and July) but for uniformity we use the January values. Furthermore, we note that there is no data available for the first year of our sample (2013) for the female and elderly proportions and we therefore impute the missing information using the 2014 values.

In addition, we use the case mix index variable CMI_{gt} to control for additional differences in the patient case mix. Case-mix indices are frequently used to measure the average severity of cases and to adjust reimbursement rates (Filistrucchi and Prüfer 2019). We can derive similar case-mix indices for the PCPs in our sample by relying on PCP reimbursement datasets. These datasets contain information on the number of registered patients (RP) and the number of weighted patients (WP). The number of weighted patients is obtained by adjusting the registered patient counts by patient demographics, patient needs, and market factors that may affect the cost of care. We can then derive the CMI measure as the ratio of the two: WP/RP. A $CMI_{gt} > 1$ implies that the PCP's patient panel is more severe and expensive to treat than the national average, whereas a $CMI_{gt} < 1$ indicates that the PCP's patient panel is less expensive to treat than the national average.

Alternative hospital choice. One concern is that patients with ACS conditions from different PCPs may find visiting the study hospital more (less) attractive depending on whether they are located close to (far from) the hospital and/or whether this is the closest suitable hospital to them. Normalizing the number of ACS attendances by dividing with the number of non-ACS attendances potentially addresses this concern but we further control for proximity between PCP g and the study hospital based on the distance between the two. More specifically, for each PCP g we determine whether the study hospital is the closest hospital alternative ($closest_g$ equal to 1) or whether an alternative hospital is closer ($closest_g$ equal to 0). All distance measures are calculated as straight line distances between the centroids of the hospital and the PCP postcodes using the Stata command *geonear* (Picard 2012). Note that we measure proximity based on the location of the PCP and not the patient's home because the exact patient address is not available. Nevertheless, most patients register with a PCP close to where they live. We believe that using this binary measure instead of the actual distance between PCPs and the study hospital is a more sensible modeling choice as this takes into account that some patients may be willing to travel a longer distance to get to the ED simply because there is no closer alternative.

Socio-economic factors. Prior literature has shown that ACS attendance and admission rates are affected by patients' socioeconomic status with disproportioanly higher ED attendance rates reported for lower socio-economic status (Oster and Bindman 2003, Johnson et al. 2012). Therefore, we control for socioeconomic differences between PCP locations. We do so via the index of multiple deprivation provided by the Department for Communities and Local Government. This

		2013	2014	2015	2016	2017
Proportion of ACS ED attendances	Mean	0.199	0.199	0.197	0.202	0.183
	SD	(0.029)	(0.031)	(0.032)	(0.029)	(0.029)
Proportion of PCPs closest to ED	Mean	0.095	0.096	0.098	0.088	0.089
	SD	(0.295)	(0.297)	(0.299)	(0.284)	(0.286)
Deprivation rank (in 1,000)	Mean	9.148	9.204	9.109	9.204	9.172
	SD	(5.364)	(5.372)	(5.335)	(5.36)	(5.387)
PCP scale (in 1,000 patients)	Mean	7.983	8.243	8.475	8.721	8.596
	SD	(3.791)	(3.874)	(4.019)	(4.124)	(3.985)
Proportion of female patients	Mean	0.498*	0.498	0.487	0.490	0.489
	SD	(0.045)*	(0.045)	(0.046)	(0.046)	(0.047)
Proportion of elderly patients	Mean	0.035*	0.035	0.035	0.035	0.034
	SD	(0.018)*	(0.018)	(0.018)	(0.017)	(0.017)
Case-mix Index	Mean	0.984	0.972	0.972	0.963	0.955
	SD	(0.072)	(0.073)	(0.071)	(0.069)	(0.075)

Table 4 Descriptive statistics of the top 84 PCPs that account for 75% of patient ACS attendances.
*** values imputed using 2014 data.**

index D_g captures dimensions such as average income, employment rates, education level, health characteristics, crime, barriers to housing and services. It is highly localized (at the postcode level) and ranks locations in England from 1 (most deprived area) to 32,844 (least deprived area). We re-scale the index to capture ranks in units of 1,000. While depriviation is in theory a time-varying variable, the data is only updated approximately every 5 years, which makes it a time-invariant factor in our analysis.

Time fixed effects. Finally, we control for any common temporal variability across PCPs by including year fixed dummy variables $Year_t$. As noted by Wooldridge (2010), omitting such variables can induce serial correlation in the implied error term.

Importantly, in this methodology we do not control for operational or clinical differences between PCPs that are under the direct control of the PCP (e.g., opening hours, patient-to-staff ratio, expertise of clinical staff, diagnostic facilities, etc). We do not include these factors as they are the driving force behind the variability in PCP performance that we would like to capture. Indeed, we would like to have a measure of PCP performance that allows us to identify how such operational and clinical factors affect performance. We demonstrate this further in §5.

Full descriptive statistics by year of the variables used are provided in Table 4. One observation is that there is relatively little variation across years compared to variation within a year, suggesting that most of these variables vary slowly over time.

3.5. Quantifying PCP performance

We estimate the model with the controls described above using the *mixed* command, Stata Version 16. Table 5 provides the model estimates for different specifications relating to equation (2). Column (1) presents the results of a model without any controls except for the year fixed effects. Column (2)

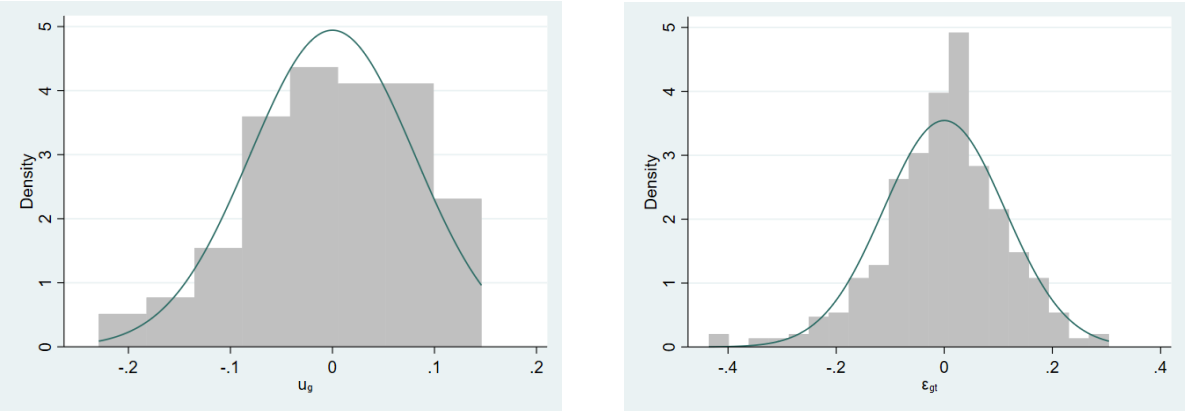
Table 5 Quantifying PCP performance with variance decomposition

