Medical Labour Supply and the Production of Healthcare*

TOM LEE, † CAROL PROPPER‡ and GEORGE STOYE§

†Institute for Fiscal Studies
(tom.lee@ifs.org.uk)
‡Imperial College
(c.propper@imperial.ac.uk)
§Institute for Fiscal Studies; University College London
(george_s@ifs.org.uk)

Abstract

Medical labour markets are important because of their size and the importance of medical labour in the production of healthcare and in subsequent patient outcomes. We present a summary of important trends in the UK medical labour market, and we review the latest research on factors that determine medical labour supply and the impact of labour on patient outcomes. The topics examined include: the responsiveness of labour supply to changes in wages, regulation and other incentives; factors that determine the wide variation in physician practice and style; and the effect of teams and management quality on patient outcomes. This literature reveals that while labour supply is relatively unresponsive to changes in wages, medical personnel do react strongly to other incentives, even in the short run. This is likely to have consequences for the

* Submitted March 2019.

We thank Monica Costa Dias and three anonymous referees for their comments. We thank the Department of Health and Social Care for access to data from the NHS Electronic Staff Record (ESR). Figures published using the ESR may be different from the official workforce statistics published by NHS Digital. The authors acknowledge financial support from the UK ESRC through the Centre for Microeconomic Analysis of Public Policy at IFS (ES/M010147/1) and the National Institute for Health Research through the Health and Social Care Workforce Policy Research Unit. C. Propper acknowledges funding by the ERC through Award 788529.

Keywords: healthcare production, medical labour force, regulation.

JEL classification numbers: I11, I18, J22, J24, J45.

© 2020 The Authors. Fiscal Studies published by John Wiley & Sons Ltd. on behalf of Institute for Fiscal Studies.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
quality of care provided to patients. We set out a series of unanswered questions
in the UK setting, including: the importance of non-financial incentives in
recruiting and retaining medical staff; how individuals can be incentivised to
work in particular specialties and regions; and how medical teams can be best
organised to improve care.

I. Introduction

Healthcare systems around the world face the twin problems of increasing
demand and increasing costs of supply. The UK is no exception. Recent
estimates from the UK suggest that an ageing population alone will increase
demand for healthcare by 3.3 percent over the next 15 years.1 In reality, the
demand for healthcare is likely to rise by even more as income increases and
technological advancements raise expectations of what the healthcare system
should provide. As the UK publicly provides healthcare, meeting this rise
in demand will require some combination of increased taxes, lower spending
elsewhere and more efficient ways of producing healthcare. Thus, the increased
demand for care means staffing (and the related financing) issues are central
to the UK policy agenda.

Healthcare is labour intensive, regardless of the nature of the health
system, its structural characteristics, levels and sources of funding, or political
underpinnings. In 2016, UK public and private health spending accounted for
nearly 10 per cent of GDP2 and individuals employed by the National Health
Service (NHS) accounted for 10 per cent of the workforce.3 The NHS directly
employs over 1.2 million persons. This includes around 620,000 professionally
qualified clinical staff. In addition, the NHS is the sole buyer of services from
over 7,500 General Practices who provide primary care (known as General
Practitioners in the UK). The NHS wage bill, funded entirely from the public
purse, was over £50 billion in 2016/17.4 This large labour force means that the
medical labour market is an important labour market for the functioning of the
economy.

Further, there is evidence of considerable and persistent variation in
healthcare spending and medical outcomes across provider organisations,
which serve broadly similar populations with broadly similar needs. For the
UK, there have been long documented differences in performances on various
dimensions across similar NHS providers. The 10 per cent of hospitals with
the lowest risk-adjusted mortality rates have risk-adjusted mortality rates that

1Charlesworth et al., 2018.
2Lee and Stoye, 2018a.
3See ONS (2018) and Skills for Care (2018). These shares are even greater in other developed countries.
For example, in the US in 2016, health spending accounted for 17.1 per cent of GDP and the healthcare
4See NHS (2017) and Department of Health and Social Care (2018).

© 2020 The Authors. Fiscal Studies published by John Wiley & Sons Ltd. on behalf of Institute for Fiscal Studies
are 21 per cent lower than the national average,5 while the top 10 per cent of GP practices perform 29 per cent better on the Quality Outcomes Framework, a scheme that rewards practices for performing well on clinical indicators, than the bottom 10 per cent.6

Similar issues are evident in the US, the setting for much of the existing evidence on the drivers of variation in medical spending and outcomes. There are large differences in healthcare spending and use across and within US regions,7 with medical outcomes largely unrelated to use.8 These differences cannot be fully explained by random fluctuations, differences in prices across regions, income, health status9 and preferences of patients.10 Recent studies have instead pointed to the fact that these differences are in part related to persistent productivity differentials across providers of care within regions.11 But while variations are well-established, there remains much less understanding of why these persistent productivity differentials exist. One obvious potential contributor is the behaviour of medical labour in response to different incentives.

The healthcare labour market is, of course, not a homogeneous mass of workers. The workforce is characterised by long training periods for some workers, extensive regulation of entry and of conduct once in the sector, a broad set of employment contracts and types of employment, and individuals frequently working in teams. The healthcare sector is also a ‘mission-orientated’ sector12 and therefore will attract individuals who are motivated by non-financial (altruistic) as well as financial concerns. In order to design policy so as to recruit and retain a suitable workforce, and to reward them so that they are more productive, it is therefore necessary to have a better understanding of staff choices and motivations. This includes an understanding not only of the role that pay and non-pay factors play in staff joining and leaving the workforce, but also of the choices that staff make during and after training (e.g. over specialty and geography). It also means knowing how to best motivate and organise staff in order to achieve the best patient outcomes.

The focus of the present paper is to review recent empirical contributions to research the behaviour of medical labour in response to incentives, both

---

5Authors’ calculations from the NHS Digital publication, ‘Summary Hospital-level Mortality Indicator (SHMI) – Deaths associated with hospitalisation, England, October 2017–September 2018’.
6Authors’ calculations from the NHS Employers/ British Medical Association publication, ‘Quality and Outcomes Framework for 2012/13’. Castelli et al. (2015) show that there are extensive differences in NHS acute hospital productivity.
8See Fisher et al. (2003), Institute of Medicine (2013) and Doyle, Graves and Gruber (2017).
9Finkelstein et al., 2016.
10Barnato et al., 2007.
11See Skinner and Staiger (2015). The magnitudes are similar to those found in the manufacturing sector (Chandra et al., 2016).
pecuniary and non-pecuniary. Our aim is to complement extant recent UK studies of policy on the NHS labour force to provide policymakers and researchers with a guide to recent economics research on labour market issues in healthcare.\textsuperscript{13} We review the literature, which falls into two broad topic areas. The first area is concerned with what can be thought of as ‘classic’ labour supply issues: the response of medical labour to prices and income changes. Topics covered include the effect of changes in wages and income on short-run responses to supply of hours and effort, geographic and sector recruitment, choice of specialty and retention. We also briefly examine papers that address the long-standing concern that the asymmetry of information between physicians and patients means that physicians are able to induce demand (known as supplier induced demand) and the role of altruism in physician decision-making.

The second area is the growing body of work that seeks explanations for what drives the well-documented productivity differences in labour in the healthcare sector. The papers we review aim to understand the role of labour in the production function for healthcare; in other words, how labour responses to a variety of incentives affect the production of healthcare. Here, we examine literature covering issues such as drivers of differences between physicians (so-called ‘physician style’), the role of teams in producing medical care, and the effect of management quality on medical-care output.

Our choice of papers is motivated by a focus on those that have robust empirical designs, meaning that the findings can, wherever possible, be interpreted as causal. As many of these studies are not from the UK, we also include some UK literature, which might have a less robust design but is relevant to the UK situation.\textsuperscript{14} We also survey findings from the Nordic countries, which have healthcare systems that are very similar to the UK in terms of public funding of healthcare, public employment in the hospital sector and public contracting with independent family doctors in the community sector. We highlight the findings that are most relevant to the UK context.

The organisation of this paper is as follows. In the next section, we begin by providing some key descriptions of the current UK medical labour market. In Section III, we review papers that deal with labour supply issues and related financial incentives. In Section IV, we review recent contributions to the role of healthcare labour in healthcare production. In Section V, we conclude by highlighting those areas for which further research is particularly needed in the UK context.

\textsuperscript{13}See, for example, Buchan, Seccombe and Charlesworth (2016) and Health Education England (2017). For an existing extensive review of the economics of issues in the medical labour market, see Nicholson and Propper (2011). This review examined a large set of issues in medical labour supply including market failures and government intervention in medical labour markets, medical labour supply elasticities and variation in productivity across organisations and markets, from both theoretical and empirical perspectives.

