MONASH University

MONASH BUSINESS SCHOOL

CENTRE FOR HEALTH ECONOMICS

WORKING PAPERS

Nursing Shortages and Patient Outcomes

Discussion Paper no. 2024-02

Elaine Kelly, Carol Propper and Ben Zaranko

Keywords: Nursing Shortages, Patient Outcomes, Mortality, Hospital Care

JEL Classification: I11, J24, J45, M50

Elaine Kelly: Health Foundation, Institute for Fiscal Studies (email: elaine.kelly@health.org.uk); Carol Propper: Imperial College Business School, Monash University, Institute for Fiscal Studies (email: c.propper@imperial.ac.uk); Ben Zaranko: Institute for Fiscal Studies (email: ben.zaranko@ifs.org.uk).

Nursing Shortages and Patient Outcomes

Elaine Kelly, Carol Propper and Ben Zaranko*

February 2024

Abstract

This paper examines the effect of nurse shortages on healthcare production. Employing novel high-frequency data we examine what effect the absence of nursing staff has on inpatient mortality and other outcomes associated with nursing care. We find significant adverse mortality impacts of shortages of nurses with degree-level qualifications but no effect of shortages of less qualified nursing assistants. Adverse mortality impacts of shortages are particularly concentrated among patients with sepsis, a condition where early detection is important for survival and where nurses have a central role in detection and subsequent control.

JEL classification: I11, J24, J45, M50.

*The research was enabled by the Imperial Clinical Analytics Research and Evaluation (iCARE) environment and used the iCARE team and data resources. Data management was provided by the Big Data and Analytical Unit (BDAU) at the Institute of Global Health Innovation (IGHI). This paper has been screened to ensure no confidential information is revealed. All authors gratefully acknowledge financial support from the UK Economic and Social Research Council (ESRC) through the Centre for the Microeconomic Analysis of Public Policy (CPP) at the Institute for Fiscal Studies (ES/M010147/1) and "The impact of medical labour on variation in patient outcomes: evidence from English public hospitals" research grant (ES/S003118/1). Propper also acknowledges funding from ERC Advanced Investigator Grant Healthcare Labour. We would like to thank Eric French, Imran Rasul, Max Warner, and seminar participants at the Institute for Fiscal Studies, University of York, Monash University, University of Sydney, Erasmus University Rotterdam, Essen Health Conference 2022, European Association of Labour Economists 2022 Conference, European Health Economics Association 2022 Conference, the Royal Economic Society conference 2022, and the Allied Social Science Associations 2023 meeting for comments. Any errors are the authors' own. Author affiliations and contacts: Kelly (Health Foundation and Institute for Fiscal Studies, elaine.kelly@health.org.uk); Propper (Imperial College Business School, Monash University and Institute for Fiscal Studies, c.propper@imperial.ac.uk, corresponding author); Zaranko (Institute for Fiscal Studies, ben.zaranko@ifs.org.uk).

1 Introduction

Healthcare professionals comprise more than 10 percent of the total workforce in many developed countries. Nurses are the largest healthcare profession and understanding their productivity is central to improving the quality and efficiency of the healthcare sector. It is also particularly policy relevant given the worldwide shortage of nurses and intense competition for nurses between countries (Scheffler and Arnold, 2019; WHO, 2020; Boniol et al., 2022). Yet despite their significance for healthcare and ongoing policy debates, there is little robust empirical evidence on the impact of nurses on healthcare delivery.

In this paper we examine the effect of nurses on healthcare output. Our focus is nurses in the hospital setting, where they comprise around 50% of all clinical staff (NHS Digital, 2023). The majority of nurses in hospitals work in teams that provide round the clock care for patients requiring hospitals stays. They play a critical role in the delivery of care, with responsibility for monitoring patients, implementing and adjusting treatment plans, and the co-ordination of overall care. Our primary measure of output is inpatient mortality, a stark but frequently used measure of the quality of healthcare production, and we complement this by looking at other patient outcomes which are plausibly related to nurses' actions.[1]

Our analysis examines the impact of the absence of nursing staff on patient outcomes. We identify this by exploiting daily shortfalls from target staffing levels which are determined several months in advance with the aim of assuring patient safety. Additional identification comes from the rotation of staff at the team level that arises because nursing care needs to be provided 24/7 but each nurse can only work a limited number of hours. These two sources of variation allow us to study the effect of an absence of a more qualified nurse, compared to that of a less qualified nurse, as well as the absence of a nurse who is more familiar with their employer, their physical location or their fellow team members than one who is less familiar. We undertake this analysis at the ward-day level, which is the unit at which staff planning is undertaken.

We abstract away from the impact of financial incentives on productivity, as pay in our setting is determined exogenously to the organisation we study, is unrelated to output, and individuals of the same grade and contract type are paid the same amount regardless of their productivity. It is also a setting in which there is a chronic shortage

[1] Mortality is a commonly used as a proxy for the quality of nursing care (e.g. Needleman et al. (2011); Griffiths et al. (2019); Peutere et al. (2024)) where potential mechanisms include missed care (Ball et al., 2014, 2018), delayed monitoring of vital signs (Redfern et al., 2019), and failure to respond to deterioration in a patient's condition (Smith et al., 2020).

of nursing staff, which has been argued to negatively impact patient safety.[2]

Our data comes from one large hospital in the English National Health Service for the calendar year 2017. It utilises around 19,000 teams defined at the 24 hour-ward level, containing just under 4,500 unique staff, responsible for 44,000 patients and providing 3.5 million hours of patient care. We control for differences in patient severity and differences across hospital sites, wards and days. Our identifying assumption is that staffing shortfalls are uncorrelated with residual determinants of patient outcomes. We undertake a large set of tests to examine this key assumption and find no evidence of such endogeneity. We find no clear relationship between staffing shortfalls and predicted mortality based on patient acuity or measures of observable patient severity, providing evidence that staffing does not respond to day-to-day variation in patient mix. Deaths on previous days have no clear impact on either subsequent staffing levels or subsequent staff sickness absences, indicating that staffing does not respond to adverse events. There is no evidence that low staffing leads to an outflow of patients from the ward, nor that sickness absences lead to such an outflow, suggesting that responses to staffing shocks do not lead to patient reallocation across wards. Finally, we find no serial correlation in patient deaths across days, indicating that our observable measures of patient severity do a good job of controlling for the probability of patient mortality.

We find that both the characteristics of who is missing and the extent of the shortage matters for patient mortality. For the average ward the absence of a nurse with university degree-equivalent level training increases the odds of a patient death by approximately 10%. In contrast, the absence of a nursing assistant, who lacks the formal training of a qualified nurse, has no impact on patient mortality. The effects of a shortfall of qualified nurses are increasing in the hours of the shortfall. We find little evidence for dynamic effects: there is, for example, no evidence of lagged effects of short-staffing or spillovers across teams.

Our findings also indicate returns to firm-specific experience. Conditional on the quantity and skill-mix of individuals on the ward, increasing the average firm-specific experience among degree qualified nurses by one year is associated with an 8.0% reduction in the odds of a patient death, the equivalent to adding three-quarters of an extra qualified nurse to the ward. This is consistent with a role for firm-specific human capital, although there are other potential interpretations, including a correlation between tenure and quality of the match between the nurse and the hospital. Again, there is no evidence of such an experience effect for nursing assistants. There are no significant re-

[2]https://www.rcn.org.uk/news-and-events/news/uk-rcn-survey-reveals-shocking-impact-of-nurse-staffing-crisis-on-patient-safety-060622, accessed 15 Jan 2024

turns to familiarity with the team location or with co-workers for either qualified nurses or nursing assistants.

Our results are driven by mortality impacts on patients of relatively low, rather than high, clinical severity. The largest impacts are for patients with a diagnosis of sepsis. This is a condition which can develop suddenly and where early detection is of critical importance for survival, and for which nurses have a central and direct role in detection of the early signs and its subsequent control (Kleinpell et al., 2013; Schorr, 2016; Kleinpell, 2017; Lasater et al., 2021). This suggests that less intensive patient monitoring and assessment by teams that are lacking degree-qualified nurses is likely to be one key mechanism driving our results.

We extend our analysis to consider impacts of staffing shortfalls on patient falls and transfers from a ward to intensive care. Falls are commonly used in studies of the association between nurse staffing and patient safety (Donaldson et al., 2005; Patrician et al., 2011; Kalisch et al., 2012). Transfers to an intensive care unit serve as one proxy for a severe deterioration in a patient's condition (Smith et al., 2020). We find no impact of shortages on falls. However, our results for intensive care transfers replicate those for mortality. Shortages of degree qualified nurses increase the probability of a transfer to intensive care from the ward, while again there is no impact of a shortage of nursing assistants.

We contribute to three separate strands of research. First, we contribute to the literature on the role of nurses in healthcare production. We build on insights from Friedrich and Hackmann (2021), who exploit the labour supply effects of a parental leave program in Denmark to quantify the impacts of nurse shortages on patient outcomes in hospitals and nursing homes; Gruber and Kleiner (2012), who examine the impacts of nursing strikes; Propper and Van Reenen (2010), who exploit variation in outside wage options for nurses induced by centralised pay regulation to examine the impacts on deaths following heart attacks; and Lin (2014), who exploits legislative changes to minimum staffing requirements in nursing homes. Our data, which is both more granular and higher frequency than that used in most previous studies, provides new evidence on the impacts of both the quantity and quality of nurses on patient outcomes at the level at which staffing decisions take place and is planned. In this literature, our paper is most closely related to Bartel et al. (2014), who examine the impact of month-to-month variation in general and unit(ward)-specific human capital on the productivity of nursing teams in Veteran Administration hospitals in the USA. Proxying nurses' general human capital by education and their unit-specific human capital by experience on the nursing unit, they find that greater amounts of both types of human capital significantly reduces patient length of stay. They also find that disruptions to team functioning (attributable to the departure of experienced nurses, the absorption of new hires, and the inclusion of temporary contract nurses) are associated with significant decreases in productivity beyond those attributable to changes in nurses' skill and experience. We extend this research by focusing on measures of clinical quality. Whilst healthcare systems often focus on reducing length of stay for cost reasons, shorter lengths of stay may not be unambiguously better for patients. We also undertake analysis at a much higher frequency level. Bartel et al. (2014) use data at the month-unit level, which means they are limited in their ability to test for endogeneity with respect to patient severity or to explore dynamics. In contrast, the high frequency of our data enables us to examine changes in teams at the level at which teams are planned (the shift level) and to examine potential sources of endogeneity in the relationships between absences and patient outcomes. The granularity of our data also allows us to go beyond Bartel et al. (2014) to examine whether there are productivity differences *within* qualification group and to differentiate between experience at the hospital, ward and team (co-worker) levels. Finally, in contrast to Bartel et al. (2014), we are able to assess which types of patient are most affected by nurses shortages.

Second, we contribute to the literature that is concerned with worker human capital and productivity. These issues have received growing attention in the healthcare setting (e.g., Doyle et al. (2010); Currie and MacLeod (2017, 2020); Chen (2021); Chan et al. (2022); Chan and Chen (2022)). Our paper also relates to the more general relationship between human capital and productivity (e.g. Moretti (2004); Gennaioli et al. (2013)).

Third, we contribute to the literature on workers who are organised into teams. In healthcare Agha et al. (2021) show that teams composed of a smaller number of specialists have better outcomes and Chen (2021) finds that higher levels of shared work experience in two person doctor teams is associated with reduced patient mortality. In education, Herrmann and Rockoff (2012) show that less experienced teachers had less impact on pupil learning. In sport, Stuart (2017) evaluates the team productivity effects of injuries of different types of hockey players; Fischer et al. (2021) investigate the productivity effects of COVID-19 on the performance of top league soccer players in Germany and Italy; and Hoey et al. (2022) examine the effect of absences of hockey players on team output.[3]

[3]There is an extensive literature on interactions between team members, including peer effects in the workplace (Mas and Moretti, 2009; Guryan et al., 2009; Cornelissen et al., 2017; Silver, 2021), learning from co-workers (Jarosch et al., 2021), credit-sharing within teams (Jones, 2021), productivity spillovers (Arcidiacono et al., 2017) and team incentives (Hamilton et al., 2003; Bandiera et al., 2009, 2013; Burgess et al., 2010, 2017; Friebel et al., 2017). With the exception of the last (as workers are not paid

The paper proceeds as follows. Section 2 describes the institutional context and our data. Section 3 presents descriptive statistics. Section 4 sets out our empirical strategy and considers possible threats to identification. Section 5 presents our mortality results, examining the impact of a shortage of nurses of different levels of general human capital and examining alternative specifications of the treatment variables. Section 6 estimates the impacts on other patient outcomes. Section 7 repeats our analysis at the patient-day level and conducts a range of further robustness tests. Section 8 concludes.