	(1)	(2)	(3)
Closest hospital		0.018	0.024
		(0.019)	(0.022)
Depriviation rank		−0.008**	−0.007*
		(0.003)	(0.003)
Scale			−0.004
			(0.004)
Female			0.846
			(0.657)
Elderly			−0.151
			(0.789)
CMI			0.011
			(0.231)
Constant	−1.624***	−1.553***	−1.958***
	(0.017)	(0.027)	(0.354)
Year FE	Yes	Yes	Yes
$\widehat{\tau}^2$	0.012	0.010	0.009
$\widehat{\tau}^2$ 95% CI	[0.006; 0.025]	[0.005; 0.022]	[0.006; 0.013]
Intraclass correlation (ICC)	45.56%	41.46%	36.96%
Model Wald χ^2	37.03	40.91	45.60
Observations	408	408	401
Number of groups	84	84	83

Clustered standard errors in parentheses.
*** p<0.001, ** p<0.01, * p<0.05.
Note: Number of observation differs due to missing case-mix information in 7 PCP-years.

adds controls for PCP location and deprivation. Column (3) adds controls for PCP scale, proportion of female and elderly patients, and CMI. Focusing on the controls of Column (3), all except one are not statistically significant at conventional levels (i.e., p-values$>5\%$), suggesting that normalizing the number of ACS attendances (by dividing with the total number of attendances) effectively controls for (observed) heterogeneity. Only the deprivation rank is statistically significant. The negative coefficient suggests that the ratio of ACS to non-ACS patients attendances is lower for PCPs located in less deprived locations. Adding the controls explains some of the between PCP variance as the ICC decreases from 45.56% in the model without controls (Column 1) to 36.96% in the most detailed model (Column 3). Nevertheless, all models suggest that over one third of the variance in PCP performance is systematic at the PCP level and the variance estimate of u_g, $\widehat{\tau}^2 = 0.009$, is more than four standard errors (0.002) away from zero.

Figure 4 presents graphically the between-PCP variation ($\widehat{u}_g$) and the within variation ($\widehat{e}_{gt}$) based on the model of Column (3). A more thorough analysis of $\widehat{e}_{gt}$ and $\widehat{u}_g$ indicates that the model is correctly specified (see Appendix §1 for details). Using this decomposition, one can think of PCPs with $\widehat{u}_g \leq 0$ as better-than-average (because the proportion of ED attendances by patients with ACS conditions registered at these practices is below average once we control for observable differences across PCPs) and, conversely, PCPs with $\widehat{u}_g > 0$ as worse-than-average practices. Furthermore, the



(a) Variation between PCPs (b) Variation within PCPs

Figure 4 Variation in PCP performance

magnitude of the between PCP variation in ACS attendances is large enough to be of practical importance. Since this is a log-linear model (where the dependent variable is the logarithm of the proportion of ACS attendances), all things being equal the difference between a PCP that is one standard deviation above the mean (i.e., a PCP with $\widehat{u}_g = 0.095$) and a PCP one standard deviation below the mean (i.e., a PCP with $\widehat{u}_g = -0.095$) is approximately 24% ($= \frac{1}{1-\pi} 2 \times 0.095$, where $\pi = 0.196$ is the proportion of ACS attendances in the sample) more ACS attendances. Since the average PCP practice has 165 ACS attendances per annum, this translates in a difference of 39 ACS attendances per annum. An alternative way to assess the magnitude of the between PCP variation is the following. If all practices that preformed worse than the 25$^{\text{th}}$ percentile of the $\widehat{u}_g$ measure (i.e., those that have $\widehat{u}_g > -0.054$) could somehow perform at the 25$^{\text{th}}$ percentile, the annual number of patients with ACS conditions attending the ED would drop by 1,376, which is equivalent to a 10.53% reduction relative to the status quo. Using a bootstrapping method based on 10,000 replications, we estimate the 95% confidence interval of this reduction to be [6.70%; 14.15%] per annum (see Appendix §4 for more details on this calculation). To translate this reduction in attendances to costs, we note that the average cost (from the perspective of a purchaser/ commissioner) of an ED attendance is estimated to be £158 (based on the average tariff for ED attendance at a type 1 and type 2 department NHS (2017)), therefore the cost reduction could be in the order of £217K annually for this ED alone. Of course, this is an underestimate of the saving potential as other EDs in the vicinity of the same PCPs would also see a reduction in ACS attendances. To get a more representative estimate of the potential cost benefit associated with such an improvement we could extrapolate these findings to the whole of England. In 2016/17, there were approximately 23,400,000 annual ED attendances (NHS 2018a). In our sample, 19.6% ED attendances are by patients with ACS conditions. Assuming this number is similar in other EDs, this would amount to 4,586,400 ACS attendances. Similarly, if these attendances would also

decline by the same amount (10.53%) the savings across England would be in the vicinity of £76 Million annually. As these numbers suggests, the systematic variability in performance between PCPs is not only statistically significant but its magnitude is large enough to be operationally meaningful.

We conclude this section by reporting the results of model specification and robustness test. The random-effect model (2) assumes that the unobserved between-PCPs variation u_g is i) Normally distributed and ii) orthogonal to the control variables. To check the first assumption, we examine the residuals of the model reported in Table 5, column 3. We do not find sufficient evidence to conclude that the Normality assumption is violated (p-value of the skewness-kurtosis test is 25%, see Appendix 1). To test the second assumption we perform a test based on Mundlak (1978). This is an alternative to the more widely used Hausman test for random vs fixed effect. In contrast to the Hausman test, this test allows for heteroskedastic errors within PCP and intragroup correlation. The test involves adding the mean of the time-varying variables of (2) (i.e., scale, proportions of female and elderly, CMI, see Appendix §1.1) as explanatory variables and testing if their coefficients are jointly different to zero. Failing to reject this hypothesis constitutes evidence for the random- and against the fixed-effect specification. The p-value of this test is 13.4% which suggests that the random-effect specification is appropriate. Furthermore, in Appendix 2.3 we estimate a fixed-effects model, which yields similar results. In model (2), we control for scale through a linear specification. The results remain similar.