\textsuperscript{14}The lack of robust studies for the UK for some areas stems from a lack of available public data.
II. Trends and issues in the UK medical labour market

The UK medical labour market is dominated by a single public employer (the NHS). The NHS employs over one million people, including doctors, nurses, midwives, healthcare professionals and support staff.\textsuperscript{15} The occupational mix of NHS staff has also changed over time. Since 1996, there has been strong growth in the number of medical staff per population employed in hospitals.\textsuperscript{16} Growth in the number of hospital doctors per head has been particularly strong, increasing by 72 per cent between 1996 and 2016. The numbers of nurses, midwives and health visitors per head also increased by 12 per cent over this period, although growth has been stagnant since 2010. In contrast, the number of GPs per person fell by 5 per cent over the same 20-year period, reflecting a shift in medical spending and activity towards the hospital sector, which has led to concerns about a shortage of GPs and funding for primary care.\textsuperscript{17}

NHS staff shortages have been a recurrent theme in the media. Importantly, these shortages do not indicate that members of staff are leaving the NHS, but that growth in the medical workforce has been outstripped by the growth in utilisation. This is demonstrated by Figure 1, which shows that the growth in the number of inpatient episodes and outpatient appointments between 2010 and 2016 was much larger than the growth in the number of hospital doctors and nurses. These pressures are likely to only intensify in the near future, with recent forecasts suggesting that the NHS will need to find 440,000 more staff by 2033/34 to keep pace with demand increases.\textsuperscript{18}

The monopsonistic role played by the NHS means that changes in the total amount of financial resources available to the NHS have a direct effect on the demand for labour. This contrasts with other countries with multiple healthcare employers, such as the US. This has important consequences both for total medical labour demand in the UK – with fiscal austerity measures reducing hours for long-run employed staff, and reducing hiring – and for the mix of permanent and temporary staff. Furthermore, a reduction in demand might also affect the longer-run supply of labour. Changes in short-run demand for labour can have long-run effects on labour supply, perhaps, for example, affecting the initial decision to train in the medical profession.

\textsuperscript{15}This number has grown substantially over time. When the NHS was founded in 1949, there were 11,735 full-time equivalent doctors working in NHS hospitals (https://digital.nhs.uk/news-and-events/latest-news/workforce-factsheet). By 2018, this number had grown almost ten times, to 109,509. However, despite this increase, the UK still has far fewer doctors and nurses per person than most developed countries. In 2015, the UK had 2.8 doctors per 1,000 people, 28 per cent lower than the EU-15 average (Lee and Stoye, 2018b).

\textsuperscript{16}Lee and Stoye, 2018b.

\textsuperscript{17}Iacobucci, 2019.

\textsuperscript{18}Charlesworth et al., 2018.
FIGURE 1
Growth in healthcare usage and NHS workforce 2010–16

Note: Year refers to the financial year. Nurses and doctors are WTE averaged across the year. FCEs and outpatient attendances come from Hospital Episode Statistics. Figures refer to nurses and doctors working in Hospital and Community Health Service (HCHS). The category ‘Nurses’ also includes health visitors.

the specialty choice of medical staff and exit decisions for individuals near retirement.\(^{19}\)

One potentially important factor in recruiting and retaining NHS staff is the structure and level of pay. Another consequence of having a dominant employer is that most members of staff in the NHS are employed on national pay scales. A benefit of this approach is that it is relatively easy to control total wage expenditure. Over the past decade, there have been a series of pay ‘freezes’ designed to limit spending under austerity, which might have affected the ability of the NHS to recruit and retain staff. In addition, over the past decade, public-sector jobs have become relatively less attractive compared with private-sector alternatives.\(^{20}\) The centralised system also affects the ability of individual NHS provider organisations to respond flexibly to local labour market challenges.

The NHS is a large employer of temporary staff. The use of such staff has historically been used to manage peaks and troughs in demand, which are a natural part of medical care (e.g. winter flu, annual leave patterns, etc). However, in recent years, temporary staff has been used increasingly to deal with long-run staff shortages. These shortages are a result of changing staffing

\(^{19}\)A current and pertinent example at the time of writing is Brexit, which affects both the resources available to the NHS and thus the overall demand for labour; it also affects the composition of the labour force and the pool from which labour can be sourced. We document the current reliance of the NHS on EU nurses later in this section.

\(^{20}\)Cribb, 2017.
patterns due to concerns over quality, high levels of attrition from nurse training and insufficient recruitment from abroad. This comes at a cost: agency workers are more costly than full-time employees and may also be of lower quality. As a result, in recent years there have been strong attempts to reduce the growth in agency spend.

The NHS operates both within the UK labour market and in an international labour market. Thus, factors that affect the general labour market in the UK, such as changes in the age of retirement for women or the relationship between regulated pay in the NHS and pay in the outside (non-NHS) labour market will affect the supply of staff. For example, previous work has shown that changes to the female state pension age led to large increases in female employment at older ages. Figure 2 illustrates that this was also likely the case for NHS nurses. In 2012–13, nearly 20 per cent of nurses aged 60 – who were, at that time, eligible for the state pension at age 60 – left the NHS that year. By 2017–18, the exit rate had fallen to only 10 per cent among nurses aged 60, who were no longer eligible for the state pension at that age. In addition, opportunities offered outside the UK will affect staff. In 2016, nearly three-quarters of NHS organisations were actively trying to recruit from abroad.

---

21Francis, 2013.
22See, for example, Monitor and NHS Trust Development Authority (2015).
23Propper and Van Reenen, 2010.
24Cribb, Emmerson and Tetlow, 2016.

© 2020 The Authors. Fiscal Studies published by John Wiley & Sons Ltd. on behalf of Institute for Fiscal Studies
This international nature of the labour market is particularly the case for nurses, for whom regulation of entry is less stringent than for doctors. Figure 3 shows the average annual change in nurses (headcount) by age group and nationality. It is clear that foreign nurses have been a key source of foreign recruitment. From 2012 to 2016, the increase in nurses was driven by young nurses from the European Union (EU), while since 2016 the largest growth has been in young non-EU nurses. However, this growth in nurses masks the failure to keep up with increased demand, which is particularly problematic in certain areas.26

The challenge for policymakers in this setting has been nicely outlined by Buchan et al. (2016): ‘Too often the approach to addressing workforce pressure points is reactive, and sometimes it is also unconnected to (or even counter to) broader issues of staffing and funding.’ Pressure points will come and go; what is needed is an understanding of how the financial and regulatory tools at policymakers’ disposal can be used to attract, retain and motivate healthcare workers to produce more from the limited resources available. Our review examines recent economic evidence relevant to these longer-run, and perennial, issues.

26At present, the UK government has identified shortages in adult nursing, general practice and primary care, pre-registration pharmacy, mental health workforce, emergency medicine, paramedics, cancer and diagnostics (Buchan et al., 2016). These shortages reflect changes in demand, which are not matched by growth in funding (e.g. cancer care and mental health), but also result from the responses of supply to the conditions of employment in the NHS and alternative labour market opportunities.
III. Medical Labour Supply

As in any other labour market, financial incentives influence the decision to enter or exit the market (the extensive margin) and the number of hours of labour to supply conditional on being employed (the intensive margin). Decisions on both margins will be affected not only by wages within the sector but also by wages in alternative occupations. For some workers, these alternative occupations might be close competitors but for others who are highly specialised, the outside options might be more limited in the short run.

In the medical-care market, interest in the extensive margin includes not only the effect of incentives on labour force participation but also on occupational choice, specialty choice and geographic location. The long training period of medical staff means that responses at the extensive margin have long-run effects on the composition and location of the labour force. Many healthcare systems spend a considerable amount of resources undertaking workforce planning and training so that they do not experience shortages in certain fields. In addition, in many countries, there is concern to ensure an even spread of medical staff to populations across geographical locations that vary in terms of amenity. The extensive margin is also affected by regulation of entry into medical professions and the behaviour of medical labour once qualified.

Even in the NHS (and other NHS-type systems), a significant proportion of medical labour is not employed directly by the public employer but is contracted to deliver services. These contracts take a variety of forms: one form is a contract for provision of services over a certain time period (known as ‘capitation payment’); another is payment per item of care delivered, where the price may be set ex ante (known as ‘prospective payment’) or ex post (known as ‘fee-for-service’); another is payment for bundles of care delivered or a payment for achieving a certain ‘care pathway’. Increasingly, healthcare policymakers tie payment to the achievement of (measurable) quality (often known as ‘pay-for-performance’). Contracts can be linear or non-linear contracts. Common examples of non-linear payments are those related to targets being met.