2 Institutional background and data

2.1 Hospital nursing teams in the English NHS

2.1.1 Staffing groups and definitions

Nursing care is delivered and planned on a team basis. Nursing staff work together in a team, defined by a particular location (hospital ward) and time (shift) to treat the patients in that ward. Two broad types of nurse are employed in a nursing team. The first are registered nurses (RNs) who have completed formal training and hold a university diploma or degree-level qualification and are registered with the body which certifies medical professionals. The second are nursing assistants (NAs), who are individuals without formal training or registration requirements and who have responsibility only for basic patient tasks. RNs are able to perform all the tasks that NAs can, but RNs perform many tasks that require specific training and cannot or should not be performed by NAs (NICE, 2014b).

In the NHS both groups are paid according to a national set of pay bands called 'Agenda for Change' (AfC), in which higher bands correspond to higher pay. NAs are typically Band 2, within which there are multiple spine points which individuals progress up based on tenure. RNs nurses are on bands 5–8, each corresponding to a different level of responsibility: band 5 contains newly qualified and less experienced nurses; band 6 contains more senior and experienced nurses, including deputy ward managers; and band 7 and 8 contains the most senior nurses and ward managers with the most responsibility. Each pay band has multiple spine points to account for tenure. Promotion between pay bands 5 to 8 is based on experience and generally requires specific training or additional qualifications.

Patients do not have a named nurse and nurses in the ward are typically responsible

for team output in our setting) these mechanisms may underlie our results, but we do not have the data to directly examine them.

for delivering care to all patients in that ward. While there are some tasks where a single nurse can safely care for multiple patients, other tasks require team members of different levels of skill and experience to work together over the course of a shift. This requires collaboration and coordination. In addition, there is a need to convey information both within teams and, when shifts change, across teams. Thus the output is a team output and, in contrast to other settings in which the output of each team member can be separately observed and recorded (e.g. Bandiera et al. (2009, 2013)), cannot be attributed to a single individual. The team is therefore the natural unit of analysis for a study of the impact of nurse staffing on patient outcomes. It is also of direct policy relevance as the team is the unit at which care is given and staffing is required and planned.

Our focus is on the effect of shortages of team members of different levels of general human capital, seniority and experience on patient outcomes. General human capital – such as overall education and experience – should raise an individual's productivity (Becker, 1964; Mincer, 1974) in nursing, as in other forms of production that require individuals with detailed knowledge and the ability to take decisions without high levels of supervision. Thus we may expect differences in outcomes when there is a shortage of qualified nurses (RNs) to when there is a shortage of less-qualified nursing assistants. Nursing staff differ also in their experience in the hospital, experience in the ward and experience with working with other members of the same team. Lack of each type of experience may also have a detrimental effect on patient outcomes.[4] Finally, given that wards have a hierarchical structure, we might expect that the absence of a more senior RN team member has a larger effect than the absence of one who is less senior.

2.1.2 Workforce planning

There are no mandated nurse:patient ratios in the UK. National guidelines state that there is no single ratio that can be applied to meet the needs of patients across the range of ward types. Instead, hospitals are required to go through a two step process (NICE, 2014a,b) to plan safe staffing. The first is to estimate the average nursing need in terms of working time in the ward during a 24-hour period, in a systemic way that is endorsed by the regulator. The second is to determine the required level and mix of nursing staff to be employed, and how they should be allocated across wards and shifts.

[4] Huckman and Pisano (2006) provide evidence of returns to hospital-specific (firm-specific) experience for surgeons carrying out cardiac surgery. Bartel et al. (2014) show returns to ward-specific experience in nursing, and Chen (2021) gives evidence of returns to team-specific experience among physicians.

This is informed by the first step and is supplemented by professional judgement at the hospital level to produce a planned level of staffing for each ward and shift. While the guidance does not mandate a particular nurse:patient ratio or skill mix, it lays down advice and requirements that constrain hospitals' choices. For example, having fewer than two registered nurses on a shift in any ward is considered a 'red flag' that should be escalated and remedied, and it is advised that a registered nurse should not be in charge of more than eight patients (NICE, 2014b).

For the first stage of this process, the hospital we study uses the 'Safer nursing care tool' (SNCT), a national, evidence-based tool endorsed by the UK's National Institute for Health and Care Excellence. The SNCT is based on expected levels of patient acuity and dependency. Each January and June, in all wards, the patients in that ward are placed into one of 5 groups based on levels of care required. This data is collected at 15:00 daily, Monday to Friday, for a minimum of 20 days. Each of the patient acuity groups has a different staff multiplier (0.99 for the lowest category, 5.96 for the highest: see Shelford Group (2015); Griffiths et al. (2020)). The SNCT estimate of staffing requirements is given by multiplying the number of patients in each category by their multiplier and then adding across the five groups. This is then combined with professional judgement on other factors that might affect safe staffing levels (such as ward layout and size, and nursing responsibilities other than patient care) to generate an estimate of the total nursing requirement for each 24-hour period (NICE, 2014b; Shelford Group, 2015). The second step takes these results and applies professional judgement and advice from the guidelines to calculate the required number of staff on the ward, split into NAs and RNs. This is then used to generate planned shifts, which can be thought of as the 'target' level of safe staffing. These are generated 2-3 months in advance of the required staffing.

Shifts may be filled by 'local' nurses (employees of the hospital, working their contracted hours in the wards to which they are attached), 'bank' nurses (nurses employed by the hospital but doing extracontractual or overtime work via the hospital 'staff bank', in their usual ward or another ward) or an agency nurses (nurses employed via an external agency and generally contracted at short notice).

The staffing guidelines recommend that there is an on-the-day assessment to ensure that the nursing staff on shift can meet the needs of patients on the ward (NICE, 2014a). This might arise because there are fewer staff than planned due to absences, or because the needs of patients on the ward are substantially higher than on an average day (NICE, 2014b). However, in reality, the ability of managers to adjust staffing at short notice is limited. The NHS – and the hospital we study – has experienced a shortage of nurses for many years. There are therefore not a large number of the hospital's directly employed staff (i.e. bank staff) who can frequently step in at short notice to cover unplanned absences. In addition, to control costs, the hospital has (along with all other NHS hospitals) introduced measures to limit the use of (expensive) agency workers. This means there may be no bank staff available, and permission to use an agency worker may not be granted in time to bring the staffing up to the planned levels.

2.2 Data sources

We exploit a novel data linkage between electronic staff roster data and electronic patient records for one NHS hospital, comprising a small number of acute hospital sites located close to one another within a large English city.[5] This hospital provides a full range of emergency and elective services, has a medical school attached, and is located in an area with an ethnically and socioeconomically diverse population.[6] We combine routinely recorded administrative information on the size and composition of nursing teams with information on the characteristics and outcomes of patients under their care. Our data covers 1st January to 31st December 2017.

2.2.1 Observations

The hospital we examine has three large sites providing short term general acute care. These sites contain 89 wards in total, but to reduce heterogeneity we focus on inpatient wards that provide 24/7 hour care for adult patients.[7] For each of these 52 wards, we construct measures of staffing, patient characteristics and patient outcomes at the 24 hour level. We define our observations at the 24-hour (henceforth referred to as day)-ward level as this is the unit at which staffing is planned. Our final sample contains 18,922 observations. 4,484 unique staff members worked in these observations, working 3,533,583 hours and 294,044 shifts over the course of the year. The staff in each observation were responsible for treating 44,485 unique patients with 66,662 separate

[5]Hospitals are known as NHS Trusts in England: for simplicity we refer to the Trust we study as a hospital.

[6]There are 25 medical schools in England, all of which are attached to one or more NHS Trusts. There were 31 NHS Trusts in England with 1,000 beds or more in the period January to March 2019. The Trust studied here belongs to both categories and is representative of these categories with regard to specialities, staff and patients.

[7]We exclude: 18 wards which do not record a patient death at any point in the year (which includes administrative wards and those that treat outpatients); 14 wards with non-continuous care, such as those that are closed on weekends or which regularly close to patients (defined as having an entire day with no patients in the ward 30 times or more over the year); and 5 maternity, paediatric and emergency department wards, which have very different staffing models and operational procedures.

hospital spells.[8] Each of these is a distinct team comprising the nursing staff members who worked together on a given calendar day (including the handover between those doing the night shift and those who take over the following morning).[9]

Whilst our primary analysis is at the team level, we also undertake analysis at the patient-day level. For this analysis, we take the sample of patient days in the 52 wards used in our team-level analysis over the calendar year, which gives a total of 276,237 observations.

2.2.2 Staffing

Our electronic roster data provides planned shifts. These data also show whether the shifts are ultimately worked or not. Shifts can go unfilled because of staff absences (due to sickness or annual leave, for instance), staff vacancies, or for other reasons, such as attendance at a training course. The data also include information on the AfC pay band of the nurse who worked each shift. We use this to identify staffing group (RNs and NAs) and seniority within RNs.

2.2.3 Measures of quantity of nurses

We measure the quantity of staffing in terms of deviations from the 'target' level of staffing, as set by the hospital management for each ward, informed by the SNCT. Specifically, we define the 'fill-rate' as the fraction of planned hours that were actually worked in ward i on day t:

$$FillRate_{it} = 100 * \frac{\text{Hours worked in ward } i \text{ on day } t}{\text{Planned hours in ward } i \text{ on day } t} \tag{1}$$

where a team with a fill-rate of 100 is fully staffed, with staffing equal to target.[10] We then group together all nursing assistants (NAs, those in AfC bands 2–4), and all registered nurses (RNs, those in AfC bands 5–8), and construct the fill rate (Equation 1) separately for NAs and RNs: the share of planned NA hours filled by NAs, $FillRateNA_{it}$, and the share of planned RN hours filled by RNs, $FillRateRN_{it}$.

[8]The sample size is less than 365 observations for each of the 52 wards (18,980 observations) because on 58 days (across 26 wards) no patients were treated and these days are excluded from our sample.

[9]As we define a team as a 'hospital ward-day', a night shift that spans midnight will contribute to the staffing of more than one team. For instance a shift that starts at 20:00 on day t and ends at 08:00 on day t+1 will contribute 4 hours to the staffing of the first team and 8 hours to the staffing of the second. The mean length of a nursing shift is 11.5 hours and more than 70% of shifts are 12 hours long.

[10]Staffing cannot exceed the target level in this setting, so the fill-rate has an upper bound of 100. The NHS runs very close to capacity at all times: bed occupancy is over 95% and there are waiting lists for planned (elective) care to ensure that beds are not left empty.

The ratio between actual and planned staffing – the 'fill-rate' – has a number of appealing features for our analysis. First, it is not sensitive to the absolute size of the ward in the question. A shortage of one staff member is a much more significant event for a team of five nurses than it is to a team of thirty: using the fill-rate captures this, because deviations are measured in percentage terms. Second, deviations are measured relative to the locally determined, ward-specific target which is considered by the hospital management to be the 'safe' level of staffing, thus allowing for heterogeneity in what is medically considered to be safe staffing across different hospital wards. Third, we can define the fill-rate separately for different groups of nurses, which allows us to study the impact of both size and composition in a comparable way across wards. In Section 5.3, we test the robustness of our results to alternative specifications of this treatment variable.

2.2.4 Measures of nurse seniority and experience

We separately identify the different grades (bands) within the RNs to distinguish between qualified nurses of different levels of seniority (and thus experience as a qualified nurse). We construct the fill rate for each band as a share of total planned RN hours. The sum of the fill rates for bands 5-8 will sum to $FillRateRN_{it}$.[11] This approach allows us to compare the impact, for instance, of a one percentage point increase in the RN fill-rate from a newly qualified (band 5) nurse versus a senior nurse manager (bands 7 and 8).

We observe the amount of time an individual has been employed in the hospital at any of its sites and use this to define experience in the hospital.[12] We also measure ward specific experience of the team. We do this three ways. First, we count the number of times that an individual has worked in a ward over the past 90 days and average this over the members of each team in our sample. We measure ward-specific experience based on the past 90 days for two reasons. One, we might expect the effects of experience to decay over time. Two, our sample covers a single calendar year and in order to construct a backwards-looking measure of experience, we have to discard a portion of

[11]For example, the band 5 fill rate $FillRateRN_{it}^{Band5}$ is equal to $100 *$ (Hours worked by Band 5 RNs in ward i on day t)/(Planned RN hours in ward i on day t). A full set of definitions is provided in the Appendix A.2

[12]For 89% of individuals in our sample (who work 87% of all shifts) tenure is directly observable in staff records, as these contain the date at which the individual's employment at the hospital started. For the remaining 11% of individuals, for whom this information is missing, experience in the hospital is imputed based on their pay band, the % of shifts they work as overtime via the staff 'bank', the % of shifts they work at night, and whether they are an agency worker. Agency workers are not employed by the hospital and in our analysis we give them a value of 0 on this measure. This is conservative if an agency worker is frequently employed by the same hospital.

the data at the start of the period (for which such a backwards-looking measure cannot be calculated). Using ward-specific experience over the previous 90 days strikes a balance between capturing how familiar workers will be with the ward, and having to discard no more than a quarter of our sample.