In addition, in appendix 2.1 we present analysis were we allow for scale to have a non-linear effect on the proportion of ACS attendances. We note that the log-transformation of the A_{gt} variable is econometrically equivalent to running a regression of the logarithm of the number of ACS attendances controlling for the number of total attendances in the right hand side of equation (2), where we also restrict the coefficient of the number of total attendances to be 1. We actually estimate this model in Appendix §2.2 without imposing the restriction that the coefficient of the number of total attendances is 1. The results are similar and the coefficient of the total number of attendances is indeed close to one (coefficient=1.076, standard error= 0.012).

4. Validation of the performance measure

By deploying the methodology outlined in the previous section we are able to construct an estimate of PCP performance $\widehat{u}_g$ using operational ED-level data. In this section we seek to validate whether this is a meaningful measure of performance.

4.1. Correlation with patient survey outcomes

To establish whether the measure constructed in the previous section is indeed related to PCP quality of care as opposed to other factors (e.g., unobserved heterogeneity) we examine whether it

is correlated with qualitative perceptions of performance, as recorded in patient surveys conducted by the NHS. These surveys consist of a variety of questions relating to access to PCP services, the patient's experience with their PCP appointments, management of care plans, and a general assessment of the overall experience with their PCP.

To examine if the measure of PCP performance $\widehat{u}_g$ correlates with patients perceptions, we focus on responses to two survey items: (i) whether patients would not recommend their PCP; and (ii) whether patients experienced problems accessing their PCP. We focus on these aspects because they are consistently surveyed over time and they are indicative of the patients' overall dissatisfaction with their PCP performance and of problems with PCP access, respectively. Until 2016, patient survey data is reported biannually (July/June and December), from 2016 only annually. We use the first measurement (July/June) for the primary analysis (a robustness test using the second measurement is presented in the Appendix, §3.3.) For the first item (whether patients would not recommend their practice), the exact survey question is "Would you recommend your GP surgery to someone who has just moved to your local area?", which patients can answer on a 5-point Likert scale from "Yes, would definitely recommend", "Yes, would probably recommend", "Not sure", and "No, would probably not recommend", "No, would definitely not recommend". In addition, there is a sixth response category "Don't know". For each PCP g in year t, we define the measure P^1_{gt} as the proportion of patients who would definitely or probably not recommend their PCP as a proportion of all surveyed patients. On average, 8.9% of the surveyed patients per PCP per year would not recommend their PCP (SD=0.068). For the second item (whether patients experienced problems accessing their PCP), the exact survey question is "Were you able to get an appointment to see or to speak to someone?" Patients have the following options, "Yes", "Yes, but I had to call back closer to or on the day I wanted the appointment", "No", and "Can't remember". For each PCP g in year t, we define the measure P^2_{gt} as the proportion of patients who stated that they could not get an appointment. On average, 13.1% of the surveyed patients per PCP per year claimed to have been unable to make an appointment (SD=0.060). Clearly, the higher the survey-based measures P^1_{gt} and P^2_{gt} are the worse the PCP performance as perceived by patients. The two measures P^1_{gt} and P^2_{gt} are positively correlated with each other (0.573, p<0.001) indicating that patients' perceptions point in the same, but not identical, direction.

To validate the measure $\widehat{u}_g$ produced in the previous section we estimate a mixed-effects model where the dependent variable is the survey-based measure of quality P^i_{gt}, with $i = \{1, 2\}$ and the model-estimated measures $\widehat{u}_g$ and $\widehat{\epsilon}_{gt}$, as estimated by equation (2), are the primary independent variables:

$$P^i_{gt} = \beta_0 + \beta_U \widehat{u}_g + \beta_\epsilon \widehat{\epsilon}_{gt} + \beta_C C_{gt} + \nu_g + \epsilon^P_{gt}, \tag{3}$$

Table 6 Validation of the performance measure

	(1) Not Recommending	(2) Not Recommending	(3) Had No Access	(4) Had No Access
$\widehat{u}_g$		0.120		0.223***
		(0.073)		(0.059)
$\widehat{\epsilon}_{gt}$		0.011		−0.009
		(0.015)		(0.020)
Closest hospital	−0.005	−0.005	0.001	0.002
	(0.020)	(0.019)	(0.017)	(0.015)
Depriviation rank	0.002	0.002	−0.001	−0.001
	(0.002)	(0.001)	(0.001)	(0.001)
Scale	−0.003	−0.003	0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.001)
Female	0.036	0.053	0.041	0.090
	(0.132)	(0.138)	(0.200)	(0.116)
Elderly	−0.142	−0.115	0.224	0.298
	(0.366)	(0.358)	(0.340)	(0.295)
CMI	0.084	0.076	−0.095	−0.119
	(0.083)	(0.084)	(0.065)	(0.061)
Constant	−0.015	−0.015	0.191	0.189**
	(0.109)	(0.112)	(0.104)	(0.073)
Year FE	Yes	Yes	Yes	Yes
Variance ν_g^P	0.003	0.003	0.002	0.002
Variance ν_g^P 95% CI	[0.001; 0.005]	[0.001; 0.005]	[0.001; 0.003]	[0.001; 0.002]
Intraclass correlation (ICC)	63.49%	62.68%	51.14%	47.26%
Model Wald χ^2	15.77	18.82	12.53	33.93
Observations	397	397	397	397
Number of groups	83	83	83	83

Clustered standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05.
Not Recommending: Proportion of patients that would not recommend their PCP to others.
Had No Access: Proportion of patients reporting problems with accessing their PCP.

where C_{gt} denotes the same vector of control variables as those included in model (2). Once again we decompose the error to a between PCP component (ν_g^P) and a within PCP component (ϵ_{gt}^P) and cluster at PCP level.

Note that the performance measures ($\widehat{u}_g$ and $\widehat{\epsilon}_{gt}$) that we use as independent variables in these regressions are model-generated estimates and therefore may be subject to measurement error. If the classical errors-in-variables assumptions hold (Wooldridge 2010), i.e. if we can assume that the measurement error is uncorrelated with the unobserved variable u_g, the estimated $\widehat{\beta}_U$ will be attenuated, rendering more conservative results. If $\widehat{u}_g$ is a sensible performance indicator, we expect the coefficient $\beta_U > 0$.