Thus, labour supply issues in medical-care markets include short-run and longer-run responses of individuals to relative wages and income but also responses to contract types (e.g. changes that relate payments to quality of care, where before they were only related to volume, or changes that relate payments to achievement of targets, whereas before they were not related to these targets). Regulation, of both entry into medical professions and the
behaviour of medical labour once qualified, means that regulatory changes also affect labour supply decisions.  

1. Short-run wages and income elasticities

a) Physicians

Early studies used aggregate time series data to estimate the labour supply curve of physicians. The general conclusion from this early research was that the labour supply curve was highly inelastic or even backward bending; in other words, hours were not very responsive to wage changes. However, there were two problems that this research did not fully overcome. First, when looking at the response of hours to wage rates, either an individual’s wage rates can be determined by the hours they work (known as endogeneity of the wage rate) or both hours and wage rates are driven by common unobserved factors (known as omitted variable bias). Recent research has used innovative identification strategies to try to overcome the endogeneity problem. Second, these papers are only able to examine the short-run elasticity.

In other labour markets, a change in remuneration will lead to a short-run effect as workers adjust their hours, followed by a longer-run effect as labour adjusts across different markets. For physicians, the longer-run effect is harder to identify because of restrictions on who can practise as a physician. If remuneration for physicians in a particular country increases, this might lead to more people wanting to work as a physician in that country. However, extensive training programmes and immigration controls for physicians mean that the adjustment process will take longer, and this is hard to identify separately from other confounding factors.

Table 1 provides a summary of key papers. Much research is US-based, where physicians tend to be self-employed. To the extent that employed (hospital-based) UK physicians have less flexibility with regards to hours, we would expect their labour supply to be more inelastic. For this reason, we also survey evidence from Norway, whose system shares many features in common with the UK.

For the US, Rizzo and Blumenthal (1994) use experience as an instrument for physicians’ wages, focusing on physicians aged 40 and under to remove retirement considerations. They find a short-run wage elasticity of 0.27 and an income elasticity of –0.17, with a larger effect for women than men. Showalter and Thurston (1997) use variation across US states in the maximum marginal tax rate to instrument for wages. They find that self-employed physicians have a small short-run wage elasticity (0.3), whereas employed physicians


TABLE 1

Physician labour supply

<table>
<thead>
<tr>
<th>Paper</th>
<th>Years</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feldstein (1970)</td>
<td>1948–66</td>
<td>US</td>
<td>Bureau of Labour Statistics</td>
<td>Reduced form</td>
<td>Physician labour supply curve is backward bending</td>
</tr>
<tr>
<td>Sloan (1975)</td>
<td>1960–70</td>
<td>US</td>
<td>Census</td>
<td>IV for wage using experience</td>
<td>Physicians do not respond to changes in wages</td>
</tr>
<tr>
<td>Rizzo and Blumenthal (1994)</td>
<td>1987</td>
<td>US</td>
<td>AMA Young Physician Survey</td>
<td>IV for wage using experience</td>
<td>Uncompensated wage elasticity of 0.27 and an income elasticity of –0.17</td>
</tr>
<tr>
<td>Showalter and Thurston (1997)</td>
<td>1983–85</td>
<td>US</td>
<td>Physicians’ Practice Costs and Income Survey</td>
<td>Variation in maximum MTR across states</td>
<td>Self-employed physicians have an uncompensated elasticity of 0.3. No effect for employed physicians</td>
</tr>
<tr>
<td>Baltagi et al. (2005)</td>
<td>1993–97</td>
<td>Norway</td>
<td>Admin data</td>
<td>Policy reform as natural experiment</td>
<td>Physician labour supply elasticity 0.3–0.34</td>
</tr>
<tr>
<td>Sæther (2005)</td>
<td>1993–97</td>
<td>Norway</td>
<td>Admin data</td>
<td>Policy reform as natural experiment</td>
<td>Physician labour supply elasticity 0.1–0.2</td>
</tr>
<tr>
<td>Andreassen et al. (2013)</td>
<td>1996–2000</td>
<td>Norway</td>
<td>Admin data</td>
<td>Policy reform as natural experiment</td>
<td>Physician labour supply elasticity 0.5</td>
</tr>
<tr>
<td>Clemens and Gottlieb (2014)</td>
<td>1997</td>
<td>US</td>
<td>Medicare</td>
<td>Reform that led to area-specific price shocks</td>
<td>A 2% increase in Medicare payment rates to physicians led to 3% increase in Medicare care provision</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Paper</th>
<th>Years</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shearer et al.(2018)</td>
<td>1996–2002</td>
<td>Quebec</td>
<td>RAMQ/CMQ time-use survey</td>
<td>Policy reform - increase in clinical services prices</td>
<td>Increase in price of all services led to physicians working less hours</td>
</tr>
<tr>
<td>Brekke et al. (2017)</td>
<td>2006–11</td>
<td>Norway</td>
<td>Admin data</td>
<td>Diff-in-diff (Increased payments after specialisation in general medicine)</td>
<td>Increase in prices paid to primary-care physicians led to increase in number of patients seen, reduction in time per patient – total hours unchanged</td>
</tr>
<tr>
<td>Chan (2018a)</td>
<td>Unknown</td>
<td>US</td>
<td>Academic hospital admission/rostering data</td>
<td>Random assignment of patients to physicians</td>
<td>At end of shift, physicians reduce effort by accepting fewer patients and spending less time with patients they are assigned</td>
</tr>
</tbody>
</table>

*Note:* RAMQ/CQQ, Regie de l’assurance maladie Quebec/College of Physicians; AMA, American Medical Association.
do not respond to differences in the post-tax wage. For Norway, recent papers exploit changes in national wage settlements. The authors find that physicians respond similarly to their US counterparts, with small short-run estimated wage elasticities ranging from 0.18 to 0.50.

Other papers examine the effect of an increase in payments to physicians on hours and/or care provided. For the US, Clemens and Gottlieb (2014) exploit a 1997 Medicare geographic boundary re-adjustment, which led to area-specific price shocks in payments for physicians. They find that a 2 per cent increase in payment rates to physicians leads to a 3 per cent increase in care provision. However, they only study how non-Medicare provision changes. As physicians have both Medicare and non-Medicare patients, it seems highly likely that they substitute away from non-Medicare care when the relative financial benefit falls, making the effect on overall labour supply ambiguous. In support of this, Shearer, Somé and Fortin (2018) find that when a policy reform in Canada increased the price of all clinical services, physicians supplied less labour. Brekke et al. (2017) exploit a reform in Norway that led to some physicians being paid a higher fee per patient than others. They find that physicians did not change the total amount of hours supplied but did spend less time with each patient, so they could see more patients over the same period of time, thereby increasing their remuneration within the total hours envelope.

Chan (2018a) focuses on the extent to which physicians change treatment decisions to preserve leisure. Exploiting the random allocation of patients to physicians on duty in hospitals, he examines how physician behaviour changes as they approach the end of their shift. He finds that towards the end of their shift physicians reduce effort by accepting fewer patients and spend less time with the patients they are assigned. This behaviour results in higher overall costs. By calibrating a discrete choice dynamic programming model, he estimates that physicians are willing to spend hospital resources more than six times their market wage in order to preserve their leisure.

These (and other) papers suggest that in the short run physicians do not alter their labour supply much in response to changes in wages. However, there is considerable evidence that physicians change what they do – and often quite rapidly – in response to changes in incentives that are brought about by changes in how, and what, they are paid for.
b) Physician response to contract changes

Table 2 provides a summary of papers looking at physician responses to contract changes. There is a large body of literature on how the type of contract affects the incentives to provide treatment. Multiple studies, both in the real world and in the laboratory, have found that physicians paid on a per-service basis prescribe more treatments than those paid per-person. Iizuka (2012) finds that physicians in vertically integrated clinics prescribe more branded drugs (in order to capture mark-up) than other physicians. Similarly, Chen, Gertler and Yang (2016) exploit a policy reform in Taiwan that restricted physician ownership of pharmacies. They find that physicians who had to give up ownership of their pharmacy prescribed 7 per cent less drugs.