Our second measure is the fraction of the individual's shifts in the last 90 days that were in the ward, again averaged over the members of each team. For our third measure, we identify those who are less familiar with the ward. If an individual is within their first 30 days on the ward, having not worked there in the preceding 90 days, we define these individuals as 'new joiners' and construct a binary variable to indicate the presence of a 'new joiner' in a team.

Finally, we construct a measure of team-specific experience in three ways. First, as the average number of shifts that an individual has worked together with their team mates, in any ward, over the past 90 days. Second, as the fraction of their fellow team members an individual has worked with at least once in the past 90 days. Third, as the fraction of fellow team members an individual has worked with at least 10 times in the past 90 days. In our analysis, we use the average of each of these measures across the members of each team.

2.2.5 Patient outcomes and severity

For our primary analysis, we use an important and high-stakes outcome: inpatient mortality at the team level. It is a commonly used performance measure for hospitals and for nursing (Needleman et al., 2011; Griffiths et al., 2019; Peutere et al., 2024) and has the further advantage of being unambiguous and of being accurately and consistently recorded across wards. Death at discharge is recorded in electronic patient records, which we link to the team responsible for treating them at the point of death. Our outcome is binary, with an indicator coded as 1 if a team (ward-day) had a patient death occur, and coded as 0 otherwise.[13] Analysis of mortality at the team (ward-day) level overcomes the issue that mortality at the patient level is relatively infrequent. It also has the advantage that concerns that unobserved patient severity may bias the results are mitigated by averaging over the severity (observed and unobserved) of all patients under the care of the team.

We supplement all-cause mortality with measures of severity that allow us to consider the impacts on mortality among groups of patients who might be more susceptible

[13]Deaths occur at all hours of the day (Appendix Figure A1), with the only noticeable spike occurring between 19:00 and 20:00 (when the handover between day and night shifts takes place) which suggests that there is no systematic misallocation of patient deaths across calendar days.

to staff shortages. We use the International Classification of Diseases (10th edition; henceforth ICD-10) codes in the electronic patient records to select one condition that is both common and particularly likely to be sensitive to the quality of nursing care on inpatient hospital wards: sepsis. Sepsis is a serious and potentially life-threatening condition, driven by an extreme immune response to an infection. It claims more lives in England each year than lung cancer (Health Education England, 2023). In the US, one-third of people who die in hospital had sepsis during that hospitalisation (Centre for Disease Control and Protection, 2023). In our sample the equivalent figure is 26.3%. The high prevalence of sepsis should make it easier to detect a statistically significant impact of nurse staffing than for other, rarer, diagnoses for which outcomes may also be nurse sensitive.

Sepsis presents through deterioration in key physical signs (e.g. rapid falls in blood pressure and increases in heart rate). Early identification is key to successful treatment (Torsvik et al., 2016), and English hospitals (including the one we study) have widely adopted a National Early Warning Score (NEWS), aimed at improving detection of patient deterioration, with a particular focus on sepsis (Inada-Kim, 2022; NHS England, 2023).[14] Timeliness of treatment following detection is extremely important: national clinical guidelines state that patients with suspected sepsis in an acute hospital setting should be escalated to a senior clinical professional immediately (Royal College of Nursing, 2023), and a broad-spectrum antimicrobial at the maximum recommended dose should be administered within an hour of the decision to treat (NICE, 2016).[15]

Nurses are crucial to the identification and treatment of sepsis, as they have the most direct contact with patients of any group of clinicians, are responsible for monitoring patients, for escalating to a ward manager or doctor following a detection of suspected sepsis, and for administering treatment (Kleinpell, 2017). An empirical relationship between nurse staffing and outcomes of sepsis patients in the US has been documented in Lasater et al. (2021) and Dierkes et al. (2022). Based on the literature and the issuance of guidance on sepsis by the NHS regulator (NICE), which focuses on the role of nurses, we posit that shortages of nurses would be expected to have a greater impact on the mortality outcomes of patients with sepsis than on patients with other conditions, such as cancer, for which timeliness of care while in hospital may be less critical to short run

[14]Other examples include the ‘Sepsis six’ in New Zealand (Kleinpell, 2017) and Rory’s Regulation in the US (Lasater et al., 2021).

[15]This is also the case in the hospital we study, where the protocol states that the first action in the case of an elevated National Early Warning Score should be to inform the nurse in charge, followed by the establishment of continuous monitoring (with observations to be recorded every 10 to 15 minutes) until the doctor arrives.

mortality.

We also use the ICD-10 codes to construct a standard and well-established measure of clinical severity, the Elixhauser Comorbidity Index score (Elixhauser et al., 1998; Quan et al., 2005; Van Walraven et al., 2009), for each patient. The score summarises a set of 30 categories of comorbidity (diseases), with each patient assigned a score that represents a weighted count of the number of those diseases the patient has been diagnosed with.[16] In our sample, patients' Elixhauser Comorbidity Index scores range from 0 to 13. The higher the score, the greater the clinical severity and mortality risk of the patient. We also use each patient's Elixhauser Cormorbidity Index score, plus a range of other demographic information, to predict patient level mortality on each given day.[17] The R^2 for the regression used to predict inpatient mortality is 0.11, which is similar to or higher than that in other studies which use a similar approach with hospital discharge data (e.g. Cooper et al. (2022)).

In addition to inpatient mortality, we construct two other nurse-sensitive outcomes: patient falls and transfers to an Intensive Care Unit (ICU). Inpatient falls are the most commonly reported safety incident in the NHS, and are a major concern in hospitals worldwide due to the risk of physical injury and psychological harm among frail patients (Morris and O'Riordan, 2017; Morris et al., 2022; Donaldson et al., 2005; Patrician et al., 2011; Kalisch et al., 2012). Fall prevention requires both the identification of patients with a fall risk, and the design and implementation of interventions to manage that risk. Nurses play a crucial role in this, and therefore the risk of falls could increase when there are unfilled shifts (Dykes et al., 2009; Morris et al., 2022). A transfer to an ICU may indicate that a patient's condition has deteriorated (e.g. due to poor monitoring by a short-staffed team)[18] or that the team on the non-ICU ward no longer feel confident in their ability to provide safe care to the patient.

Table 1: Summary statistics

	Mean	SD
Staffing: quantity		
Fill-rate (all)	95.7	5.4
Nursing Assistant (NA) fill-rate	94.8	10.9
Registered Nurse (RN) fill-rate	96.4	5.8
Planned NA hours	53.4	30.7
Shortfall relative to planned NA staffing (hours)	3.2	6.1
Planned RN hours	141.5	91.5
Shortfall relative to planned RN staffing (hours)	4.8	7.6
RN hourly shortfall above ward median (binary)	0.32	0.47
RN hourly shortfall $\geq$ 1 SD above ward mean (binary)	0.15	0.36
RN hourly shortfall $\geq$ 4 hours (binary)	0.36	0.48
RN hourly shortfall $\geq$ 8 hours (binary)	0.33	0.47
RN hourly shortfall $\geq$ 12 hours (binary)	0.22	0.42
RN hourly shortfall $\geq$ 16 hours (binary)	0.11	0.31
RN hourly shortfall $\geq$ 20 hours (binary)	0.07	0.25
RN hourly shortfall $\geq$ 24 hours (binary)	0.03	0.18
Staffing: contract type and seniority		
Band 5 RN fill-rate	68.4	15.8
Band 6 RN fill-rate	22.8	13.3
Band 7–8 RN fill-rate	5.2	5.8
Staffing: firm-specific experience		
Mean years of experience in the hospital group (all)	6.3	2.0
Mean years of experience in the hospital group (NAs)	5.6	3.2
Mean years of experience in the hospital group (RNs)	6.7	2.5
Staffing: ward-specific experience		
Mean shifts in ward in past 90 days (NAs)	27.8	9.7
Mean shifts in ward in past 90 days (RNs)	31.5	5.9

	Mean	SD
Staffing: team-specific experience		
Mean shifts with other team members in past 90 days (NAs)	13.5	6.3
Mean shifts with other team members in past 90 days (RNs)	14.9	5.1
Patient characteristics		
Age	64.16	9.84
Prop. (%) aged 65+	0.54	0.24
Prop. (%) aged 85+	0.14	0.17
Female %	0.46	0.18
Non-white %†	0.52	0.16
Elixhauser Comorbidity Index	3.69	1.17
Length of hospital stay (days)‡	28.55	17.07
Number of patients treated in the ward	20.65	8.41
Patient outcomes		
Death (binary)	0.063	0.244
Death rate per 1,000 patients	3.59	16.22
Death of low-severity patient (binary)*	0.036	0.186
Death of high-severity patient (binary)*	0.030	0.171
Death of patient with sepsis diagnosis (binary)**	0.018	0.131
Death of patient without sepsis diagnosis (binary)	0.047	0.213
Patient transferred to Intensive Care Unit (ICU) (binary)	0.099	0.298
Patient transferred to non-ICU ward (binary)	0.603	0.489
Fall (binary)	0.070	0.256
Predicted number of patient deaths***	0.058	0.069
Observations (teams)	18,871	
Clusters (hospital wards)	52	
Calendar days	365	

† Non-white includes patients with ethnicity recorded as 'unknown'. ‡ Calculated as the average for the patients treated in a given unit on a given day (so that patients with a stay of more than one day are counted multiple times), rather than the average for the universe of patients. * Low-severity patients are defined as those with an Elixhauser index score of less than 5; high-severity patients are defined as those with an Elixhauser index score of at least 5. ** A patient is coded as having sepsis if they have a primary or secondary ICD-10 diagnosis code of A400, A403, A409, A410, A411, A414, A415, A418, or A419. *** Predicted mortality is calculated at the patient level and then aggregated up to the team (hospital ward-day) level. Values are between 0 and 1, and are predicted using patient age; age squared; age cubed; sex; ethnicity; Elixhauser Cormorbidity Index; pairwise interactions between each of these; dummy variables for 17 separate diseases; how long the patient had been in hospital up to that point; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

3 Summary statistics

3.1 Staffing and team composition

The average team (ward-day) had 53.4 rostered NA hours and 141.5 rostered RN hours, or around 194.9 hours in total. This is equivalent to around 16 staff members working 12-hour shifts, with around 4 assistants and 12 qualified nurses in the average team. Not all of these staff members will be working at any one time and shifts overlap in order to ensure continuous care.

[16] The measure has been validated and found to perform better than other commonly used comorbidity indices in predicting patient mortality (Southern et al., 2004; Menendez et al., 2014).

[17] Mortality is predicted at the patient-day level using patient age; age squared; age cubed; sex; ethnicity; Elixhauser Cormorbidity Index; pairwise interactions between each of these; dummy variables for 17 separate diseases; how long the patient had been in hospital up to that point; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

[18] Bapoje et al. (2011) and Marquet et al. (2015) both find evidence that a substantial proportion of ICU transfers were preventable with earlier interventions.

Our analysis exploits variation in team level staffing hours as a percentage of planned hours (i.e. the fill rate). Table 1 reports summary statistics for the fill-rate by staffing group and also reports the average level of different types of experience of the team. In just over half (51.2%) of teams, the actual number of hours worked was less than the number of planned/rostered hours, so just over half of teams had a fill-rate of less than 100. The mean fill rate for the average team was 95.7% (i.e. 95.7% of planned hours were actually worked), equivalent to an average of 8 hours of planned staffing unfilled, or two-thirds of an average shift unfilled. The fill rate was slightly lower for NAs (94.8%) than for RNs (96.4%). These figures differ because these are distinct staffing groups with different rates of absence; it could also be that more effort is made to fill rostered RN shifts due to their critical role in patient care. There is little correlation between the NA and RN fill-rates (coefficient=0.066). Nurses in pay band 5 comprise the largest staff group, accounting for more than two-thirds (68.4%) of rostered RN hours. Band 7 and 8 staff – the most senior nurses responsible for managing their ward – work around 5% of all rostered RN hours. The majority of RN hours – 70.5% of those planned – were worked by 'local' staff in their regular ward to which they are attached and in which they do their contracted hours. A further 16.3% were worked by bank staff and a further 9.6% by agency workers.