Table 6 presents the results for the proportion of patients who decline to recommend their PCP in Columns (1) and (2) and the results for the proportion of patients who claim to not have been able to get an appointment with their PCP in Columns (3) and (4). We find that the PCP's random effect $\widehat{u}_g$ is positively associated with the survey measures P_{gt}. Focusing on the model

of Column (2), the results show that if the estimated PCP's log-transformed proportion of ACS attendances is one standard deviation higher than the average, the proportion of patients who decline to recommend their PCP increases by $\sqrt{\tau^2} \times 0.120$, i.e. $\sqrt{0.009} \times 0.120 = 0.011$, which constitutes an increase of 12.8% relative to the mean proportion of patients declining to recommend their PCP (0.011/0.089). The p-value of the estimated coefficient is 10.2% (two-sided) indicating that the relationship may, in reality, be zero. More importantly as indicated by Column (4), if the PCP's log-transformed proportion of ACS attendances is one standard deviation higher than the average, the proportion of patients who claim to have experienced access problems increases by $\sqrt{0.009} \times 0.223 = 0.021$ leading to a 16.1% increase relative to the mean proportion of patients with access problems (0.021/0.131). The p-value of the estimated coefficient is less than 0.1%.

In contrast, to the highly informative PCP random effect $\widehat{u}_g$, we note that the stochastic fluctuation in PCP performance (the within-PCP variation $\widehat{\epsilon}_{gt}$) does not have any operationally or statistically significant impact on the perception of PCP performance. This gives credibility to the claim that ϵ_{gt} indeed only captures the residual (i.e., random) noise in performance as defined in equation (2).

Taken together, these results indicate that the measure $\widehat{u}_g$ estimated in §3 is a valid indicator of performance as it is positively related with patient survey outcomes. Furthermore, we believe that assessing PCP performance using the variance decomposition method outlined above may be more reliable compared to using patient surveys for at least three reasons. First, the measure based on ED data is objective and suffers less from human bias (e.g., imperfect recall, confirmation bias, regional variation in propensity to complain, etc.) that may creep-in in any qualitative survey. Second, surveys are only conducted infrequently (every 6-12 months), capture a small cross section of patients, and are expensive to administer. In contrast, the variance decomposition method could, in principle, be estimated frequently (as new data becomes available) without additional administrative costs. Third, the variance decomposition method allows one to estimate the impact of inferior PCP performance on the patient attendances at the ED. For example, this could be used to estimate ED cost savings associated with a policy that helps underperforming PCPs to catch up with their better performing peers. This is a point we demonstrated in §3.5 and expand on in §5. Such analysis would be impossible to perform with survey data alone, which we also demonstrate in Appendix 7.

4.2. A measure based on ED-sensitive conditions: A placebo test

In this section we provide an additional investigation that suggests that the performance measure constructed using the variance decomposition methodology proposed in §3 is indeed capturing systematic PCP differences as opposed to unobserved patient-level heterogeneity. We do so by

repeating the analysis of §3 but instead of using patients with ACS conditions we focus on a subset of patients that attend the ED with potentially life threatening acute conditions that are less influenced by PCP quality of care (which we call ED-sensitive conditions). For these conditions, the variance decomposition methodology should fail to detect a large amount of systematic variation between PCPs (once we control for observable heterogeneity). In addition, one would expect any between-PCP variation detected based on ED-sensitive conditions to be unrelated to the patients perceptions of quality of care as recorded by the PCP-patient survey.

To compile a set of conditions for which ED treatments are required and are unrelated with PCP performance, we rely on the set of emergency care–sensitive conditions identified by Vashi et al. (2019). This set contains life-threatening conditions such as road accidents and poisoning, which are arguably completely unrelated to PCP quality of care. The set also includes conditions such as myocardial infraction, respiratory arrest, and pulmonary embolism. Such conditions may be somewhat influenced by PCP preventative care but do have a significant stochastic component that is independent of PCP actions. From this set we exclude any conditions that overlap with ACS conditions, for example acute complications relating to diabetes or COPD that require urgent ED treatment, as the onset of such episodes is – at least partially – sensitive to the management of chronic conditions by PCPs in the community. We further exclude any non-life threatening conditions for which the NHS advises patients to seek care in settings other than EDs (NHS 2018b). A complete list of ED-sensitive conditions appears in Appendix §6. 6.52% of ED attendances in our sample qualify as ED-sensitive.

For each PCP we calculate the number of ED-sensitive attendances E_{gt} as a proportion of all ED attendances and decompose it in a manner similar to the primary analysis:

$$ln(E_{gt}) = \gamma_0 + \gamma_C C_{gt} + \gamma_t Year_t + u_g + \epsilon_{gt}. \tag{4}$$

The results of model (4) are presented in Table 7. In the model with the most detailed controls, column (3), the ICC is equal to 27.21%, indicating that 27.21% of the variance is attributed to systematic differences between PCPs. This is indeed lower compared to the case of ACS conditions, in which 36.96% of the variance is attributed to systematic differences between PCPs.

Finally, we assess whether the performance measure based on ED-sensitive attendances is correlated with patient perception of quality as recorded in patient surveys using the same approach as in the §4.1. The results are presented in Table 8 and show that the PCP's estimated performance (the random effect $\widehat{u}_g$) as measured by ED-sensitive attendances is not associated with the survey measures P_{gt}. These findings suggest that the between PCP variability in ED-sensitive attendances is not related to PCP quality of care.

Table 7 Decomposing variation in PCP performance based on ED-sensitive conditions

	(1) $\ln(E_{gt})$	(2) $\ln(E_{gt})$	(3) $\ln(E_{gt})$
Closest hospital		0.022	0.003
		(0.058)	(0.066)
Depriviation rank		0.008*	0.006
		(0.004)	(0.004)
Scale			0.002
			(0.005)
Female			−2.047
			(1.111)
Elderly			2.149
			(1.620)
CMI			−0.114
			(0.301)
Constant	−2.820***	−2.894***	−1.833**
	(0.030)	(0.046)	(0.571)
Year FE	Yes	Yes	Yes
$\widehat{\tau}^2$	0.034	0.032	0.020
$\widehat{\tau}^2$, 95% CI	[0.012; 0.090]	[0.012; 0.085]	[0.012; 0.034]
Intraclass correlation (ICC)	39.01%	37.84%	27.05%
Model Wald χ^2	19.98	25.63	36.30
Observations	408	408	401
Number of groups	84	84	83

Clustered standard errors in parentheses.
*** p<0.001, ** p<0.01, * p<0.05.

4.3. Comparing measures based on ACS admissions vs ACS attendances

We believe that assessing PCP performance based on attendances of patients with ACS conditions is substantially different and, we would argue, better than measures that rely on hospital admissions of patients with ACS conditions which take place through the ED. Such measures have been used in the literature as means to compare PCPs performance at the national level (e.g. Blunt 2013, Harrison et al. 2014, Barker et al. 2017). First, ACS admissions are relatively rare events compared to attendances. The PCPs in our sample have, on average, 165 annual ACS ED attendances (SD:150) but only 41 annual ACS admissions (SD: 38). Relying on admissions would only allow to detect very serious PCP failures. It would not be well suited to detect more frequent but less harmful failures such as appointment delays and lack of out-of-hour provisions that result in ED attendances but not ED admissions. In contrast, ACS attendances are approximately four times more frequent and would be better suited for the latter. Second, ACS attendances have the advantage of being exogenous to the hospital as opposed to ACS admissions which could be affected by the hospital's occupancy, hospital targets, and ED physicians' preferences, all of which are known to affect admissions (Galarraga et al. 2015, Freeman et al. 2020).