These papers all provide evidence that physicians are, at least in part, motivated by financial incentives that lead them to alter the treatments they provide to patients when they have the ability to do so. However, the extent to which physicians change these treatment decisions might in part be tempered by concerns about providing treatment that delivers worse outcomes for patients (or that these decisions may be detected by peers or regulators). The literature typically calls this physician ‘altruism’. However, despite its considerable theoretical importance, little is known empirically about physician altruism. There are some studies that try to estimate altruism in the real world; however, they struggle to truly identify the importance of altruism with regards to other factors. Most recent developments have taken place in the laboratory setting. Godager and Wiesen (2013) run an experiment showing that altruism is a key feature of physician decision-making. They also find that there is substantial heterogeneity; some physicians are far more altruistic than others. Kesternich, Schumacher and Winter (2015) examine the importance of professional norms in determining treatment decisions. They find that physicians, who are primed pre-experiment by answering questions about the Hippocratic Oath, perform more altruistically than physicians who are not primed. In a similar fashion, Godager, Hennig-Schmidt and Iversen (2016) find that physicians whose treatment decisions are disclosed to their peers perform more altruistically.

---

33 Only academic empirical analyses are included. Meta-analyses and policy reports are only discussed in the text.
34 McGuire (2000) provides an overview of the early literature. Some of this examines physician induced demand (PID), which is the ability of physicians to exploit information asymmetries with patients in order to provide more treatment (and therefore charge more for their services) than they would be able to in the case where patients are fully informed about optimal treatment choices. This literature concludes that the observed treatment patterns suggest that PID is prevalent. Johnson and Rehavi (2016) provide more recent evidence by looking at the decisions of patients who are themselves physicians (and are therefore more knowledgeable about optimal treatment decisions than the average patient) when they are giving birth. They find that the C-section rate is 10 per cent lower when physicians are giving birth, with only a quarter of the effect attributable to sorting across hospitals and obstetricians.
### TABLE 2

*Physician labour supply (contract changes)*

<table>
<thead>
<tr>
<th>Paper</th>
<th>Years</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kantarevic et al. (2011)</td>
<td>1992–2008</td>
<td>Canada</td>
<td>Admin prescribing records</td>
<td>Policy reform</td>
<td>Family Health Groups increase physician productivity (relative to FFS). No effect on quality</td>
</tr>
<tr>
<td>Hennig-Schmidt et al. (2011)</td>
<td>N/A</td>
<td>Lab</td>
<td>Experiment</td>
<td>Random assignment</td>
<td>Medical students provide more services under fee-for-service than capitation; high severity patients receive larger health benefits under fee-for-service, low severity under capitation</td>
</tr>
<tr>
<td>Brosig-Koch et al. (2016)</td>
<td>N/A</td>
<td>Lab</td>
<td>Experiment</td>
<td>Random assignment</td>
<td>Physicians respond less to payment schemes than medical and non-medical students</td>
</tr>
<tr>
<td>Iizuka (2012)</td>
<td>2003–05</td>
<td>Japan</td>
<td>Admin prescription data</td>
<td>Reduced form</td>
<td>Doctors in vertically integrated clinics prescribe more branded drugs</td>
</tr>
<tr>
<td>Chen et al. (2016)</td>
<td>1997–2000</td>
<td>Taiwan</td>
<td>Admin prescription data</td>
<td>Policy reform</td>
<td>Restricting physician ownership of pharmacies led to reduction in drugs prescribed – no effect on health outcomes</td>
</tr>
<tr>
<td>Whynes et al. (1999)</td>
<td>1996</td>
<td>UK</td>
<td>GP postal survey</td>
<td>Stated Preference</td>
<td>All physicians respond to financial incentives; larger effect on more entrepreneurial physicians</td>
</tr>
<tr>
<td>Godager and Wiesen (2013)</td>
<td>N/A</td>
<td>Lab</td>
<td>Experiment</td>
<td>Random assignment</td>
<td>Substantial variation in the degree of physician altruism</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Paper</th>
<th>Years</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kesternich et al. (2015)</td>
<td>N/A</td>
<td>Lab</td>
<td>Experiment</td>
<td>Random assignment</td>
<td>Professional norms restrain physicians’ self-interest and introduce altruism towards patients (although at cost of reduced efficiency)</td>
</tr>
<tr>
<td>Godager et al. (2016)</td>
<td>N/A</td>
<td>Lab</td>
<td>Experiment</td>
<td>Random assignment</td>
<td>Disclosing information about physician performance makes physicians perform closer to the social optimum</td>
</tr>
<tr>
<td>Gravelle et al. (2010)</td>
<td>2005</td>
<td>UK</td>
<td>ISD – GP Contractor Database</td>
<td>Policy reform</td>
<td>Physicians gamed the performance-related pay system to maximise their income</td>
</tr>
</tbody>
</table>
Taken together, this literature therefore suggests that ‘altruism’ plays an important part in the decisions that physicians make, but this term may capture more than just the physician’s concern over their patients’ well-being.

One type of contract that has become increasingly common across the world is performance-related pay (P4P) for physicians. P4P is a popular reform model in many healthcare systems, reflecting policymakers’ concerns that capitation payments and prospective payment per case give incentives to skimp on quality. The introduction of pay-for-performance contracts provides an opportunity to understand what happens when physicians (and hospitals) are given incentives to increase quality. Issues that are pertinent are whether quality does increase, what levels of payments are required to increase quality, whether incentivising a subset of activities leads to effort diversion from other unmeasured activities, and whether ‘gaming’ is prevalent.

The UK is a leader in the use of P4P. While some schemes have been implemented in hospitals, the largest scheme has established performance-related payments for GPs. Known as the Quality and Outcomes Framework (QOF), the scheme was implemented in 2004. It includes targets for 76 clinical indicators covering 10 important medical conditions. GPs are rewarded if the percentage of patients who meet the indicator is above a certain threshold. While there is some evidence that the scheme led to higher quality in the first year, there is considerably less evidence of an ongoing effect.

Poor scheme design is one key reason identified for the increase in quality being only temporary. In the case of the QOF, the baseline level of the indicators was underestimated, so that quality did not have to increase by much for GP practices to meet the thresholds. As a result, in the first year, performance-related payments were £1 billion, £300 million more than expected. This illustrates a more general problem with such schemes: the scheme designers typically have considerably less information than those who are affected by the scheme. Another design flaw is that the thresholds remain static over time. It is likely that improvements in quality would have taken place in the absence of the reform, so that GPs are often getting paid without making an active effort to improve quality.

These flaws have been identified more generally. Van Herck et al. (2010) provide a systematic review of different schemes across a variety of settings. They conclude that static targets are a common design flaw. Another common drawback of performance-related schemes is that they only focus on a subset of activities that physicians undertake. By incentivising particular clinical indicators, it is likely that performance will fall elsewhere. One other common

37 Roland and Guthrie, 2016.
38 Clinical data are automatically extracted from patient records, although GPs can exclude patients for patient dissent, intolerance of medication or clinical inappropriateness.
40 For the QOF, Gillam et al. (2012) find that performance for non-incentivised conditions has become relatively worse since the introduction of the scheme.
outcome from P4P schemes is ‘gaming’ – that is, the use of resources to make it appear that targets are met, whilst in fact they are not. For the QOF, there is some evidence that doctors might have ‘gamed’ the system by categorising patients as excludable from the set that counted towards their targets, so they could receive increased payments without improving quality.\textsuperscript{41} However, this gaming appears to have been relatively limited. In the case of the QOF, the lack of gaming could be attributed to the low levels of the threshold for payments, meaning targets could be met without much effort, either to improve outcomes or to ‘game’.

In general, although the issue is still open to question, there are several reviews that suggest the evidence does not strongly support the use of P4P (at least as currently implemented in a variety of schemes) as a good use of resources in a healthcare context.\textsuperscript{42} Interestingly, the scepticism about P4P schemes contrasts with the current move, in many healthcare systems, to link payments for delivery more tightly to patient outcomes.

c) Wage elasticities: nurses

Table 3 provides a summary of papers looking at responses of nurses to wages. The early literature, summarised in Shields (2004), finds that nurses are not particularly responsive to changes in remuneration, but few of these papers convincingly address concerns over endogeneity of wages.\textsuperscript{43}

More recent work uses instruments or exploits natural experiments to address this issue. For the UK, Frijters, Shields and Price (2007) look at the extensive margin effect of remuneration on 28,000 nurses in the UK, instrumenting for nurses’ wages with pre-nurse educational qualifications. They estimate that a 10 per cent increase in nurses’ wages would reduce the annual exit rate by 0.7 per cent. Elliott et al. (2007) look at the effect of relative wages in the NHS and in the outside (non-NHS) labour market to explain geographic variation in vacancy rates. In the UK, nurses’ wages are set centrally but local geographic variation in the outside option available to nurses means that the attractiveness of working for the public sector as a nurse differs geographically. They find that there is a relationship between wage