Importantly, there is considerable day-to-day within-ward variation in both the level and composition of staffing. Appendix Table A1 shows that there is more variation in key staffing variables (such as the NA and RN fill-rate) *within* wards than *between* them (illustrated graphically in Appendix Figure A2). This is key to our identification strategy. There is also considerable variation in the degree of hospital-, ward- and team-specific experience amongst the team. In part, this variation is because of staff rotations, arrivals and departures which contribute to turnover. But in addition most staff members work across multiple wards, as shown in Appendix Figure A3. Junior staff, in particular, tend to work across multiple sites and wards within the hospital. The average NA worked across 1.7 hospital sites and 6.8 unique wards. The average newly qualified RN (band 5) worked across 1.5 hospital sites and 4.5 unique wards.

At the patient-day level, the summary statistics are much the same: the average patient-day has a RN fill-rate of 96.2% with a shortfall of 5.4 RN hours relative a target of 145.2 hours (versus 96.4%, 4.8 and 141.5 at the team level, respectively). These are reported in Appendix Table A2.

3.2 Patients and patient outcomes

The average team was responsible for treating 20.7 patients over the course of the 24-hour ward-day (Table 1). The average patient was 64.2 years old, 46% were female, and 52% were recorded with an ethnicity other than White (including 'unknown'). The average severity, as measured by the Elixhauser Comorbidity Index score (described in Section 2.2.5), was 3.69. 6.3% of the teams in our sample (1,197 of 18,922) recorded a patient death, close to the 5.8% predicted by our estimates of mortality from the individual patient data. Around 10% of teams transferred a patient to an Intensive Care Unit (ICU) and 7% recorded a patient fall.

At the patient-day level, the average age was 64.5; 47% were female; the average Elixhauser Comorbidity Index score was 3.49; and 10% were diagnosed with sepsis (Appendix Table A2). A death occurred on 0.4% of patient days (the equivalent figures for falls and transfers to ICUs were also 0.4%).

4 Empirical strategy

4.1 Baseline model

We begin by estimating a simple production model in which the probability of a patient death is a function of a shortage of nurse staffing. We allow for different effects of shortages of NAs and RNs, to reflect their differing roles and levels of human capital.[19] We begin by assuming production is additively separable in the two types of staffing, with no explicit interactions or complementarities between different types of nursing staff.

In this simplest model, we assume that healthcare G is produced according to $G(L_{NA}, L_{RN}) = \beta_1 L_{NA} + \beta_2 L_{RN}$. To implement this, we estimate the following baseline fixed-effects logit model:

$$Pr(y_{it} = 1|z_{it}) = Pr(y^*_{it}|z_{it} > 0)$$

$$y^*_{it} = \beta_1 FillRateNA_{it} + \beta_2 FillRateRN_{it} + \delta X_{it} + \lambda_t + \eta_i + \varepsilon_{it} \tag{2}$$

[19] An absent team member has a both a direct impact on team production by causing a team to be one member short, but may also have an indirect effect if they are replaced by someone with lower skills or familiarity with the production process (Herrmann and Rockoff, 2012; Bartel et al., 2014). Hoey et al. (2022) distinguish between the two in a setting where teams are always fully staffed. In our context they are not and therefore we do not make a distinction between direct and indirect effects.

where y_{it} is a binary indicator of a patient death in ward i on day t (where the ward-day is the team) and z_{it} is a vector of explanatory variables. y^*_{it} is a latent variable where we observe $y_{it} = 1$ when $y^*_{it} > 0$.

Staffing quantity is measured by the fraction of planned NA and RN hours that were actually worked, denoted by $FillRateNA_{it}$ and $FillRateRN_{it}$, respectively (as defined in Section 2). X_{it} measures time-varying characteristics of patients in ward i on day t. We include the mean age of patients; the mean age squared; the female share; the share of patients who are non-white; the mean patient Elixhauser Comorbidity Index score; pairwise interactions between each of these variables; and the average length of hospital stay for the patients in the ward. We also include the number of patients treated in ward i on day t to control for variation in workload. λ_t captures time effects (day of week; month; and a dummy variable indicating a public holiday). Ward fixed effects (η_i) capture time-invariant differences in physical capital and the type of patient across wards (such as their medical specialty, patient mix, hospital management quality, and physical layout). Standard errors (ε_{it}) are clustered at the ward level, but we show that our results are robust to specifications that use alternative structures.

β_1 and β_2 are the coefficients of interest, and represent the impact of a change in the quantity of nursing assistants and qualified nurses, respectively, on the probability of a team experiencing a patient death. We present our results as odds ratios for ease of interpretation.

Below we extend this model to examine whether there are dynamic effects, effects of various measures of experience, non-linear effects, impacts on other patient outcomes, robustness to OLS estimation of our main results and repeat our key analyses at the patient-day level.

4.2 Threats to identification

Identification in our baseline specification (2) requires the staffing shortfall to be exogenous to other (unobserved) determinants of patient outcomes. Let nurse staffing in ward i on day t be denoted by L_{it}:

$$L_{it} = L^*_{it} - \hat{L}_{it} \tag{3}$$

where L^*_{it} denotes the *planned* level of staffing in the ward (i.e. the target level of staffing, as determined by the staffing algorithm and management judgement) and $\hat{L}_{it}$ denotes the staffing shortfall – the gap between planned and actual staffing. In this set-up, the fill-rate can be written as:

$$FillRate_{it} = 100 * \frac{L^*_{it} - \hat{L}_{it}}{L^*_{it}} \tag{4}$$

Nursing rosters, reflected in L^*_{it}, are generated 2–3 months in advance, and so are effectively fixed, and should be captured by ward and time fixed effects. Variation in the fill-rate is therefore driven by variation in $\hat{L}_{it}$. Our identification comes from deviations that arise because the appropriate number and mix of staff is not present due to absences, departures and more general staffing shortages. These deviations are captured and summarised by $\hat{L}_{it}$. The crucial identifying assumption in (2) is that $\hat{L}_{it}$ (and, by extension, the fill-rate) is exogenous to other residual determinants of patient outcomes.

In this section, we consider three sets of relationships that could pose a threat to this key assumption: staffing levels and patient mix; staff shortages and ward performance; and nursing staffing and other healthcare inputs, in particular the staffing of other occupations.

4.2.1 Staffing levels and patient mix

Planned staffing levels (L^*_{it}) are exogenous to eventual deviations from expected patient volumes and severity, as they are determined several months before actual staffing takes place. But *actual* staffing (L^*_{it} - $\hat{L}_{it}$), and so the fill-rate, could be endogenous to patient severity if managers respond to an influx of sicker patients by adjusting either the quantity or type of staff on the ward. By definition, a more severe patient mix means a greater risk of a death. If staffing levels or quality adjust upwards in response to patient severity and if there is a negative relationship between staffing and the risk of death, this would bias our estimates towards zero and mean that we underestimate the true effect of higher nurse staffing on deaths.

The level of staff shortages in the NHS, and the difficulty of obtaining an external agency worker at the last minute, mean that the feasibility of adjusting staff mix to patient severity is limited. Nonetheless, we examine this assumption empirically in the following ways.

First, if there is no relationship between staffing and patient severity, we would expect to find no relationship between nurse staffing levels and *predicted* mortality, as the latter is based on patient severity. To test this we estimate a fixed-effects OLS regression, with the predicted number of patient deaths as the dependent variable regressed on the same independent variables as our main specification, equation (2). We estimate a fixed effects OLS model because predicted mortality is a continuous measure. The first column of Table 2 shows that the coefficients on both the NA and RN fill-rate are close to and not statistically significant from zero.

Second, to examine whether wards move patients out in response to deviations from expected staffing, which would indicate that the fill-rate might affect patient mortality by influencing the number or type of patients in the ward, we estimate the same fixed-effects OLS regression replacing predicted patient mortality with an indicator for an indicator for whether a ward transfers a patient to another (non-ICU) ward. Results in the second column of Table 2 indicate a well defined zero: there is no evidence that patient volumes are correlated with deviations from planned staffing.[20] This fits with a context where staff shortages are endemic and bed occupancy rates stand at around 90%, making it hard to move either staff or patients around.[21]

Table 2: Impact of nurse staffing on predicted mortality and patient transfers

	Outcome:	
	Predicted mortality	Patient transferred to a non-ICU ward (binary)
	(1)	(2)
Nursing Assistant (NA) fill-rate	-0.00005 (0.00003) [0.138]	0.00024 (0.00031) [0.434]
Registered Nurse (RN) fill-rate	-0.00004 (0.00007) [0.604]	-0.00019 (0.00066) [0.778]
Patient and time controls	✓	✓
Ward fixed effects	✓	✓
N	18,871	17,051
Adjusted R-squared	0.229	0.087
Mean of outcome variable	0.058	0.603

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Coefficients shown are from a fixed-effects OLS regression with hospital ward fixed effects (rather than a logit model, to allow for a continuous predicted mortality measure). Robust standard errors, clustered at the hospital unit level, are shown in parentheses; p-values are shown in square brackets. Predicted mortality is calculated at the patient level and then aggregated up to the team (hospital ward-day) level. Values are between 0 and 1, and are predicted using patient age; age squared; age cubed; sex; ethnicity; Elixhauser Cormorbidity Index; pairwise interactions between each of these; dummy variables for 17 separate diseases; how long the patient had been in hospital up to that point; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday. ICU = Intensive Care Unit. ICUs are excluded from the analysis of the impact on patient transfers to a non-ICU ward. The unit of analysis is the hospital ward-day. All regressions control for the mean patient age; the mean patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise interaction terms between age, female share, non-white share and Elixhauser Cormorbidity Index; the mean hospital length of stay of patients treated in the unit that day; the number of patients treated in the unit that day; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

Third, in the spirit of a balance test, we regress various staffing measures on the

[20] In Section 6 we repeat this analysis using our main fixed-effects logit specification and again find no statistically significant effect (results displayed in Figure 3).

[21] https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/nhs-hospital-beds-data-analysis accessed 29th January 2024

patient characteristics that are in equation (2). We find no evidence of a relationship between mean patient severity (as measured by the Elixhauser Comorbidity index) and the overall fill rate, the RN fill rate or the RN staffing shortfall ($\hat{L}_{it}$). These results are in columns 1, 2 and 3 of Table 3, respectively.[22]

Finally, we examine whether the margin that could be most easily used to respond to changes in severity – agency staffing – is a function of patient severity, and find no evidence that agency staffing is higher when average patient severity is greater (column 4 of Table 3). There is some evidence that agency staffing is higher when there is a greater *volume* of patients in the ward, but we control for numbers of patients in our analysis.

Taken together, this evidence supports our assumption that variation in fill-rates are quasi-random with respect to patient characteristics.

4.2.2 Staff shortages and ward performance

A second threat to our identification strategy arises if staff shortages (and so changes in the fill-rate) are endogenous to recent or expected ward performance. That is, staffing is lower because of 'no-shows' that are themselves a response to recent patient outcomes or a prediction of future outcomes. This would bias our results as $\hat{L}_{it}$ and the fill-rate would reflect variation in performance captured by those staff responses as well as any direct effect of staff shortages on outcomes.

Table 4 examines the effect of deaths in the ward on day $t-1$, $t-2$ and $t-3$ on the level and composition of staffing on day t. There is some weak evidence that a death on the previous day is associated with a lower RN fill-rate (column 2) at the 10% level. However, the effect is very small: a death on day $t-1$ is associated with a 0.23 percentage point reduction, which is equivalent to only 4% of a standard deviation, and small relative to a sample mean fill rate of 96.4%. No such associations are evident for deaths two or three days previously (and the coefficient changes sign in the case of the latter). There are no statistically significant effects on NA staffing, nor on the composition of RN staffing, nor on the level of agency staffing.

We also examine whether a recent death in the ward is associated with a greater number of subsequent sickness absences (Appendix Table A4). For qualified nurses we find no statistically significant relationship between deaths in the previous three days

[22] There is some weak evidence that the RN fill-rate is marginally lower (the RN staffing shortfall marginally higher) on days with older patients. This effect is very small and, in any case, would cause a (spurious) positive correlation between nurse staffing and patient mortality, thereby biasing our estimates upwards towards zero.