We demonstrate this point more formally in Appendix §5, where we show that compared to attendances, the variation in admissions that can be attributed to between PCP performance is

Table 8 Validation of the performance measure based on ED-sensitive conditions

	(1) NotRec	(2) NoAccess
$\widehat{u}_g$	−0.043	−0.079
	(0.036)	(0.040)
$\widehat{\epsilon}_{gt}$	−0.004	−0.004
	(0.009)	(0.010)
Closest hospital	−0.005	0.001
	(0.020)	(0.016)
Depriviation rank	0.002	−0.001
	(0.002)	(0.001)
Scale	−0.003	0.001
	(0.001)	(0.001)
Female	0.042	0.058
	(0.134)	(0.154)
Elderly	−0.130	0.258
	(0.363)	(0.320)
CMI	0.083	−0.098
	(0.082)	(0.065)
Constant	−0.016	0.185∗
	(0.110)	(0.089)
Year FE	Yes	Yes
Variance ν_g^P	0.003	0.002
Variance ν_g^P 95% CI	[0.001; 0.005]	[0.001; 0.003]
Intraclass correlation (ICC)	63.29%	50.29%
Model Wald χ^2	18.458	16.23
Observations	397	397
Number of groups	83	83

Clustered standard errors in parentheses.
*** p<0.001, ** p<0.01, * p<0.05.
NotRec: Proportion of patients refraining to recommend their PCP. NoAccess: Proportion of patients experiencing access problems at their PCP

much lower (less than 24% for admissions compared to 37% in the case of attendances). Furthermore, we also show that the measure constructed using attendances is not positively related to the measures constructed using patient surveys, further suggesting that attendances as opposed to admissions are a more useful measure of PCP performance.

5. Illustrative Application: The Impact of PCP Staffing on ED Attendances

Having access to an objective measure of PCP performance, such as the one estimated in §3, would be helpful in enabling health authorities to improve PCP care. If augmented with PCP operational and clinical data (e.g., PCP staffing levels, operating hours, information about the clinical experience of medical providers, diagnostic facilities, etc) and perhaps qualitative information (e.g., by interviewing relevant stakeholder), it may allow identification of best practice that would enable other PCPs to improve care for their patients and avoid unnecessary ED attendances by patients with ACS conditions. Furthermore, one could use the methodology described in §3 to estimate

potential cost savings at the hospital ED level associated with any improvements implemented in primary care. In this section we present an illustrative example of how this could be done in practice.

The operational aspect we focus on is the patient-to-staff ratio, i.e. total number of patients registered with practice g per physician FTE employed by the practice g, denoted as PpP_g. Patient-to-staff ratios have been shown to be associated with quality of care differences in the hospital setting (Kane et al. 2007, Needleman et al. 2002, Pronovost et al. 2002) and in this section we explore whether this is also the case in the primary care setting. The PCPs in our sample exhibit considerable variability in their patient-to-staff ratios. On average, there are 2,049 patients registered for each FTE physician employee, with a standard deviation of 1,092 patients. The number of patients registered with a PCP to the number of FTE physicians working in that practice is an indicator of the physician's average workload and the practice congestion level. Everything else being equal, we expect a PCP with a higher patient-to-staff ratio to be more congested than a PCP with a lower patient-to-staff ratio. Congestion may lead to access problems and perhaps inferior quality of care for patients. Therefore we expect that higher patient-to-staff ratios would be associated with higher u_g, i.e. lower performance. To assess if this is indeed the case we calculate the average number of patients per FTE employee over the duration of our data and estimate the following model:

$$\widehat{u}_g = \gamma_0 + \gamma_P PpP_g + \epsilon_g^U. \tag{5}$$

In this model specification we do not need to include any control variables already included in model (2) since the performance measure $\widehat{u}_g$ is, by construction, orthogonal to these control variables. The results are presented in Column (1) of Table 9 and are illustrated graphically in Figure 5a. For every additional 1,000 patients per FTE, the measure $\widehat{u}_g$ increases by 0.028 ($p<0.01$) implying that increasing patient-to-staff ratio is indeed associated with more ED attendances by patients with ACS conditions.

We subsequently explore whether this effect is moderated by the size of the practice measured in number of patients registered. From an operations management perspective, it is likely to be the case that economies of scale (e.g., pooling) could allow larger practices to offer a good level of service to their patients at a higher patient-to-staff ratio compared to smaller practices. To investigate this we interact the patient-to-staff ratio with the practice scale ($scale_g$), measured in 1,000s of patients and averaged over the duration of the data.

$$\widehat{u}_g = \delta_0 + \delta_1 PpP_g + \delta_2 PpP_g \times Scale_g + \delta_3 Scale_g + \epsilon_g^U. \tag{6}$$

The results are presented in Column (2) of Table 9 and illustrated graphically in Figure 5b – indeed, the effect of the patient-to-staff ratio on ACS attendances seems to be driven by smaller practices.

Table 9 Relationship between performance measure and patient-to-staff ratio

	(1) $\widehat{u}_g$	(2) $\widehat{u}_g$
PpP	0.028**	0.080**
	(0.010)	(0.026)
Scale		0.016*
		(0.008)
PpP × Scale		−0.008*
		(0.004)
Constant	−0.058*	−0.161*
	(0.024)	(0.062)
Observations	83	83
R^2	0.073	0.124

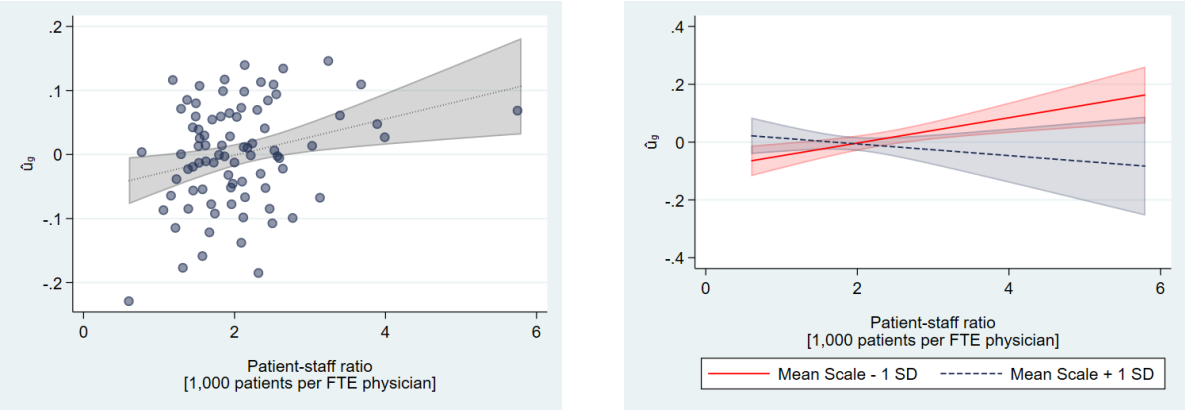
Robust standard errors in parentheses.
*** p<0.001, ** p<0.01, * p<0.05.