\textsuperscript{41}Gravelle, Sutton and Ma, 2010.
\textsuperscript{42}For example, see Mendelson et al. (2017) who distinguish between ambulatory (service provided in the community) and inpatient (in hospital) settings. They conclude there is some evidence that ambulatory processes of care improve with performance-related pay, but no evidence that patient health outcomes improve in either setting. The lack of an effect of schemes could be due to substantial heterogeneity in the design/setting of different schemes, but Markovitz and Ryan (2017) conclude that ‘extant heterogeneity in the effects of P4P does not fundamentally alter current assessments about its effectiveness’.
\textsuperscript{43}For the UK, Phillips (1995) uses a two-stage model of labour force participation to examine factors that influence nurse employment across both the intensive and extensive margin for a sample of 300 nurses. He finds that increasing wages would raise employment at the extensive margin but would have little effect on the number of hours worked. Rice (2005) uses a variety of variables, including qualifications and labour market experience, to instrument for wages. He finds labour supply elasticities ranging from 0.29 to 0.38.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Years</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phillips (1995)</td>
<td>1980</td>
<td>UK</td>
<td>Woman and Employment Survey</td>
<td>Control on observables</td>
<td>Nurses respond more on extensive margin than intensive margin to wage increases</td>
</tr>
<tr>
<td>Rice (2005)</td>
<td>1991–99</td>
<td>UK</td>
<td>BHPS</td>
<td>Several instruments for wages</td>
<td>Nurse labour supply elasticity is around 0.3</td>
</tr>
<tr>
<td>Frijters et al. (2007)</td>
<td>1997–2001</td>
<td>UK</td>
<td>QLFS</td>
<td>Instrument for wage with education</td>
<td>A 10% increase in a nurses’ wages reduces annual exit rate by 0.7%</td>
</tr>
<tr>
<td>Elliott et al. (2007)</td>
<td>2000</td>
<td>UK</td>
<td>QLFS</td>
<td>Regional variation in wages</td>
<td>Health authorities where NHS nurse wages are relatively higher have lower vacancy rates</td>
</tr>
<tr>
<td>Crawford et al. (2015)</td>
<td>1997–2012</td>
<td>UK</td>
<td>ASHE</td>
<td>Within region change in relative wages</td>
<td>Increasing NHS nurse pay has a large effect in London (where relative wages are lower) and small effects elsewhere</td>
</tr>
<tr>
<td>Staiger et al. (2010)</td>
<td>1990–92</td>
<td>US</td>
<td>Veterans Affairs Personnel and Accounting Integrated Data (PAID)/AHA Nursing Personnel Surveys (NPS)</td>
<td>Area-specific shock to nurses working in Veterans Affairs hospitals</td>
<td>Nurse labour supply is not responsive to wage changes</td>
</tr>
<tr>
<td>Holmás (2002)</td>
<td>1993–97</td>
<td>Norway</td>
<td>Admin data</td>
<td>Policy reform as natural experiment</td>
<td>Increase in wages reduces nurse attrition rate; attrition rate higher in more crowded hospitals</td>
</tr>
<tr>
<td>Askildsen et al. (2003)</td>
<td>1993–98</td>
<td>Norway</td>
<td>Admin data</td>
<td>Policy reform as natural experiment</td>
<td>Nurse labour supply elasticity is 0.21</td>
</tr>
<tr>
<td>Andreassen et al. (2017)</td>
<td>1997–99</td>
<td>Norway</td>
<td>Admin data</td>
<td>Policy reform as natural experiment</td>
<td>Nurse labour supply is not responsive to changes in wages</td>
</tr>
</tbody>
</table>

Note: BHPS, British Household Panel Survey; QLFS, Quarterly Labour Force Survey; ASHE, Annual Survey of Hours and Earnings; AHA, American Hospital Association.
differentials and vacancy rates. Health authorities where NHS nurse wages are relatively higher have lower vacancy rates. In a similar vein, Crawford, Disney and Emmerson (2015) look at the effect of increasing central pay on regional nurse employment. They find that increasing public-sector nurse pay has a large effect in London (where relative wages are lower) and small effects elsewhere in the country, where there are fewer outside options.

For the US, Staiger, Spetz and Phibbs (2010) exploit a 1991 policy reform that led to area-specific shocks in the wages of nurses working in Veterans Affairs hospitals in the US. They find the nurse wage elasticity to be highly inelastic. There are several papers that make use of Norwegian policy reforms, which increased nurse wages, to estimate nurse labour supply elasticities. They all find that nursing labour supply does not respond much to changes in remuneration in the short run.

These papers all examine wages, but a potentially important factor in understanding nurse labour supply is the conditions of work and, in particular, the use of shifts. Nurses may be more responsive to changes in this (and other) features of employment.

2. Entry, exit and geographical location

Changes in medical labour market regulation have been used to examine a number of issues relating to entry, exit and geographical location. The latter is of interest because governments are generally concerned about the spatial distribution of medical care and, hence, of medical labour. Table 4 provides a summary of recent papers that estimate the effect of changes in regulation on entry, exit and location.

Chatterji, Li and Marschke (2018) examine whether malpractice financial liability affects the geographical location of newly qualified physicians. They exploit the staggered introduction of caps on malpractice damages across US states. They find that states that introduce caps on damages have more physicians starting their careers there. Peterson, Pandya and Leblang (2014) also find that states that introduce more stringent legislation have less physician migration. Lieber (2014) uses a more sophisticated identification strategy, which exploits the fact that regulatory changes in other states are more exogenous than those in one’s own state, which could be influenced by insider power. He finds that when a neighbouring state places caps on non-economic


45For example, Mas and Pallais (2017) find that call centre workers are willing to give up 20 per cent of their salary to avoid a schedule set by an employer on short notice and 8 per cent for the option to work from home. This suggests that non-pay levers might be important in boosting nurse labour supply.

46For an extensive overview of earlier research on this topic (including speciality choice, which has not received much recent interest), see Nicholson and Propper (2011).
## Table 4

**Entry, exit and geographical location**

<table>
<thead>
<tr>
<th>Paper</th>
<th>Years</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lieber (2014)</td>
<td>1993–2004</td>
<td>US</td>
<td>Area Resource File (Physician)</td>
<td>Policy reform (diff-in-diff across states)</td>
<td>When a neighbouring state places caps on non-economic damages there is a 4% fall in physician labour supply and a 4% fall in malpractice claims</td>
</tr>
<tr>
<td>Peterson et al. (2014)</td>
<td>1973–2010</td>
<td>US</td>
<td>AMA (Regulation)/Department of Human Services (Physician)</td>
<td>Policy reform (diff-in-diff across states)</td>
<td>States with more stringent licensing laws for foreign physicians have less physician immigration; these states’ licensing boards are more self-financed than others</td>
</tr>
<tr>
<td>Agarwal (2017)</td>
<td>2010–11</td>
<td>US</td>
<td>National Graduate Medical Education census</td>
<td>Matching model (IV for unobserved programme quality with Medicare’s reimbursement rates)</td>
<td>Changing number of places and remuneration leads to large re-sorting effects</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Paper</th>
<th>Years</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kleiner et al. (2016)</td>
<td>1999–2010</td>
<td>US</td>
<td>Nurse Practitioner (Regulation)/ACS (Labour Supply)/Insurance (Prices)</td>
<td>Policy reform (diff-in-diff across states)</td>
<td>States with stronger regulation on nurse prescription have higher prices/lower nurse hours/higher physician hours - no effect on mortality/malpractice</td>
</tr>
</tbody>
</table>

*Note: ACS, American Community Survey; NCBSN, National Council of State Boards of Nursing.*
damages, there is a 4 per cent fall in physician supply as well as a 4 per cent reduction in malpractice claims.\textsuperscript{47}

The long accreditation process for many professions means that the domestic supply of labour stays fairly stable over time. Cortés and Pan (2014) estimate the long-run effect of an increase in the stock of nurses. They exploit variation in the migrant nurse labour supply, instrumenting for the number of migrant nurses with the historical geographical distribution of migrants. They find that an increase in foreign nurses leads to native nurses leaving the profession. Interestingly, this does not appear to be driven by price effects but rather a reduction in the quality of the workplace environment. This illustrates the importance of factors other than pay, such as workplace composition, affecting behaviour. DePasquale and Stange (2016) look at the effect of reducing regulation on the accreditation of foreign workers. They exploit the staggered adoption of the Nurse Licensure Compact (NLC), which made it easier for qualified nurses to practise in a different state to the one in which they qualified. They find no evidence that the labour supply of nurses changes after a state adopts the NLC. Kleiner et al. (2016) look at regulation that introduces greater restrictions on nurse prescribing, as this could influence the division of labour between medical professions, and therefore labour supply more generally. They find that introducing greater restrictions on nurses prescribing drugs leads to nurses working fewer hours while physicians work more. There is no effect on the extensive margin (i.e. the numbers of nurses and physicians stay the same).