Table 3: Relationship between staffing and patient characteristics

	Staffing outcome:			
	Overall fill-rate	RN fill-rate	RN staffing shortfall (hours)	Agency RN fill-rate
	(1)	(2)	(3)	(4)
Mean age	-0.227*	-0.331*	0.338	-0.213*
	(0.134)	(0.178)	(0.294)	(0.124)
Mean age^2	0.001	0.002	-0.002	0.001
	(0.001)	(0.001)	(0.002)	(0.001)
Female share	-0.587	-0.287	-3.760	1.267
	(3.647)	(4.874)	(6.185)	(3.541)
Non-white share	-1.214	-0.628	4.778	-1.016
	(3.443)	(3.945)	(5.507)	(3.285)
Mean Elixhauser Comorbidity Index	0.441	0.319	0.405	-0.159
	(0.635)	(0.804)	(1.408)	(0.570)
Age * Female	0.033	0.040	0.035	0.006
	(0.063)	(0.094)	(0.126)	(0.061)
Age * Non-white	0.051	0.061	-0.133	0.048
	(0.045)	(0.056)	(0.088)	(0.051)
Age * Elixhauser	-0.003	-0.001	-0.011	0.004
	(0.008)	(0.010)	(0.017)	(0.008)
Female * Non-white	-2.378	-4.360	4.810	-3.713
	(2.208)	(3.008)	(4.072)	(2.276)
Female * Elixhauser	-0.065	0.001	-0.339	0.036
	(0.350)	(0.460)	(0.698)	(0.326)
Non-white * Elixhauser	-0.371	-0.456	0.777	-0.158
	(0.300)	(0.357)	(0.527)	(0.294)
Number of patients in ward	0.046	0.061	-0.003	0.081***
	(0.029)	(0.038)	(0.038)	(0.026)
Mean hospital length of stay	-0.001	0.003	-0.006	0.014*
	(0.008)	(0.010)	(0.014)	(0.007)
NA fill-rate	×	✓	✓	✓
Local and Bank RN fill-rate	×	×	×	✓
Time controls	✓	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓	✓
N	18,871	18,871	18,871	18,871
Adjusted R-squared	0.386	0.042	0.042	0.728
Mean of outcome variable	95.7	96.4	4.84	9.55

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Coefficients shown are from a fixed-effect regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses. The unit of analysis is the hospital ward-day. Time controls include a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

and subsequent sickness. The coefficients are also small and vary in sign. For NAs, while there is a statistically significant increase in sickness two days after a death on the ward, it is not large, and the coefficients for days $t-1$ and $t-3$ are negative and not statistically significant. We also do not find any evidence of serial correlation in deaths, which lessens concerns that nursing staff might be able to predict a difficult shift, as well as concerns about serial correlation in the error term (Appendix Table A5). Taken together these tests strongly suggest that staff shortages are not a response to recent or expected patient mortality outcomes.

Table 4: Impact of deaths on previous days on levels of nurse staffing

	Outcome:		
	NA fill-rate	RN fill-rate	Band 5 RN fill-rate
	(1)	(2)	(3)
Death in the ward on day T-1	0.054 (0.245) [0.827]	-0.232* (0.120) [0.058]	0.094 (0.273) [0.731]
Death in the ward on day T-2	-0.029 (0.394) [0.941]	-0.072 (0.154) [0.640]	-0.284 (0.343) [0.411]
Death in the ward on day T-3	-0.119 (0.217) [0.585]	0.239 (0.150) [0.118]	0.195 (0.261) [0.458]
Mean of outcome variable	94.8	96.4	68.4
	Outcome:		
	Band 6 RN fill-rate	Band 7-8 RN fill-rate	Agency RN fill-rate
	(4)	(5)	(6)
Death in the ward on day T-1	-0.232 (0.262) [0.381]	-0.095 (0.095) [0.322]	0.225 (0.262) [0.395]
Death in the ward on day T-2	0.268 (0.269) [0.323]	-0.057 (0.148) [0.703]	-0.018 (0.230) [0.938]
Death in the ward on day T-3	0.016 (0.190) [0.934]	0.028 (0.135) [0.834]	0.186 (0.246) [0.452]
Mean of outcome variable	22.8	5.2	9.5
Patient and time controls	✓	✓	✓
Ward fixed effects	✓	✓	✓

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Coefficients shown are from a fixed-effects regression with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses. The unit of analysis is the hospital ward-day. All regressions control for the mean patient age; the mean patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise interaction terms between age, female share, non-white share and Elixhauser Cormorbidity Index; the mean hospital length of stay of patients treated in the unit that day; the number of patients treated in the unit that day; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

4.2.3 Other staffing adjustments

A final concern is that deviations in nurse staffing might be correlated with variation in other healthcare inputs. For example, doctor staffing could be correlated with nursing team composition if there are common shocks (due to e.g. weather or illness), or if hospitals bolster teams short-staffed of nurses by bringing in more doctors. This would mean that $\hat{L}_{it}$ and the fill-rate would be correlated with other factors that influence patient outcomes.

We are unable to examine this with data as the doctor rostering system in our setting is completely orthogonal to the rostering system for nurses, most doctors are not attached to a single ward (as in the English setting more generally), and the doctor rostering data is not available to researchers. However, we are confident that this is not a major threat to our identification strategy. The two rostering systems operate completely separately and both are set several months in advance by different and siloed management teams who do not co-ordinate with one another. The number of doctors is not, for example, factored into the SNCT that is used to inform nurse staffing. Further, clinical leaders in the hospital we study stated that it is exceedingly unlikely that a shortage of nurses would lead to additional doctors being drafted in to work on a ward. In any case, a junior doctor would be unlikely to fill in for a missing nurse; if anything, a senior nurse is more likely to step into the shoes of a missing or inexperienced junior doctor. Finally any common shocks, any effects of seasonal rotation, and any permanent differences in doctor quality across wards should be captured by our fixed effects specification.

5 Results: mortality

5.1 Baseline results

The first column of Table 5 reports the results from our baseline model, regressing a binary indicator of a patient death on the nursing assistant (NA) and registered nurse (RN) fill-rate (as defined in Section 2), as well as the controls discussed above. A one-unit (one percentage point) increase in each of these variables is equivalent to an additional 0.53 and 1.41 hours, respectively, for the average team (see Table 1), though this will vary according to the size of the ward.

An increase in nursing assistant staffing has no statistically significant impact on the odds of a patient death. In contrast, higher levels of qualified nurse staffing are associated with lower patient mortality: a one percentage point increase in the RN fill-rate reduces the odds of a patient death by 1.3% (with an odds ratio of 0.987). Our estimates suggest that an additional 12-hour shift by an RN would reduce the probability of an actual patient death by around 0.6 percentage points, equivalent to around 10% of the sample mean.

We repeat this analysis using a fixed-effect OLS regression to directly compare to the results for predicted mortality reported in Section 4.2. Appendix Figure A5 reports these results, as well as those for an alternative specification that uses panel-corrected AR(1)

Table 5: Impact of nurse staffing on patient mortality

	Outcome: patient death				
		Additional staffing measure:			
	Baseline model	Seniority	Firm-specific experience	Ward-specific experience	Team-specific experience
	(1)	(2)	(3)	(4)	(5)
Nursing Assistant (NA) fill-rate	1.003 (0.004) [0.360]	1.003 (0.004) [0.348]	1.003 (0.003) [0.342]	1.001 (0.004) [0.758]	1.001 (0.004) [0.757]
Registered Nurse (RN) fill-rate	0.987** (0.006) [0.038]		0.988** (0.006) [0.032]	0.983** (0.007) [0.022]	0.983** (0.007) [0.022]
RN fill-rate, of which:					
Band 5 nurses		0.989* (0.006) [0.087]			
Band 6 nurses		0.984** (0.007) [0.020]			
Band 7 and 8 nurses		0.976** (0.011) [0.025]			
Average firm-specific experience among NAs (years)			1.017 (0.015) [0.258]		
Average firm-specific experience among RNs (years)			0.920*** (0.018) [<0.001]		
Average shifts in the ward in past 90 days (NAs)				1.001 (0.005) [0.891]	
Average shifts in the ward in past 90 days (RNs)				0.996 (0.008) [0.618]	
Average shifts with team members in past 90 days (NAs)					1.004 (0.010) [0.701]
Average shifts with team members in past 90 days (RNs)					0.990 (0.012) [0.405]
Patient volume and characteristics	✓	✓	✓	✓	✓
Time controls	✓	✓	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓	✓	✓
N	18,922	18,922	18,001	13,598	13,598
Clusters	52	52	52	52	52
Mean of outcome variable	0.063	0.063	0.063	0.063	0.063

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Coefficients shown are odds ratios from a fixed-effect logit regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses; p-values are shown in square brackets. The unit of analysis is the hospital ward-day. All regressions control for the mean patient age; the mean patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise interaction terms between age, female share, non-white share and Elixhauser Cormorbidity Index; the mean hospital length of stay of patients treated in the unit that day; the number of patients treated in the unit that day; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

standard errors, and those for an alternative specification with wild bootstrapped confidence intervals.[23] There is, as noted above, no relationship between predicted mortality and staffing. There is, on the other hand, a statistically significant negative relationship between RN staffing and patient mortality. The standard error corrections do not change our baseline results.

5.2 Dynamic effects

Our baseline model captures only the contemporaneous impacts of a shortage of nurse staffing on the outcomes of patients. This follows much of the previous medical and nursing literature, which typically examines same-day effects of staffing disruptions (Needleman et al., 2011; Griffiths et al., 2016, 2019; Redfern et al., 2019; Dall'Ora et al., 2020). A focus on contemporaneous effects is also driven by the fact that length of stay in our sample (and for English hospitals more generally) is short. The median length of stay for patients in our sample is 26 hours; the median patient is treated by 2 teams; and 36% of patients were treated by only one team, limiting the scope for substantive dynamic effects.[24]

Nonetheless, it could be that the negative effects of being treated by a short-staffed team do not materialise immediately and instead lead to the patient dying some time later under the care of a different team. One way to look at this is to examine whether nurse staffing in a ward on a particular day has an impact on mortality outcomes for patients treated by the team on subsequent days. We examine this in Appendix Table A6. The outcome variable is a binary indicator equal to one if a patient treated in ward i on day t dies under the care of a different team within 1, 2 or 3 days. We find no evidence of an impact of the NA fill-rate or RN fill-rate on subsequent deaths among patients treated by the team.

However, an examination of dynamic effects in our setting is complicated by the high degree of correlation in staffing across days (driven by shifts that span two calendar days and so generate mechanical correlation; sickness absences that span more than a 24 hour period; and the fact that it might take time to replace someone who unexpectedly leaves the team). The serial correlation in RN staffing is documented in Appendix Table A7. If we think about a staffing disruption as a 'treatment', wards are treated multiple times, on a semi-regular basis, and these treatments are serially correlated over time. This means that if we regress deaths on day t on lagged and leading measures of the fill-rate,

[23] The latter is informed by Carter et al. (2017), who note that when the *effective* number of clusters is small, the critical values from a normal distribution can be too small with cluster heterogeneity. We follow Lee and Steigerwald (2018) and report wild bootstrap critical values.

[24] This also means that an exploration of dynamic effects makes more sense at the team, rather than the patient or patient-day, level.

the coefficient on the RN fill-rate on day $t-1$ or $t+1$ will partly reflect the impact of the RN fill-rate on day t, with which the lagged and leading variables are correlated.[25]

Thus to allow for multiple treatments and look more directly at potential dynamic effects, we estimate a 'stacked' event study of the form:

$$y_{it} = \sum_{h=-3}^{3} \sum^{j} \psi_h \cdot E_{ij} \cdot 1(j-t=h) + \theta \chi_{it} + \alpha_i + \gamma_t + \epsilon_{it} \tag{5}$$

where y_{it} is the outcome of interest (in this case, a binary indicator of a patient death in ward i on day t). E_{ij} is a binary indicator equal to one if there was a staffing disruption in ward i on day j. $h \in \{-3,-2,-1,0,1,2,3\}$ is the days between the the outcome on day t and the disruption on day j. The combined expression $E_{ij} \cdot 1(j-t=h)$ is therefore equal to one if there was a disruption on day j *and* the event happened h days after date t (if $j-t=h$). It is possible for this expression to be equal to one for multiple values of h if the ward experienced multiple staffing disruptions (e.g. both one day before and two days before). For this reason, there is no omitted base period. This is also what allows us to separately identify the impact of staffing disruptions at different times relative to the outcome on day t.

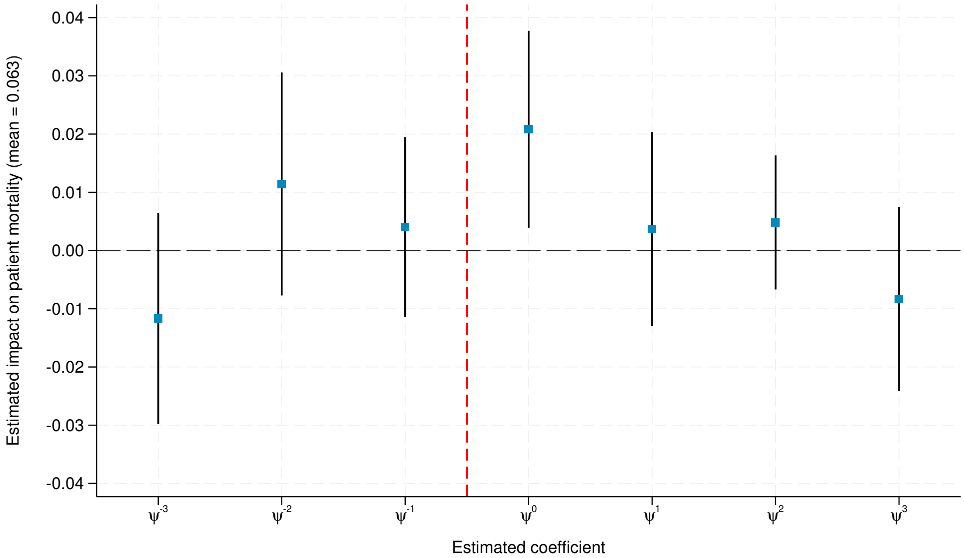
ψ_h represents the impact of a staffing disruption on day $t+h$ on patient mortality on day t: these are the coefficients of interest. χ_{it} is the same set of time-varying patient controls as in the baseline model (Equation (2)), plus the planned level of RN staffing in ward i on day t. α_i is a set of ward fixed effects, and γ_t is the set of time fixed effects as in Equation (2).