For a PCP one standard deviation below the mean in terms of scale (i.e., a PCP with 4,441 registered patients) the impact of an additional 1,000 patients per FTE on $\widehat{u}_g$ is 0.044 (p<0.01). In contrast the effect of an increase in the number of patients per FTE employee is statistically indistinguishable from zero for a PCP one standard deviation above the mean (i.e., a PCP with 12,340 registered patients).

Furthermore, if we combine these results with those of the model described by (2), these numbers imply that if the PCP operating at a scale of one standard deviation below the mean (i.e., a PCP with 4,441 registered patients) could decrease the number of patients per FTE employee by 1,000, it would see a reduction in ACS attendances at the ED of approximately 5.5% ($= \frac{1}{1-\pi}(0.0798 - 0.008103 \times 4.441)$, where $\pi = 0.196$ is the proportion of ACS attendances). Of course such a change would not be costless as it would necessitate hiring an additional 2.06 FTEs for a PCP operating at the average staffing level of 2.17 FTEs. In contrast, if two such PCPs (operating at a scale of one standard deviation below the mean) were to merge, this would raise the number of registered patients to 8,882 (which is close to the mean). Such a merger would reduce the number of ACS attendances by approximately 0.4% ($= \frac{1}{1-\pi}(0.01583 - 0.008103 \times 2.049) \times 4.441$) without requiring additional staff. However, it does assume that it is possible for two PCPs to merge and do so in an operationally effective way.

6. Conclusion and Further Research

This paper presents a quantitative methodology that exploits routinely collected ED operational data to identify PCPs whose patients place a disproportionably larger / lower burden on ED services. The methodology relies on comparing the number of ED attendances of patients with ACS conditions as a proportion to the total number of ED attendances and uses variance decomposition methods to identify systematic variation between PCP. We show that such systematic variation



(a) Estimated relationship as per column (1) of table 9 (b) Estimated relationship as per column (2) of table 9

Figure 5 Estimated relationship between patient-to-staff ratio and the performance measure

between PCPs is statistically and operationally important – our analysis suggests that between 37%-46% of the variation in ED attendances of patients with ACS conditions is due to systematic differences at the PCP-level.

As with any study, our findings should be interpreted in the light of the study's limitations. The first has to do with the limited amount of data used to illustrate the methodology – although the patients of each PCP are free to visit any ED we rely on data from a single ED. Nevertheless it should be straightforward to extend this methodology to multiple EDs, even to the whole of England. A second limitation is that the performance measure constructed by this work looks at historical PCP performance which it assumes is time invariant. In the five years of data in our sample this may not be a prohibitive limitation, and in any case, we had to make this assumption in order to have a sufficient number of observations per PCP, but for studies that monitor performance over longer horizons and across multiple hospitals it might be instructive to increase the frequency of the panel data (e.g., from annual to quarterly) and allow for time-varying systematic variation in PCP performance.

These imperfections notwithstanding, our findings have important implications. We demonstrate that the performance measure constructed has validity – the PCPs that score poorly are those for whom patients are more likely to complain with regards to timely access – and that it is better able to identify PCP variation than measures based on ED admissions used in extant literature. Therefore, this methodology can be used to quantify variation in the quality of primary care services provided within a region and, if appropriately extended to include data from multiple EDs, across different regions within England. Furthermore, we present an exploratory analysis that demonstrates how the measure constructed by this work can be used to identify and quantify the impact of operational drivers such as staffing on PCP performance. Further work could build on this methodology to identify clinical pathways and operations practices that enhance PCP

performance. More ambitiously, this methodology could be used to modify PCP reimbursement and thus, provide financial incentives for PCPs to reduce clinically unwarranted ED attendances. Such changes should, over time, improve the efficiency of providing care, and are particularly important to consider as the current model of providing primary care is under increasing pressure due to factors such as aging populations, reduced funding, and staff shortages.

References

Adams JL, Mehrotra A, Thomas JW, McGlynn EA (2010) Physician cost profiling—reliability and risk of misclassification. *New England Journal of Medicine* 362(11):1014–1021.

Barker I, Steventon A, Deeny SR (2017) Association between continuity of care in general practice and hospital admissions for ambulatory care sensitive conditions: cross sectional study of routinely collected, person level data. *The BMJ* 356:j84.

Bastani H, Goh J, Bayati M (2019) Evidence of upcoding in pay-for-performance programs. *Management Science* 65(3):1042–1060.

Batt RJ, Terwiesch C (2015) Waiting patiently: An empirical study of queue abandonment in an emergency department. *Management Science* 61(1):39–59.

Bavafa H, Canamucio A, Marcus S, Terwiesch C, Werner RM (2021) Capacity rationing in primary care: Provider availability shocks and channel diversion. *Management Science, forthcoming* .

Berchet C (2015) Emergency care services. *OECD Health Working Papers, OECD Publishing, Paris.* (83), URL `https://www.oecd-ilibrary.org/content/paper/5jrts344crns-en`.

Berry Jaeker J, Tucker A (2017) Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Science* 63(4):1042–1062.

Blunt I (2013) Focus on preventable admissions. *London: Nuffield Trust* URL `https://www.health.org.uk/sites/default/files/QualityWatch_FocusOnPreventableAdmissions.pdf`, last accessed: 05.10.2018.

Brown C, Lilford R (2006) Cross sectional sudy of performance indicators for english PCT's: Testing construct validity and identifying explanatory variables. *BMC Health Services Research* 6:81.

Busby J, Purdy S, Hollingworth W (2015) A systematic review of the magnitude and cause of geographic variation in unplanned hospital admission rates and length of stay for ambulatory care sensitive conditions. *BMC Health Services Research* 15(1):324.