A key issue with any identification strategy that exploits changes in regulation is whether the regulatory change is exogenous.\textsuperscript{48} Additionally, while it is also clear that regulation affects labour supply, the net welfare effect is not clear, as many of the studies of the effect of wage or regulation changes on labour supply do not consider the effect on patient outcomes. Linking regulatory changes to patient outcomes\textsuperscript{49} would shed light on whether regulation is correcting market failures or is simply a way of incumbents extracting more rent.\textsuperscript{50}

Agarwal (2017) examines the effect of using quantity- and price-based solutions to incentivise physicians to train in rural areas. To do so, he creates a matching model for the centralised US medical residency (trainee physician)

\textsuperscript{47}He interprets this as evidence that ‘malpractice prone’ physicians sort into lenient states.

\textsuperscript{48}For example, Peterson et al. (2014) find that US states’ licensing boards that were more self-financing were more likely to introduce more stringent legislation. This is consistent with a model where incumbent physicians with more bargaining power introduce more legislation to reduce competition in their labour market, in order to extract higher rents.

\textsuperscript{49}Propper and Van Reenen, 2010.

\textsuperscript{50}Kleiner (2016) provides a more extensive discussion of the trade-off between consumer protection and rent extraction.
allocation mechanism.\footnote{A key issue with identifying the parameters of the model is the potential endogeneity between salaries and unobserved programme characteristics. Agarwal uses Medicare’s reimbursement rates to competitor residency programmes (based on 1985 regulations) as an instrument for unobserved programme quality.} He finds that increasing remuneration in rural areas leads to no effect on the number of doctors training in rural locations. However, it has a large effect on the quality of doctors. There is a re-sorting effect, whereby some high-quality physicians, who with equal wages would rather train in a city, are incentivised to train in a rural area. For quantity regulations, the effect depends on whether the number of urban positions is reduced or the number of rural positions is increased. Reducing the number of urban positions leads to a small increase in the number of rural trainees with no effect on quality. Increasing the number of rural positions leads to a large rise in rural trainees as well as an increase in quality (due to re-sorting).

In addition to choices about location, medical staff must make decisions at different stages in their career about the specialty that they wish to focus on. There has been no recent work on this, potentially because of a lack of detailed data on the career decisions of physicians and the exogenous shocks that could influence specialty choice.

In summary, the consensus view is that there is only a small effect on the total amount of labour supplied in the short run in response to wage changes. For physicians, who are relatively well paid, it is perhaps unsurprising that there is a large income effect, which can cancel out the substitution effect. But these small wage and income elasticities are somewhat in contrast to the (well-documented) often large and rapid changes that physicians make in their behaviour in response to marginal changes in payment schedules. Explanations for these differences are likely to lie in labour market frictions that limit changes in total hours for employed physicians and nurses, the importance of conditions other than wages (e.g. shift work), and the long training periods for medical staff, particularly physicians. All of these constraints contrast with the high levels of discretion physicians have over treatment patterns, which make altering treatments in response to changes in relative prices comparatively easy. Thus, changing treatment patterns is an easier way for physicians to increase leisure, limit hours and/or increase their income.

\section*{IV. Labour and the production of healthcare}

The extensive variation observed in productivity in the healthcare sector that we noted in the introduction has driven interest in the role played by medical labour in driving this variation.\footnote{This can be thought of as an interest in the role of medical labour in the healthcare production function. In the production function, medical labour (physicians, nurses), capital (technology) and other inputs (drugs, etc.) are combined to produce healthcare.}
1. Physicians

Recent research has focused on the role played by ‘physician style’, the association of style with patient outcomes, and the determinants of style. The aspects of style or the contributors to variation in style include learning-by-doing, willingness to experiment and clinical environment (the latter including peer effects). Table 5 provides an overview of these papers.

a) Physician style

Epstein and Nicholson (2009) seek to understand why the variation in caesarean section (C-section) rates is twice as large within the US market as the variation between markets. Using hospital discharge and physician characteristics data from Florida and New York, they examine how characteristics such as local area, residency training, experience, gender and race affect treatment styles. They find that observable physician characteristics have very little explanatory power. However, time-invariant unobservable factors at the physician level (which they argue can be thought of as ‘style’) account for 30 per cent of the variation in risk-adjusted C-section rates.

Currie, MacLeod and Van Parys (2016) use hospital discharge data from Florida to examine what factors influence physician treatment decisions for patients who have had a heart attack. They define style in terms of willingness to undertake more aggressive invasive procedures. In their context, these more invasive procedures benefit all patients but also incur higher costs, so they should only be used on the most appropriate patients. They link treatment patterns to characteristics of the physicians and find that physicians who are young, male and trained in one of the top 20 schools are not only more likely to carry out invasive procedures but are also better at selecting the most appropriate patients for these procedures.

Frank and Zeckhauser (2007) conceptualise the framework that physicians use when making clinical decisions, drawing a distinction between rational individually tailored treatment decisions and ready-to-wear decisions based on norms. They highlight the fact that there are substantial costs (communication, cognition, coordination and capability) incurred when a hyper-rational decision-making process is used. In the context of drug prescribing, therefore, it might be optimal to experiment with new drugs (i.e. non-tailored treatments) for some patients (e.g. the sicker chronic patients) and stick to norms for other patients (e.g. those who are acute patients, and so are ill only once). Using data on the prescribing habits of US primary care physicians, they find that norm-based decisions are fairly common. Often this behaviour appears optimal, although some physicians display suboptimal idiosyncratic behaviour, whereby they continue to prescribe certain drugs, even though a large body of evidence suggests there is a superior alternative.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Year</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currie et al. (2016)</td>
<td>1992–2014</td>
<td>Florida</td>
<td>Admin discharge data (patient).State medical licence database (physician)</td>
<td>Random assignment of patient to physician</td>
<td>Young, male, top 20 trained physicians have more ‘aggressive’ treatment style; they are also better at selecting the most appropriate patients for these costly procedures</td>
</tr>
<tr>
<td>Franck and Zeckhauser</td>
<td>1999–2003</td>
<td>US</td>
<td>NAMCS (patient/physician).Quality Improvement for Depression (patient)</td>
<td>Random assignment of patient to physician</td>
<td>Ready-to-wear treatments are often a sensible response to complex decision-making environment</td>
</tr>
<tr>
<td>Currie and MacLeod</td>
<td>2013–16</td>
<td>US</td>
<td>Insurer (patient).IQVIA (physician)</td>
<td>Random assignment of patient to physician</td>
<td>Optimal treatment depends on physician quality; high-skilled patients who experiment more (within guidelines) achieve better patient outcomes</td>
</tr>
<tr>
<td>Cutler et al. (2019)</td>
<td>2007–10</td>
<td>US</td>
<td>AMA Physician survey/Medicare patient survey/Dartmouth Atlas of Medical Care</td>
<td>Vignette</td>
<td>Physician style important in explaining observed variation in survey response; 12% of Medicare spending explained by physician beliefs that cannot be justified by patient preferences of evidence of clinical effectiveness</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Paper</th>
<th>Year</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baicker et al. (2007)</td>
<td>1993–2001</td>
<td>US</td>
<td>National Practitioner Data Bank/Medical Liability Monitor (malpractice costs)/Medicare (services)</td>
<td>Control on observables</td>
<td>A rise in malpractice premiums (at state level) leads to increased use of diagnostic services; no effect on more invasive procedures</td>
</tr>
<tr>
<td>Shurtz (2013)</td>
<td>1992–2008</td>
<td>Florida</td>
<td>Admin discharge data (patient). Practitioner Profile/Medical Profession Liability (physician)</td>
<td>Random assignment of patient to physician/event study</td>
<td>C-section rates increased by 4% after a medical error and continue to increase afterwards</td>
</tr>
<tr>
<td>Molitor (2018)</td>
<td>1998–2012</td>
<td>US</td>
<td>Medicare</td>
<td>Random assignment of patient to physician/exogenous physician migration</td>
<td>Environmental factors explain between 60 and 80% of regional disparities in physician behaviour</td>
</tr>
<tr>
<td>Chan (2018b)</td>
<td>Unknown</td>
<td>US</td>
<td>Academic hospital admission/rostering data</td>
<td>Random assignment of trainees to supervisory physicians</td>
<td>Trainee doctors play an oversized role (relative to expertise) in decision-making process in order to learn more effectively and improve future decisions</td>
</tr>
<tr>
<td>Silver (2016)</td>
<td>2005–13</td>
<td>New York</td>
<td>Admin discharge data</td>
<td>Random assignment of physicians to different shifts</td>
<td>Physician peer effects have a variance one-quarter to one-third as large as physician fixed effects within a hospital</td>
</tr>
</tbody>
</table>
Currie and MacLeod (2018) put forward a theoretical framework that captures both rational and norm-based decision-making. In their model, physicians differ both in the extent they like to experiment (measured by the concentration of different drugs they prescribe) as well as in their overall level of skill (higher-skilled physicians learn quicker about the efficacy of new drugs). In their model, experimenting leads to a short-run cost to patients as new drugs are tested on them. High-skilled physicians are able to use this information to quickly update their prior beliefs, leading to a long-run benefit to patients, whereas lesser-skilled physicians over-prescribe ineffective drugs as their beliefs do not update quickly enough. Empirically, they find that higher-skilled physicians are more experimental and that receiving treatment from an experimental provider reduces the probability of hospitalisation for a patient. They also find that guidelines place a useful bound on experimentation.