In this event study we focus on a large staffing disruption, which we define as a staffing disruption as occurring when: (1) the difference between actual RN staffing and target/planned RN staffing is ≥ 16 hours and (2) this was 'escalated' by the manager (i.e. they unsuccessfully posted the shifts to staff outside of the ward via the staff bank and/or external agencies).[26] The results are in Figure 1. While somewhat noisy, they are consistent with a limited role for dynamic effects. The estimated coefficient for ψ_0 (the contemporaneous impact of a staffing disruption on patient mortality) is positive and statistically significant at the 90% level. Estimated coefficients for all other values of h are close to and not statistically significantly different from zero. We interpret this as supporting our focus on contemporaneous impacts in our primary analysis.

[25] The results from this analysis are reported in Appendix Figure A6. The coefficients point to a limited role for dynamic effects, but the interpretation is complicated by the serial correlation in staffing described in the text.

[26] In Section 5.3 we also show robustness of our main results to this specification of the treatment variable.

Figure 1: Event study design: impact of RN staffing shortage on patient mortality



Note: 90% confidence intervals shown. RN= Registered Nurse. Estimated coefficients along the horizontal axis correspond to those in equation 5. The treatment (a 'staffing shortage') is defined as occurring when: 1) the difference between actual RN staffing and target/planned RN staffing is $\geq$ 16 hours; and 2) this was 'escalated' by the manager (i.e. they unsuccessfully posted the shifts to staff outside of the ward via the staff bank and/or external agencies). The robustness of our main results to this measure is displayed in Figure 2.

5.3 Alternative specifications of the treatment variable

5.3.1 The effect of seniority/experience

So far we have examined the differential impact of Nursing Assistants (NAs) versus Registered Nurses (RNs). We now distinguish between RNs of different seniority (i.e. in different pay bands). Nurses in higher pay bands are more experienced, generally have more qualifications, and have more responsibilities, including for team organisation, than the RNs in lower bands.[27] The second column of Table 5 distinguishes between qualified nurses of different seniority. Our point estimates indicate that a one percentage point increase in the share of planned RN (band 5–8) hours worked by a band 5 nurse would reduce the odds of a patient death by 1.1%, increasing to 1.6% for band 6, and 2.4% for band 7 and 8. While these estimates are not statistically significantly different from one another, the ordering and relative magnitudes of the point estimates are indicative of greater returns to higher seniority. For the average team, adding an extra nurse in the

[27]For example, the protocol for early detection of patient deterioration in the hospital we study indicates that the first action should be escalation to the nurse in charge.

most senior pay bands reduces the odds of a patient death by more than twice as much as adding a newly qualified (band 5) nurse.

The third column of Table 5 examines the impact of firm-specific experience, controlling for the quantity of NA and RN staffing. The estimates show that a greater amount of firm-specific experience among RNs is associated with a reduction in patient mortality. Increasing the average firm-specific experience among RNs in the team by one year reduces the odds of a patient death by 8.0% (equivalent to roughly three-quarters of the impact of adding an extra RN).[28] Increasing average firm-specific experience by one standard deviation (2.5 years, see Table 1) reduces the odds of a patient death by 18.8%. In contrast, a greater level of firm-specific experience among nursing assistants has no statistically significant impact.

The fourth column of Table 5 examines the impact of recent work experience in the ward. The coefficients indicate that recent ward-specific experience – as measured by the number of shifts in the ward in the past 90 days – has no statistically significant effect. The fifth and final column of Table 5 considers the effect of familiarity of team members. The results indicate that the average amount of shared experience in the team has no statistically significant effect on team performance, for both NAs and RNs. We test a range of alternative measures of ward-specific experience and team familiarity, including the average percentage of shifts staff worked in the ward, presence of joiners and leavers, and percentage of team members staff had worked with before. In all cases, there is no statistically significant impact on patient mortality.[29]

Overall, these results suggest that there is some evidence of increasing returns to seniority within RNs and to firm-specific experience, but no such returns to increasing ward-or team-specific experience. However, as longer tenure may reflect a better quality match between the nurse and the hospital, causality is less clear for firm-specific experience than for our estimates of shortages of qualified compared to unqualified nurses.

5.3.2 Other measures of staff shortages

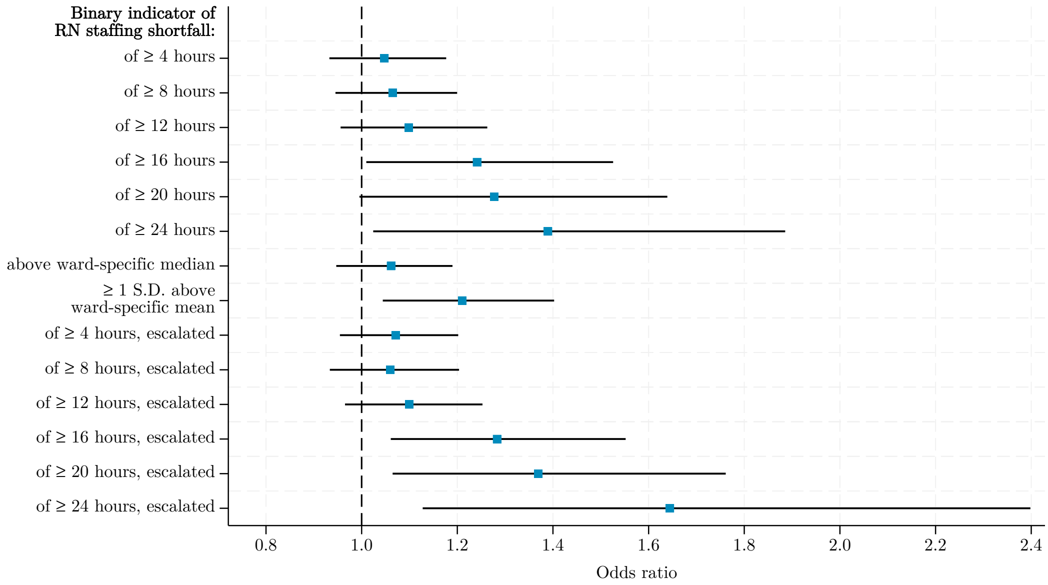
Our main analysis uses the fill-rate as the measure of nurse staffing. Here, we test the robustness of our results to other specifications of the treatment variable. Figure 2 shows the estimated coefficients on a range of binary indicators of a disruption to RN staffing. The first coefficient on the Figure denotes the estimated impact of a team having an

[28] We find very similar results if we additionally control for the seniority mix of RNs – in this specification, increasing the average firm-specific experience among RNs in the team by one year reduces the odds of a patient death by 7.2%. See Appendix Table A8.

[29] These results are available upon request.

RN staffing shortfall of at least four hours; the next five coefficients show the estimated impact of a team having a RN staffing shortfall of at least eight, 12, 16, 20 or 24 hours, respectively. We estimate these regressions separately and control for the planned level of RN staffing.

Figure 2: Team-level analysis: alternative specifications of the treatment variable



Note: Figures denotes estimated odds ratios, with 95% confidence intervals. The dependent variable in each case is a binary indicator equal to one if a patient death occurred. Robust standard errors are clustered at the hospital ward-day level (i.e. team level). RN = Registered Nurse. An 'escalated' staffing shortfall is one where the manager unsuccessfully posted the shift to staff outside of the ward (via the staff bank or an external agency). All regressions include the same set of controls as in the baseline specification in Table 5, as well as controls for Nursing Assistant staffing and the planned number of RN hours. RN staffing shortfall is defined as the difference (in hours) between the planned number of RN hours in a ward on a given day, and the actual number of RN hours worked.

All estimated odds ratios are above one, indicating that a staffing shortfall is associated with an increase in patient mortality, but are only statistically significant at the 5% level for shortfalls of 16 hours and above.[30] A difference between actual and planned RN staffing or at least 16 hours is associated with a 24% increase in the odds of a patient death, rising to 39% for a shortfall of at least 24 hours. We also examine the impact of a shortfall that was 'escalated' by the manager, as defined in Section 5.2, but not

[30]We also test the robustness of our results to a continuous measure of the RN staffing shortfall: the estimated odds ratio is above 1 (1.007) but is not statistically significant (p=0.126). These results are reported in full in Appendix Table A9.

filled. These 'escalations' (as they require effort on the part of the manager) could signal that the manager was particularly keen to fill the shift, perhaps because a particularly important staff member was missing. We find similar, but slightly more pronounced effects, for these shortages, denoted by the final six coefficients of Figure 2. The Figure also shows that teams with an RN staffing shortfall at least one standard deviation above the ward-specific mean have a 21% increase in the odds of a patient death, but having a shortfall above the ward-specific median has a positive but non-statistically significant impact.

These results suggest that our main results are robust to alternative specifications of absence and that the impacts of short-staffing are non-linear: the effect of a registered nurse shortage kicks in when unfilled shifts on a ward amount to more than 12 hours. Larger shortages than this have a monotonically increasing negative effect on patient mortality, though the point estimates are not significantly different from each other.

6 Other outcomes

We consider a number of additional outcomes measures. These allow us to (i) provide indirect evidence of potential pathways for our observed impacts on mortality, and (ii) corroborate that qualified nurse staffing shortages negatively affect patient outcomes.

6.1 Heterogeneity by severity and diagnosis

We first examine whether the impacts vary by the clinical severity of the patient. Finding a relationship could indicate where resource-constrained teams focus their efforts. We measure severity by the Elixhauser score. The median Elixhauser Comorbidity Index score among patients who die in hospital in our sample was 5. We define 'low-severity' patients as patients with an Elixhauser score of less than 5, and 'high-severity' patients as those with those with a score of more than 5.[31]

We then examine whether the mortality impacts vary by patient diagnosis of sepsis. As noted above, this outcome is plausibly highly sensitive to the quantity and quality of nurse staffing. Shortages of nurses would therefore be expected to have greater impacts on the mortality outcomes of patients with sepsis than on those who are in hospital for a condition such as a heart attack or a stroke (where rapid clinical attention is required upon onset but mortality outcomes are less sensitive to subsequent nurse monitoring) or on patients with cancer (for whom deaths typically represent the end of a longer

[31] 3.6% of teams experienced the death of a low-severity patient, and 3.0% experienced the death of a high-severity patient (Table 1). Low-severity patients account for 54.4% of deaths in our sample.

pre-planned treatment pathway and are plausibly less sensitive to nursing care on any given day).

The results are presented in Table 6 and graphically in Figure 3. The first two columns of Table 6 indicate that the impact of a shortage of qualified nurses falls almost entirely on low-severity patients. This suggests our results do not reflect 'harvesting', the bringing forward the deaths of patients that would have otherwise occurred over the next few days. The estimates also suggest that our baseline results are driven by impacts on patients with sepsis. The coefficient on deaths of patients diagnosed with sepsis is statistically significant at the 1% level, and indicates a 1 percentage point increase in the RN fill-rate reduces the odds of a sepsis death by 3.8% (with an odds ratio of 0.962, column 4 of Table 6), meaning that an extra RN working a 12-hour shift reduces the odds of a sepsis death by more than a quarter (28.3%). In contrast, our estimates suggest no impact on deaths of patients who are not diagnosed with sepsis (column 3). Further analysis suggests that there are returns to firm-specific experience for reducing sepsis and non-sepsis mortality, but that the returns are greater for patients diagnosed with sepsis (right-hand panel of Figure 3).

6.2 Other nurse-sensitive outcomes

Figure 3 also presents estimates from the extension of our analysis to look at several other patient outcomes plausibly sensitive to nursing care. First, we examine whether nurse staffing affects the likelihood of a patient being transferred to an Intensive Care Unit (ICU). Staffing and ICU transfer may be linked through two mechanisms: (i) staffing affects quality of care and therefore the probability that patient deteriorates and develops a need for ICU treatment; (ii) staffing determines whether the ward manager feels able to safely care for patients with the highest levels of need. In either case, transfer to ICU is a bad outcome for both the patient and the hospital, as ICU care tends to be more intrusive, more resource intensive, and delays discharge.