Busby J, Purdy S, Hollingworth W (2017) How do population, general practice and hospital factors influence ambulatory care sensitive admissions: A cross sectional study. *BMC Family Practice* 18(1):1–9.

Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M (2009) Effects of pay for performance on the quality of primary care in england. *New England Journal of Medicine* 361(4):368–378.

Clarke GM, Conti S, Wolters AT, Steventon A (2019) Evaluating the impact of healthcare interventions using routine data. *The BMJ* 365:l2239.

Cowling T, Harris M, Majeed A (2015) Evidence and rhetoric about access to UK primary care. *The BMJ* 350:h1513.

Cowling TE, Harris M, Majeed A (2016) Extended opening hours and patient experience of general practice in england: Multilevel regression analysis of a national patient survey. *BMJ Quality & Safety* 26:360–371.

Dolton P, Pathania V (2016) Can increased primary care access reduce demand for emergency care? Evidence from England's 7-day GP opening. *Journal of Health Economics* 49:193–208.

Downing A, Rudge G, Cheng Y, Tu YK, Keen J, Gilthorpe MS (2007) Do the UK government's new quality and outcomes framework (QOF) scores adequately measure primary care performance? A cross-sectional survey of routine healthcare data. *BMC Health Services Research* 7(1):1–7.

Dusheiko M, Gravelle H, Jacobs R, Smith P (2006) The effect of financial incentives on gatekeeping doctors: evidence from a natural experiment. *Journal of Health Economics* 25(3):449–478.

Dusheiko M, Gravelle H, Martin S, Rice N, Smith PC (2011) Does better disease management in primary care reduce hospital costs? Evidence from english primary care. *Journal of Health Economics* 30(5):919–932.

Filistrucchi L, Prüfer J (2019) Faithful strategies: How religion shapes nonprofit management. *Management Science* 65(1):188–208.

Freeman M, Robinson S, Scholtes S (2020) Gatekeeping under congestion: An empirical study of referral errors in the emergency department. *Management Science, forthcoming* .

Freeman M, Savva N, Scholtes S (2017) Gatekeepers at work: An empirical analysis of a maternity unit. *Management Science* 63(10):3147–3167.

Galarraga JE, Mutter R, Pines JM (2015) Costs associated with ambulatory care sensitive conditions across hospital-based settings. *Academic Emergency Medicine* 22(2):172–181.

Gentil ML, Cuggia M, Fiquet L, Hagenbourger C, Le Berre T, Banâtre A, Renault E, Bouzille G, Chapron A (2017) Factors influencing the development of primary care data collection projects from electronic health records: A systematic review of the literature. *BMC Medical Informatics and Decision Making* 17(1):1–21.

Gibson OR, Segal L, McDermott RA (2013) A systematic review of evidence on the association between hospitalisation for chronic disease related ambulatory care sensitive conditions and primary health care resourcing. *BMC Health Services Research* 13(1):1–13.

Green L, Savin S, Savva N (2013) "Nursevendor problem": Personnel staffing in the presence of endogenous absenteeism. *Management Science* 59(10):2237–2256.

Harrison MJ, Dusheiko M, Sutton M, Gravelle H, Doran T, Roland M (2014) Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: Controlled longitudinal study. *The BMJ* 349:g6423.

Hong CS, Atlas SJ, Chang Y, Subramanian S, Ashburner JM, Barry MJ, Grant RW (2010) Relationship between patient panel characteristics and primary care physician clinical performance rankings. *JAMA* 304(10):1107–1113.

Huntley A, Lasserson D, Wye L, Morris R, Checkland K, England H, Salisbury C, Purdy S (2014) Which features of primary care affect unscheduled secondary care use? A systematic review. *BMJ Open* 4(5):e004746.

Ibanez M, Toffel M (2020) How scheduling can bias quality assessment: Evidence from food-safety inspections. *Management Science* 66(6):2396–2416.

Ibanez MR, Clark JR, Huckman RS, Staats BR (2018) Discretionary task ordering: Queue management in radiological services. *Management Science* 64(9):4389–4407.

Ismail SA, Gibbons DC, Gnani S (2013) Reducing inappropriate accident and emergency department attendances: A systematic review of primary care service interventions. *British Journal of General Practice* 63(617):e813–e820.

Johnson PJ, Ghildayal N, Ward AC, Westgard BC, Boland LL, Hokanson JS (2012) Disparities in potentially avoidable emergency department (ED) care: ED visits for ambulatory care sensitive conditions. *Medical Care* 1020–1028.

Jones HE, Spiegelhalter DJ (2011) The identification of "unusual" health-care providers from a hierarchical model. *The American Statistician* 65(3):154–163.

Jürges H, Köberlein J (2015) What explains drg upcoding in neonatology? The roles of financial incentives and infant health. *Journal of Health Economics* 43:13–26.

Kane RL, Shamliyan TA, Mueller C, Duval S, Wilt TJ (2007) The association of registered nurse staffing levels and patient outcomes: Systematic review and meta-analysis. *Medical Care* 1195–1204.

KC D, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* 55(9):1486–1498.

KC D, Terwiesch C (2017) Benefits of surgical smoothing and spare capacity: An econometric analysis of patient flow. *Production and Operations Management* 26(9):1663–1684.

Kim SH, Chan CW, Olivares M, Escobar G (2015) ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* 61(1):19–38.

Kuntz L, Mennicken R, Scholtes S (2015) Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science* 61(4):754–771.

Lavoie JG, Wong ST, Ibrahim N, O'Neil JD, Green M, Ward A (2019) Underutilized and undertheorized: The use of hospitalization for ambulatory care sensitive conditions for assessing the extent to which primary healthcare services are meeting needs in british columbia first nation communities. *BMC Health Services Research* 19(1):1–10.

MacKichan F, Brangan E, Wye L, Checkland K, Lasserson D, Huntley A, Morris R, Tammes P, Salisbury C, Purdy S (2017) Why do patients seek primary medical care in emergency departments? An ethnographic exploration of access to general practice. *BMJ Open* 7(4):e013816.

Maisey S, Steel N, Marsh R, Gillam S, Fleetcroft R, Howe A (2008) Effects of payment for performance in primary care: Qualitative interview study. *Journal of Health Services Research & Policy* 13(3):133–139.

McCarthy ML, Zeger SL, Ding R, Levin SR, Desmond JS, Lee J, Aronsky D (2009) Crowding delays treatment and lengthens emergency department length of stay, even among high-acuity patients. *Annals of Emergency Medicine* 54(4):492–503.

McDonald R, Roland M (2009) Pay for performance in primary care in england and california: Comparison of unintended consequences. *The Annals of Family Medicine* 7(2):121–127.