Cutler et al. (2019) use vignettes to determine the treatment decisions that cardiologists and primary-care physicians would make under hypothetical scenarios. They examine which factors are good predictors of hypothetical treatment decisions. They find that although factors such as patient demographics, contract type and clinical setting can predict some of the variation in responses, there is a large amount of variation that depends on individual physician style. Some of this variation is justifiable; however they find that 12 per cent of expenditure is explained by physician beliefs that cannot be justified either by patient preferences or by evidence of clinical effectiveness.

Some recent studies examine how physician style changes when subject to different working environments. This can be in the form of working across different organisations (e.g. hospitals), facing different external (market) environments or working with different colleagues (peer effects).

Molitor (2018) studies the effect of the hospital environment in determining the considerable regional variation in invasive procedures for patients who have had a heart attack. He exploits cardiologist migration between (large) regions in the US to examine if cardiologists change their style in a new clinical setting. Using data on Medicare patients, he finds that environmental factors explain between 60 and 80 per cent of regional disparities in physician behaviour, indicating that physician style adapts significantly to the clinical setting. Put another way, the style physicians bring with them is soon eroded in a new setting.53

Baicker, Fisher and Chandra (2007) examine the effect of malpractice premiums on the type of treatments provided. Exploiting variation in premiums

53One issue not totally addressed by this design is endogeneity of region. While Molitor argues that movement between large regions might be more exogenous than moving employers within a city, migrating physicians might be more adaptable/amenable to new settings than physicians who do not move. This would give an upwards bias to the estimates of the effect of environmental factors.
at state level, they find that increases in malpractice premiums led to increased use of diagnostic services, as physicians became more concerned about litigation from failing to diagnose a patient correctly. There was no effect on the use of more invasive procedures, indicating that the increased diagnostic testing did not reveal more unnoticed conditions or that there were behavioural responses by physicians to avoid higher-risk patients. This study cannot account for the endogeneity of malpractice premiums or for physician movement across states in response to changes in premiums.\textsuperscript{54} Shurtz (2013) examines if individual obstetricians change their style after medical errors that led to malpractice litigation for cases where the claim was ultimately successful. He finds that C-section rates increase 4 per cent immediately after the adverse event and continue to rise over time. This suggests that physicians learn from genuine errors rather than from fear of malpractice claims.\textsuperscript{55}

At the hospital level, recent interest has focused on the role of peers and teams. Chan (2018b) exploits variations in physician rostering. He hypothesises that physicians may change their treatment decisions when working with different colleagues, both those at the same level of experience and those in more supervisory roles. He examines the aggregation of information and subsequent physician decisions among physician teams with different individual levels of knowledge. He finds that the allocation of influence over treatment decisions between senior and junior trainees is proportional to the amount of knowledge they have. However, trainees have a disproportionately high influence, relative to their supervisors, over treatment decisions, based on what might be expected from their respective levels of expertise. Chan interprets this as evidence that trainee physicians are given an oversized role in order to experience the decision-making process, so that they can make decisions more effectively in the future.\textsuperscript{56} Silver (2016) also focuses on peer effects using variation in the other physicians who work during an individual physician’s shift. He finds significant peer effects, which have a variance one-quarter to one-third as large as the physician’s own effects.

### 2. Nurses and healthcare production

The majority of recent papers with a strong identification strategy focus on physician treatment decisions in determining patient outcomes. There has, however, been a long-standing interest in the role of nurses and there is a

\textsuperscript{54}The latter is shown in Lieber (2014).

\textsuperscript{55}This paper addresses the endogeneity issues in Baicker et al. (2007) by using a physician-level study design to isolate the physician response. In this design, the clinical setting is the same before and after the adverse event, whereas changes in state-level premiums are likely to affect more than individual physician style.

\textsuperscript{56}This highlights the importance of a dynamic framework (e.g. Currie and MacLeod, 2018) to understand patterns in treatment decisions.
large body of literature that documents the associations between nurse inputs and patient outcomes. A large set of papers examines the association between nurse numbers and output in the US. For example, Rogowski et al. (2013) find that nurse under-staffing relative to national guidelines is widespread and that hospitals that are understaffed have worse health outcomes. Aiken et al. (2011) find that hospitals with a higher nurse-to-patient ratio have worse outcomes. They also find that, conditional on the number of nurses, hospitals with a better workplace environment (measured by a survey) and more qualified nurses perform better. More recent studies have exploited the growing availability of data to establish whether these associations are causal. Table 6 provides a summary of papers that estimate the importance of nurse inputs in explaining treatment decisions and subsequent outcomes.

Gruber and Kleiner (2012) exploit nursing strikes in New York, finding that strikes led to in-hospital mortality rising by 18.3 per cent and an increase of 5.7 per cent in the 30-day readmission rate. Friedrich and Hackmann (2017) examine the introduction of a new parental leave programme in Denmark, which led to a 10 per cent fall in the nursing labour supply. They find no effect on hospital mortality, although nursing home mortality increases by 13 per cent as a result of the reform. Lin (2014) uses changes in the minimum staffing requirements for nursing homes to instrument for the effect of the number of nurses on quality. He finds that an increase of one standard deviation in the number of nurses leads to a 16 per cent increase in the quality of care provided.

Economic analysis makes clear that alternative opportunities affect the behaviour of labour. Using this intuition, Propper and Van Reenen (2010) examine the effect of the wedge between the regulated wages of nurses in English hospitals and the freely set outside labour market wages for individuals with qualifications similar to those of nurses on hospital production. Using data from all English NHS acute hospitals for the years 1997–2005, they find that hospitals located in areas with higher relative outside wages perform less well. In these hospitals, patients admitted following a heart attack are more likely to die. The study exploits short-run changes in the relative wages paid to NHS nurses and so suggests that nurse labour supply, in the short run, is responsive to outside opportunities. They also find that, on balance, regulated wages for nurses have an overall negative effect for patients admitted following a heart attack. Thus, while wage regulation may allow governments to control overall upwards pressure on wages, it is a blunt instrument and may have unintended consequences.

One limitation of these papers is that they do not have the data to examine the pathways by which reductions in nurse labour supply affect patient outcomes. A reduction in the nursing labour supply could lead to hospitals changing their behaviour in a variety of ways (e.g. by employing contract nurses, calling in more doctors, rationing or postponing treatment). Bartel et al. (2014) examine one aspect of this: the effect of nurse quality in the context of team
<table>
<thead>
<tr>
<th>Paper</th>
<th>Year</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gruber and Kleiner</td>
<td>1984–2004</td>
<td>New York</td>
<td>Admin discharge data (patient)/ FCMS (strikes)</td>
<td>Use strikes as exogenous variation in nurses</td>
<td>Nursing strikes led to an increase in in-hospital mortality by 18.3% and 30-day readmission by 5.7%</td>
</tr>
<tr>
<td>(2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10% reduction in nursing labour supply led to no effect on hospital mortality; 13% increase in nursing home mortality</td>
</tr>
<tr>
<td>Lin (2014)</td>
<td>1999–2003</td>
<td>US</td>
<td>Admin nursing home data (OSCAR)</td>
<td>Use reform to minimum staffing requirements as exogenous variation in nurses</td>
<td>One standard deviation increase in nurses leads to an increase in quality (reduction in deficiencies) of 16%</td>
</tr>
<tr>
<td>Bartel et al. (2014)</td>
<td>2002–06</td>
<td>US</td>
<td>Veterans Affairs admission data (PTF)/Staff record (PAID)</td>
<td>Use sick leave/staff turnover as exogenous variation in team-specific human capital</td>
<td>Disruptions to team functioning (short and long term) led to significant decreases in productivity</td>
</tr>
</tbody>
</table>
production. They use ward-level data from Veterans Affairs hospitals in the US, which contain detailed information on nurse rostering, to isolate the effect of changes in nurse quality on team output. They focus on short-run disruptions (sickness) and long-run disruptions (staff turnover), some of which they argue are plausibly exogenous. They find that higher nurse quality (as measured by education and ward-specific experience) is associated with shorter hospital stays (the only output measure they have in the data). They also find that, conditional on nurse quality, higher team-specific human capital results in shorter lengths of stay.