We find that higher levels of RN staffing is associated with a reduction in transfers to an ICU; the same is not true of NA staffing. We find no evidence of an impact on transfers to non-ICU wards, consistent with a very limited role for spillovers onto other (non-intensive care) teams, and very high bed occupancy rates in the NHS.[32] Similarly, we find that higher levels of firm-specific experience among RNs is associated with a reduction in transfers to ICUs (right-hand panel of Figure 3), but find no such impact on transfers to non-ICU wards. Finally, we examine the impact on patient falls. We find

[32]We exclude ICUs from the sample in our analysis of patient transfers to ICU and non-ICU wards.

Table 6: Impact of nurse staffing on patient mortality: heterogeneity by patient severity and diagnosis

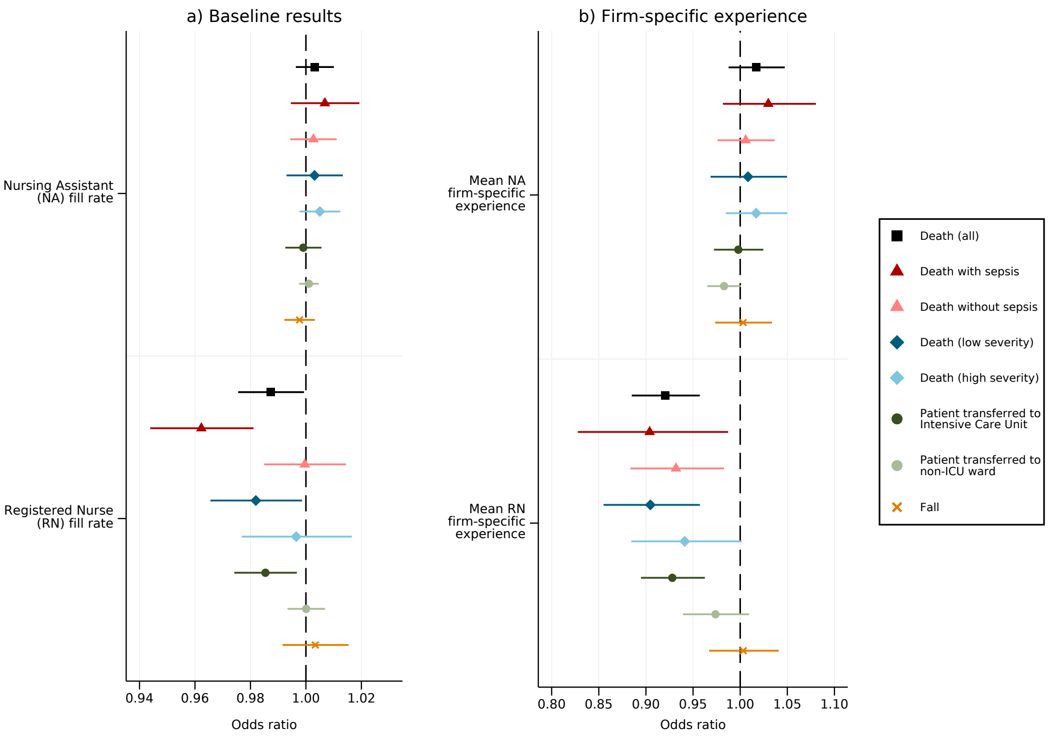
	Outcome:			
	Death of patient classified as:		**Death of patient diagnosed:**	
	Low-severity	High-severity	Without sepsis	With sepsis
	(1)	(2)	(3)	(4)
Nursing Assistant (NA) fill-rate	1.003	1.005	1.003	1.007
	(0.005)	(0.004)	(0.004)	(0.006)
	[0.548]	[0.181]	[0.527]	[0.278]
Registered Nurse (RN) fill-rate	0.982**	0.996	1.000	0.962***
	(0.008)	(0.010)	(0.008)	(0.010)
	[0.034]	[0.730]	[0.956]	[<0.001]
Patient volume and characteristics	✓	✓	✓	✓
Time controls	✓	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓	✓
N	18,567	18,563	18,558	17,840
Clusters	51	51	51	49
Mean of outcome variable	0.036	0.030	0.047	0.018

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Coefficients shown are odds ratios from a fixed-effect logit regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses; p-values are shown in square brackets. The unit of analysis is the hospital ward-day. All regressions control for the mean patient age; the mean patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise interaction terms between age, female share, non-white share and Elixhauser Cormorbidity Index; the mean hospital length of stay of patients treated in the unit that day; the number of patients treated in the unit that day; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday. The median Elixhauser Cormorbidity Index among patients who died in our sample was 5. A low-severity patient is defined as one with an Elixhauser score of less than 5; a high-severity patient as one with an Elixhauser score of 5 or more. A patient is coded as having sepsis if they have a primary or secondary ICD-10 diagnosis code of A400, A403, A409, A410, A411, A414, A415, A418, or A419.

no relationship between the NA and RN fill-rates or the level of firm-specific experience and the number of patient falls.

Figure 3: Impact of nurse staffing and firm-specific experience on other outcomes



Note: Figure denotes estimated odds ratios, with 95% confidence intervals. Robust standard errors are clustered at the hospital ward-day level (i.e. team level). All regressions include the same set of controls as in the baseline specification in Table 5. The results for headline patient mortality in the left-hand and right-hand panel correspond to the results in column 1 of and column 3 of Table 5, respectively. A patient is coded as having sepsis if they have a primary or secondary ICD-10 diagnosis code of A400, A403, A409, A410, A411, A414, A415, A418, or A419. The median Elixhauser Comorbidity Index among patients who died in our sample was 5. A low-severity patient is defined as one with an Elixhauser score of less than 5; a high-severity patient as one with an Elixhauser score of 5 or more. ICU = Intensive Care Unit. ICUs are excluded from the analysis of patients transfers to ICUs and non-ICUs.

7 Analysis at the patient-day level and further robustness checks

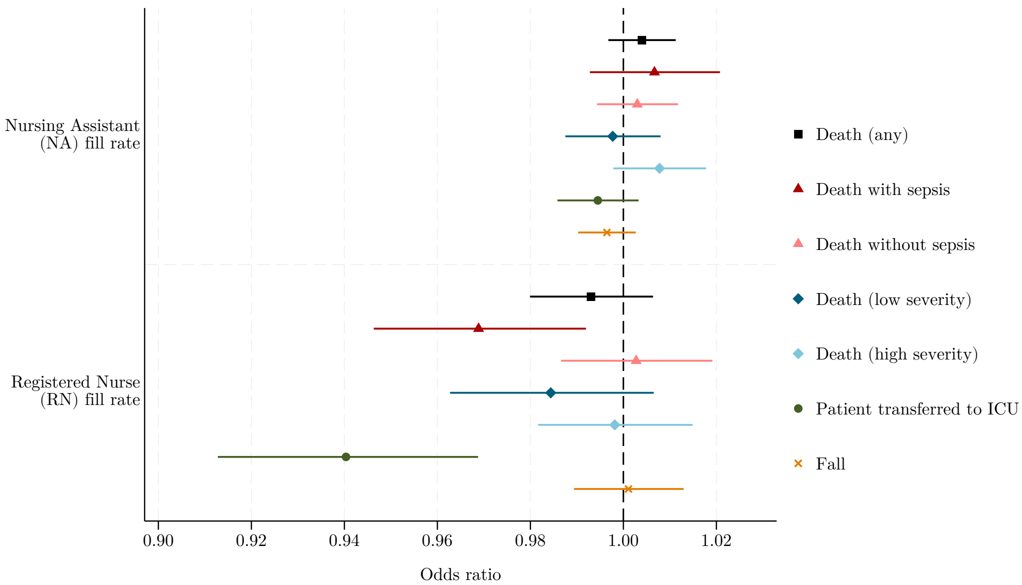
7.1 Analysis at the patient-day level

The key advantages of conducting analysis at the team level are that it mitigates concerns over unobserved patient severity by averaging over all patients under the care of the team and lends greater policy relevance to our results, because staffing is planned and organised at the team level. In our context, because the only variation in staffing across patients is at the ward level, analysis at the patient level will not differ substantively from analysis at ward level provided our controls for patient severity at the ward level are appropriate. To examine this we conduct analysis at the patient-day level. We use the patient-day, rather than the patient-spell, level because the length of a patient spell is itself potentially endogenous to nurse staffing levels.

The staffing level for each patient-day is calculated as the average across the wards in which the patient was treated that day, weighted by the number of minutes spent in each ward. The results are displayed in Figure 4. We find no statistically significant relationship between NA staffing and any of our outcome variables. We find a negative, but not statistically significant, association between the RN fill-rate and headline patient mortality (with an odds ratio of 0.993, as reported in Appendix Table A10, versus 0.987 in our team level analysis). We do, however, find a statistically significant impact on deaths of patients diagnosed with sepsis (with an odds ratio of 0.969 reported in Appendix Table A11, versus 0.962 at the team level) and on the likelihood of being transferred to an ICU, consistent with the team level results described above. The point estimates also suggest bigger effects on patients classified as low-severity than on those classified as high-severity, though these are not statistically significant. We find no evidence of an impact on patient falls.

In sum, the coefficients on RN staffing in our patient-day analysis have the same sign as, and are of similar magnitude to, those estimated in our primary, team-level analysis, but are less precisely estimated. This is what we might expect given that mortality is difficult to predict with observable patient characteristics, and our team-level specification averages unobservable characteristics over all patients under the care of the team, thereby increasing precision.

Figure 4: Patient-day level analysis



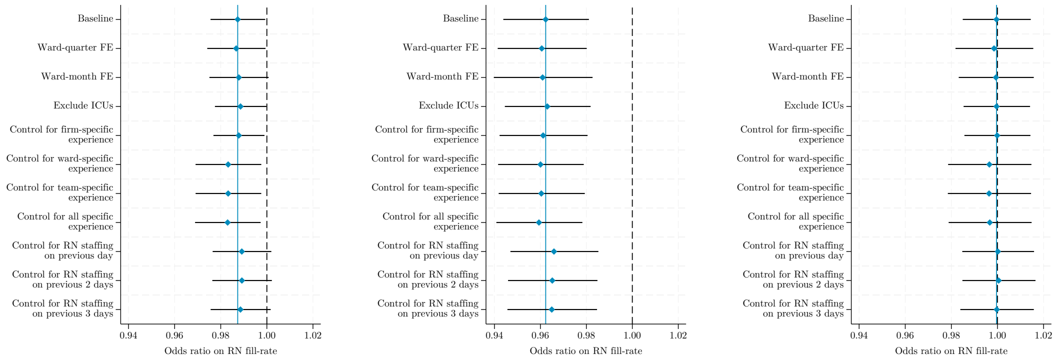
Note: Figure denotes estimated odds ratios, with 95% confidence intervals. Robust standard errors are clustered at the hospital ward-day level (i.e. team level). The unit of analysis is the patient-day. Staffing variables are calculated as the average across the wards the patient was treated in that day, weighted by the number of hours spent in each. All regressions control for patient age; patient age^2; patient age^3; sex; a dummy for being non-white; the patient's Elixhauser Cormorbidity Index; pairwise interaction terms between age, sex, non-white dummy and Elixhauser Cormorbidity Index; a set of dummies for 17 separate diseases, derived from diagnosis codes; the day within the patient's hospital spell; a dummy for each day of the week; a dummy for month of the year; a dummy indicating whether the day was a bank holiday; and a set of dummies for each hospital ward the patient was treated in that day. A patient is coded as having sepsis if they have a primary or secondary ICD-10 diagnosis code of A400, A403, A409, A410, A411, A414, A415, A418, or A419. The median Elixhauser Cormorbidity Index among patients who died in our sample was 5. A low-severity patient is defined as one with an Elixhauser score of less than 5; a high-severity patient as one with an Elixhauser score of 5 or more.

7.2 Further robustness checks

We undertake a further set of robustness tests, and estimate a further set of alternative specifications, and report the results in Figure 5 for all patient mortality, deaths of patients with sepsis, and deaths of patients without sepsis. Our main results condition on separate ward and month fixed effects. Here, we also allow for ward-specific time trends by including ward-month fixed effects and ward-quarter fixed effects. We exclude ICUs from our analysis, owing to their greater average size and higher levels of staffing intensity. We control for firm-specific, ward-specific and team-specific experience alongside the NA and RN fill-rate (separately and together), and we control for RN staffing levels on previous days. In all cases, there is very little change in the estimated coefficient on the RN fill-rate. In a few cases, the estimated impact on all patient mortality ceases to be significant at the 95% level, becoming significant at the 90% level instead, but the overall picture – and the differential impacts on sepsis and non-sepsis deaths – is unchanged.