Mundlak Y (1978) On the pooling of time series and cross section data. *Econometrica* 46(1):69–85.

National Audit Office (2014) Out-of-hours gp services in england. *London: National Audit Office* URL `https://www.nao.org.uk/report/hours-gp-services-england-2/`.

Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K (2002) Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine* 346(22):1715–1722.

NHS (2016) Registering with a gp surgery outside the area you live. URL `https://www.nhs.uk/using-the-nhs/nhs-services/gps/patient-choice-of-gp-practices/`, last accessed: 05.10.2018.

NHS (2017) National tariff 2017/19: document and policies. URL `https://www.england.nhs.uk/publication/past-national-tariffs-documents-and-policies/`, last accessed: 16.07.2021.

NHS (2018a) Hospital accident and emergency activity, 2016-17. URL `https://digital.nhs.uk/data-and-information/publications/statistical/hospital-accident--emergency-activity/2016-17`, last accessed: 16.07.2021.

NHS (2018b) When to visit an urgent treatment centre (walk-in centre or minor injury unit). URL `https://www.nhs.uk/nhs-services/urgent-and-emergency-care-services/when-to-visit-an-urgent-treatment-centre-walk-in-or-minor-injury-unit/`, last accessed: 27.08.2021.

Oster A, Bindman AB (2003) Emergency department visits for ambulatory care sensitive conditions: Insights into preventable hospitalizations. *Medical Care* 41(2):198–207.

Paddock SM (2014) Statistical benchmarks for health care provider performance assessment: a comparison of standard approaches to a hierarchical bayesian histogram-based method. *Health Services Research* 49(3):1056–1073.

Paddock SM, Adams JL, de la Guardia FH (2015) Better-than-average and worse-than-average hospitals may not significantly differ from average hospitals: an analysis of medicare hospital compare ratings. *BMJ Quality & Safety* 24(2):128–134.

Picard R (2012) GEONEAR: Stata module to find nearest neighbors using geodetic distances. *Statistical Software Components* .

Powell A, Savin S, Savva N (2012) Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing & Service Operations Management* 14(4):512–528.

Pronovost PJ, Angus DC, Dorman T, Robinson KA, Dremsizov TT, Young TL (2002) Physician staffing patterns and clinical outcomes in critically ill patients: A systematic review. *JAMA* 288(17):2151–2162.

Racz MJ, Sedransk J (2010) Bayesian and frequentist methods for provider profiling using risk-adjusted assessments of medical outcomes. *Journal of the American Statistical Association* 105(489):48–58.

Roland M, Guthrie B (2016) Quality and outcomes framework: What have we learnt? *The BMJ* 354:i4060.

Rosano A, Loha CA, Falvo R, Van der Zee J, Ricciardi W, Guasticchi G, De Belvis AG (2013) The relationship between avoidable hospitalization and accessibility to primary care: A systematic review. *The European Journal of Public Health* 23(3):356–360.

Savva N, Tezcan T, Yıldız Ö (2019) Can yardstick competition reduce waiting times? *Management Science* 65(7):3196–3215.

Schoen C, Osborn R, Huynh PT, Doty M, Davis K, Zapert K, Peugh J (2004) Primary care and health system performance: Adults' experiences in five countries: Differing performance levels among countries highlight the potential for improvement and cross-national learning. *Health Affairs* 23(Suppl1):W4–487.

Scott A, Sivey P, Ouakrim DA, Willenberg L, Naccarella L, Furler J, Young D (2011) The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database of Systematic Reviews* (9).

Shleifer A (1985) A theory of yardstick competition. *The RAND Journal of Economics* 16(3):319–327.

Soltani M, Batt R, Bavafa H, Patterson B (2020) Does what happens in the ED stay in the ED? The effects of emergency department physician workload on post-ED care use. *Working Paper* .

Song H, Andreyeva E, David G (2021) Time is the wisest counselor of all: The value of provider–patient engagement length in home healthcare. *Management Science, forthcoming* .

Song H, Tucker A, Murrell K (2015) The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science* 61(12):3032–3053.

Song H, Tucker AL, Graue R, Moravick S, Yang JJ (2020) Capacity pooling in hospitals: The hidden consequences of off-service placement. *Management Science* 66(9):3825–3842.

Song H, Tucker AL, Murrell KL, Vinson DR (2018) Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices. *Management Science* 64(6):2628–2649.

Spiegelhalter DJ (2005) Funnel plots for comparing institutional performance. *Statistics in Medicine* 24(8):1185–1202.

Staats B, Gino F (2012) Specialization and variety in repetitive tasks: Evidence from a japanese bank. *Management Science* 58(6):1141–1159.

Tan TF, Netessine S (2014) When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science* 60(6):1574–1593.

Thomas JW, Grazier KL, Ward K (2004) Economic profiling of primary care physicians: Consistency among risk-adjusted measures. *Health Services Research* 39(4p1):985–1004.

van Loenen T, van den Berg MJ, Westert GP, Faber MJ (2014) Organizational aspects of primary care related to avoidable hospitalization: A systematic review. *Family Practice* 31(5):502–516.

Vashi AA, Urech T, Carr B, Greene L, Warsavage T, Hsia R, Asch SM (2019) Identification of emergency care–sensitive conditions and characteristics of emergency department utilization. *JAMA Network Open* 2(8):e198642–e198642.

Vuik SI, Fontana G, Mayer E, Darzi A (2017) Do hospitalisations for ambulatory care sensitive conditions reflect low access to primary care? An observational cohort study of primary care usage prior to hospitalisation. *BMJ Open* 7(8):e015704.

Wang J, Zhou YP (2018) Impact of queue configuration on service time: Evidence from a supermarket. *Management Science* 64(7):3055–3075.

Watson AG, McCoy JV, Mathew J, Gundersen DA, Eisenstein RM (2019) Impact of physician workload on burnout in the emergency department. *Psychology, Health & Medicine* 24(4):414–428.

Weber EJ, Mason S, Freeman JV, Coster J (2012) Implications of england's four-hour target for quality of care and resource use in the emergency department. *Annals of Emergency Medicine* 60(6):699–706.

Weinick RM, Burns RM, Mehrotra A (2010) Many emergency department visits could be managed at urgent care centers and retail clinics. *Health Affairs* 29(9):1630–1636.

Wooldridge JM (2010) *Econometric analysis of cross section and panel data* (MIT Press).

Xu Y, Tan T, Netessine S (2020) The impact of workload on operational risk: Evidence from a commercial bank. *SMU Cox: IT & Operations Management (Topic)* .