3. Management quality and healthcare production

Another factor that is important in any production process is management quality. Management quality can be thought of either as an additional input or as influencing the productivity of other inputs (i.e. similar to a technology that affects how much output is generated for a given set of inputs). Recent research has shown that management quality improves productivity in the manufacturing sector of the economy57 and there is growing evidence that it also improves productivity in public-sector organisations.58

The idea that management and organisational practices can drive variation in quality and efficiency across hospitals is not new in the healthcare literature. In-depth qualitative studies59 document that hospitals with lower heart attack (AMI) mortality rates tend to have clear and well-communicated goals throughout the organisation, make systematic use of problem-solving tools (such as ‘root cause analysis’), have greater reliance on data as well as stronger communication and coordination routines relative to low-performing hospitals. Case-study evidence on hospitals such as Virginia Mason60 ThedaCare61 and Intermountain62 suggest that managerial and organisational processes can contribute to better clinical performance by focusing attention and resources towards the issue of the quality of care.

The NHS has undertaken a plethora of reforms, beginning in the 1980s, but increasing in pace in the 1990s and 2000s, that devolve responsibility for care delivery to the local level, and with this bringing a greater focus on the role for local management.63 But there has traditionally been limited empirical evidence in large samples on the relevance of management and organisational practices for hospitals, because of the inherent difficulty in measuring different

57 Bloom and Van Reenen, 2007.
58 Janke, Propper and Sadun, 2019.
60 Kenney, 2015.
61 Toussaint, Conway and Shortell, 2016.
63 Janke et al., 2019.
inputs, and the inability to control simultaneously for hospital and patient population characteristics.

One recent approach is to employ the World Management Survey (WMS), which has been used to show that managerial practices matter in a variety of sectors outside healthcare, to examine the effect of management practices on healthcare production. Table 7 provides a summary of these papers.

Bloom et al. (2017) show the distribution of the WMS management score in approximately 2,000 hospitals across nine countries. The data show large variations in the adoption of managerial practices, both across and within countries. Crucially, this variation is systematically associated with clinical outcomes and the magnitudes of the correlations are non-trivial. They find that a change of one standard deviation in the management score (which corresponds to an increase of one point on the 1 to 5 scale of the WMS score) is associated with a decline of 18 per cent of a standard deviation in AMI deaths in the US and 29 per cent of a standard deviation in Brazil. McConnell et al. (2013) document a negative and significant association between management (measured using the WMS survey instrument) and AMI mortality rates in the context of 597 cardiac units in the US. Bloom et al. (2015) look at management practices in NHS hospitals in 2006 and find a positive association between higher management quality and hospital performance in terms of survival rates from general surgery, lower staff turnover, lower waiting lists, shorter lengths of stay and lower infection rates.

Examination of the variation in management and organisational practices across providers has indicated the importance of three types of factors associated with this variation: (a) the characteristics of the environment faced by hospitals (e.g. differences in competition and regulatory context); (b) the characteristics of the hospitals (e.g. size, ownership, network affiliation, teaching status); (c) the characteristics of the managers working at the hospital (e.g. their tenure and managerial training).

Bloom et al. (2015) exploits variation in hospital competition brought about by the reluctance of politicians to close hospitals in marginal political constituencies. Using this, they show that higher managerial quality is causally associated with hospitals operating in more competitive markets. In other words, competition between hospitals increases managerial quality, which is associated with better performance. Other work has looked at associations between management quality and hospital type. Tsai et al. (2015) find that a variety of characteristics are associated with management quality across 2,000 hospitals in nine countries. Teaching hospitals, hospitals with boards that care

64Bloom and Van Reenen, 2007.
65See https://worldmanagementsurvey.org/.
66The UK has been at the forefront of policies to introduce competition into a highly regulated and centralised healthcare system. As many of these reforms target organisations (primarily hospitals) rather than individuals, these reforms are not discussed here. Propper (2018) provides a review.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Year</th>
<th>Location</th>
<th>Data</th>
<th>Identification</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bloom et al. (2017)</td>
<td>2006–13</td>
<td>Nine countries</td>
<td>WMS/Regulator quality measures</td>
<td>Control on observables</td>
<td>Higher management quality associated with substantially better clinical outcomes both across and within countries</td>
</tr>
<tr>
<td>Bloom et al. (2015)</td>
<td>1997–2005</td>
<td>US</td>
<td>WMS/Regulator quality measures</td>
<td>Use political marginality as instrument for amount of competing hospitals</td>
<td>Adding rival hospital increases management quality by 0.4 standard deviation; increases survival rates from emergency heart attacks by 9.7%</td>
</tr>
<tr>
<td>Tsai et al. (2015)</td>
<td>2009</td>
<td>US/England</td>
<td>WMS/Survey</td>
<td>Control on observables</td>
<td>Positive correlation between hospital size/ownership/corporate governance and management</td>
</tr>
</tbody>
</table>

Note: WMS, World Management Survey.
more about quality and hospitals that use performance-related pay for senior management roles all had, on average, higher management quality. Across the same sample of 2,000 hospitals, Bloom et al. (2017) show an association between management quality and proximity to a university that supplies joint business and clinical education.

V. Conclusion

In the face of rising demand for healthcare, determining the size and shape of the medical workforce is a key priority for policymakers across the world. Better understanding of how to recruit and retain medical staff, and of the effects of individual and team productivity on patient outcomes, is key to providing high-quality care to the population now and in the future. In this paper, we have reviewed studies of medical labour supply and its role in the production of healthcare. While it has not been possible to survey all work in a vast area, we have summarised recent advances that examine the labour supply decisions of medical personnel, and we have attempted to explain how and why productivity varies across these personnel.

Much of the literature is from the US and while many of the issues relating to the medical workforce are similar across healthcare systems, there are some important areas where future research could fill gaps in knowledge that are particularly pertinent for the UK. First, future research should try to identify and quantify non-pay factors that are related to medical labour supply. Recruiting and retaining a medical workforce of sufficient size to meet increasing demand for healthcare is, and will continue to be, a huge policy issue in the UK. The existing literature, which is largely US focused, suggests that physician labour supply is relatively insensitive to wages. An even smaller evidence base suggests that a similar story is true for nurses. This raises the important question of which other factors have large impacts on the labour supply of medical workers. For example, could labour supply be increased by improvements in wider working conditions, or by reforming shift work to make working more flexible? Answering these questions will be important in recruiting and retaining a larger workforce in future.

Second, it is crucial to understand how to incentivise training in particular specialties, and in ensuring geographical balance in the ability to hire doctors and nurses. While demand for most forms of healthcare is increasing in the UK, demand for care (and the associated labour demand) will be rising faster in certain medical specialties. For example, the rapid ageing of the population means that a large rise in the number of doctors specialising in geriatric medicine and in general practice will be required over the next 20 years to fully meet demand. However, these are areas that face real issues in recruiting sufficient doctors, with the numbers of GPs per head actually falling over the past five years. The existing evidence on how and why doctors choose to work
in particular specialties is very scarce, as a result of both little exogenous variation in choice and a lack of available data on the career choices of doctors. Similarly, a better understanding of why individuals choose to locate in particular areas is needed to ensure that there are no regional shortages of doctors and nurses. Understanding these choices should be a research and policy priority.

Third, the composition of the medical labour force has changed over the past 20 years and is likely to continue to change. This is true in terms of both the mix of the different types of staff who work together – with a shift towards more doctors relative to nurses, and substitution of healthcare and nursing assistants for traditional roles such as doctors and nurses – and the qualifications required of medical personnel. For example, educational requirements for nurses have become more stringent over time and, coupled with reforms to the profession associated with the introduction of the ‘Agenda for Change’, this might have positive knock-on effects for the quality of nursing. It is particularly important to study the effect of these changes on the labour force composition – and the effect on patient outcomes – in light of recent US-based work on the productivity and styles of labour that highlights the importance of labour force quality in the production of healthcare. The UK institutional setting – without the complicated healthcare networks and contracts present in the US – provides a promising setting to better understand these issues. While UK study to date has been limited by the lack of administrative data linking data on personnel to patient outcomes, such data are becoming available. Exploiting these data would allow researchers to examine how different types of workers interact and complement or substitute for one another, and how this can be used to improve patient outcomes.

References

Medical labour supply and the production of healthcare