Figure 5: Robustness analysis

(a) Deaths (all) (b) Deaths with sepsis (c) Deaths without sepsis



Note: Figure denotes estimated odds ratios on the Registered Nurse (RN) fill-rate, with 95% confidence intervals. The 'Baseline' estimate (at the top) uses the specification from column 1 of Table 5, with a full set of controls for patient volumes, patient characteristics, month and day-of-the-week effects, and hospital ward fixed effects. All other estimates are variations on the baseline model. Estimates 2 and 3 use ward-quarter and ward-month fixed effects, respectively, rather than hospital ward fixed effects. Estimate 4 excludes Intensive Care Units (ICUs) from the sample. Estimate 5 additionally controls for mean experience working in the hospital group (firm) among Nursing Assistants (NAs) and RNs. Estimate 6 controls for mean experience working in the ward in the past 90 days among NAs and RNs. Estimate 7 controls for mean experience working with other team members among NAs and RNs. Estimate 8 controls for firm-specific experience (as in Estimate 5), ward-specific experience (as in Estimate 6) and team-specific experience (as in Estimate 7) together. Estimate 9 controls for the RN fill rate on day T-1; Estimate 10 controls for the RN fill rate on T-1 and T-2; and Estimate 11 controls for the RN fill rate on T-1, T-2 and T-3. A patient is coded as having sepsis if they have a primary or secondary ICD-10 diagnosis code of A400, A403, A409, A410, A411, A414, A415, A418, or A419.

Finally, we consider an alternative specification with a different implied team production function. Equation 2 implicitly assumes a Cobb-Douglas production function, where each type of labour – NA and RN, in the baseline model – is additively separable. We consider an alternative model of the form $G(L_{NA}, L_{RN}) = \beta_1 L_{NA} + \beta_2 L_{RN} + g(L_{NA}, L_{RN})$. To capture possible interactions between shortages in the different types of labour we let $g(L_{NA}, L_{RN}) = \gamma_1 L_{NA} L_{RN}$. We find no statistically significant interactions between shortages of the two broad nursing labour types (odds ratio on the interaction term of of 1.000, full results reported in Appendix Table A13).

In a further analysis, given our finding of higher returns to individuals in bands 7 and 8, who are also the ward managers, we allow for the absence of these senior nurse managers to have an additional effect on top of a shortage of (band 5 and 6) qualified nurses, by including an interaction term between the band 5 and 6 RN fill-rate and a dummy indicating the presence of a senior nurse on the ward. The coefficient on this interaction term is close to zero and not statistically significant (results reported in Appendix Table A14). We therefore conclude that production is separable in the different types of nursing labour.

8 Discussion and Conclusions

Our analyses have shown the following. First, nursing teams which have insufficient qualified staff, relative to the managerially-defined safe level for that team, have poorer patient outcomes, while the absence of unqualified nursing assistants has no impact on these outcomes. Second, and perhaps initially surprisingly, these negative effects are concentrated on patients who are less severely ill. Third, these negative effects are concentrated on patients with a diagnosis of sepsis. Fourth, there is some evidence that the effect is non-linear in the number of hours by which the team is short-staffed. Finally, we find some tentative evidence that the absence of the most senior staff has a larger effect on negative patient outcomes.

Our results on the effect of a lack of qualified nurses on patient outcomes corroborate the findings from associative studies in the medical literature or the handful of studies with a causal design, the latter of which were conducted at a much less granular level (Bartel et al., 2014; Friedrich and Hackmann, 2021). Our overall mortality results are comparable to those found in associative studies such as Needleman et al. (2011) who found that the hazard of death increased by 2% for patients exposed to shifts with RN staffing 8 hours or more below target and Griffiths et al. (2019) who found that the hazard rose by 3% for every day patients experienced treatment from a team with RN staffing below the ward mean. Our results indicate (using the patient day results for comparability) that the absence of one staff member for an 8 hour shift would increase the odds of a patient death by 4%.[33]

Importantly, the high frequency of our data means that we can show that these effects occur at the ward-day level, with little spillover to following days. Thus we can show that staffing on the day in which care is delivered matters, rather than simply the absence of staff over the longer periods which have previously been analysed (such as a month or

[33] The odds ratio on the RN fill-rate at the patient-day level is 0.993 (reported in Figure 4 and Appendix Table A10). A one percentage point increase would therefore add approximately 1.4 hours, and an extra staff member working a 8-hour shift would increase staffing by $8/1.4 \approx 5.7$ percentage points of planned RN hours. $0.9993^{5.7} \approx 0.96$, equivalent to a 4% reduction.

a year). We also show that the negative effects primarily kick in at absences which are longer than one shift. This makes our results highly pertinent to staff planning, which operates at at the shift/ward-day level.

In contrast to Bartel et al. (2014), who examine only length of stay, we do not find a significant effect of either ward experience or team familiarity. This may be due to the fact that nurses in our setting move wards quite frequently compared to those in the VA hospitals that Bartel et al. (2014) study. Where nurses move a lot between wards, management may well develop policies which limit the effect of unfamiliarity with the specific ward or other coworkers.

Our finding that many of the additional deaths are for patients suffering from sepsis points to reductions in the quality of monitoring and the timeliness of treatment and escalation to a doctor as key channels through which nurse staffing shortfalls affect patient mortality, and is consistent with the limited existing evidence in this area (Kleinpell et al., 2013; Schorr, 2016; Kleinpell, 2017; Lasater et al., 2021). Our heterogeneity analysis allows us to show that the negative effects on patient outcomes are concentrated in less severe patients. This supports evidence that short-staffed teams ration care by focusing on the most (observably) sick patients (Schubert et al., 2008; Jones et al., 2015; Bail and Grealish, 2016; Hegney et al., 2019). Other, less observably sick, patients then receive less care as a result, increasing their mortality rate. Such rationing has been shown to occur in a variety of healthcare settings, including triage in the Emergency Department (Aacharya et al., 2011), and in hospitals at weekends when staffing levels are reduced (Sutton et al., 2018).

This finding also has implications for the total costs (in terms of lost life) from nursing shortages. Our results do not suggest that shortages of nurses are simply bringing forward deaths of patients that had a high chance of dying in the near future in any case. Instead, deaths are occurring among those who (on observable characteristics at least) were less severely ill and had a lower risk of dying soon, or who might have survived with better monitoring and early detection. This implies a greater overall cost to social welfare. A simple back-of-the-envelope calculation, provided in Appendix section A.4, combines our estimates with the value of a (quality-adjusted) life year, as used by the body in the UK which assesses whether medicines should be provided by the NHS (the National Institute for Health and Care Excellence, NICE), to assess the cost-effectiveness of higher nurse staffing as a mortality-reduction strategy. This suggests that each patient who died in our sample – among whom the median age was 77 – who might have otherwise survived with higher levels of nursing staff, would need to have otherwise lived for a further three or four years for the elimination of all short staffing to be a cost-effective intervention. However, the value of life implicitly used by NICE is quite low (£20,000–30,000): using a higher value comparable to that often used in the USA (100,000 USD) would suggest that this would be a cost-effective strategy if patients who would have died in the absence of the intervention instead go on to live for an additional two or so years.

One limitation of our data is that we cannot directly observe what happens when a team is understaffed. One interpretation of our results is that problems occur when there is too little human capital on the ward. For example, the hospital we study has a protocol for nurse monitoring of patients that states that the first action in the case of patient deterioration is to inform the nurse in charge: if the nurse in charge is too junior, mistakes may be made. More subtly, when someone is missing, tasks will have to be reallocated between the team members who are present. Problems may occur when the tasks that need to reallocated are those usually performed by qualified nurses, because while RNs can perform the tasks usually performed by NAs, the opposite is not true (NICE, 2014a). Discussion with ward managers, nursing leaders and the literature identifies that task reallocation is the job of the most senior staff in the ward (Pinnock, 2012; Clark et al., 2015). More generally, the Royal College of Nursing has highlighted the importance of senior nurses in management of staff (Royal College of Nursing, 2009, 2022). Thus, while the point estimates are not significantly different, the fact that we find that patient outcomes are worse when the most senior qualified nurses are absent supports the idea that tasks are less well done when the most experienced and senior staff on the ward are absent.

Finally, given the current shortage of nurses in many healthcare systems, increasing the supply of qualified nurses will be expensive and take time. Nevertheless, our results do not support the drive to increase the use of nursing assistants in hospital wards (McKinsey, 2014), nor moves towards measures of 'care hours per patient day' that aggregate qualified and non-qualified staff hours into a single measure (Carter, 2016). While they are less expensive and take less time to train, nursing assistants do not appear to be good substitutes for qualified staff. Our results instead suggest that hospitals should undertake greater investment in existing qualified staff, enabling these staff to gain greater hospital experience, and take measures to reduce exits before such experience can be put to use. Alongside this, our results point to the possibility of considerable gains from protocols and training aimed at improving the ability of the broader nursing team to detect patient deterioration at an early stage. The healthcare regulator in England (NICE) has been supporting the roll-out of such protocols in the NHS.[34] While there

[34] https://www.nice.org.uk/advice/mib205/chapter/The-technology

is evidence that adherence to such guidelines and protocols is mixed (e.g. Blencowe et al. (2017); Dierkes et al. (2022)), NICE guidance suggests that such interventions are cost-effective in our NHS setting and thus supports efforts to increase the use of such approaches to address the consequences of nurse shortages.

References

Aacharya, R. P., C. Gastmans, and Y. Denier (2011, December). Emergency department triage: An ethical analysis. BMC Emergency Medicine 11(1).

Agha, L., K. M. Ericson, K. H. Geissler, and J. B. Rebitzer (2021, September). Team Relationships and Performance: Evidence from Healthcare Referral Networks. Management Science, mnsc.2021.4091.

Arcidiacono, P., J. Kinsler, and J. Price (2017, January). Productivity Spillovers in Team Production: Evidence from Professional Basketball. Journal of Labor Economics 35(1), 191–225.

Bail, K. and L. Grealish (2016, November). 'Failure to Maintain': A theoretical proposition for a new quality indicator of nurse care rationing for complex older people in hospital. International Journal of Nursing Studies 63, 146–161.

Ball, J. E., L. Bruyneel, L. H. Aiken, W. Sermeus, D. M. Sloane, A. M. Rafferty, R. Lindqvist, C. Tishelman, and P. Griffiths (2018, February). Post-operative mortality, missed care and nurse staffing in nine countries: A cross-sectional study. International Journal of Nursing Studies 78, 10–15.

Ball, J. E., T. Murrells, A. M. Rafferty, E. Morrow, and P. Griffiths (2014, February). Care left undone' during nursing shifts: Associations with workload and perceived quality of care. BMJ Quality & Safety 23(2), 116–125.

Bandiera, O., I. Barankay, and I. Rasul (2009). Social Connections and Incentives in the Workplace: Evidence From Personnel Data. Econometrica 77(4), 1047–1094.

Bandiera, O., I. Barankay, and I. Rasul (2013). Team Incentives: Evidence from a Firm Level Experiment. Journal of the European Economic Association 11(5), 1079–1114.

Bapoje, S. R., J. L. Gaudiani, V. Narayanan, and R. K. Albert (2011, Feb). Unplanned transfers to a medical intensive care unit: causes and relationship to preventable errors in care. Journal of Hospital Medicine 6(2), 68–72.

Bartel, A. P., N. D. Beaulieu, C. S. Phibbs, and P. W. Stone (2014, April). Human Capital and Productivity in a Team Environment: Evidence from the Healthcare Sector. American Economic Journal: Applied Economics 6(2), 231–259.

Becker, G. S. (1964). Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition. NBER.

Blencowe, N. S., S. Strong, UK National Surgical Research Collaborative, et al. (2017, January). Multicentre observational study of adherence to Sepsis Six guidelines in emergency general surgery. British Journal of Surgery 104(2), e165–e171.

Boniol, M., T. Kunjumen, T. S. Nair, A. Siyam, J. Campbell, and K. Diallo (2022, June). The global health workforce stock and distribution in 2020 and 2030: A threat to equity and 'universal' health coverage? BMJ Global Health 7(6), e009316.

Burgess, S., C. Propper, M. Ratto, S. v. H. Kessler Scholder, and E. Tominey (2010, September). Smarter Task Assignment or Greater Effort: The Impact of Incentives on Team Performance. The Economic Journal 120(547), 968–989.

Burgess, S., C. Propper, M. Ratto, and E. Tominey (2017, October). Incentives in the Public Sector: Evidence from a Government Agency. The Economic Journal 127(605), F117–F141.

Carter, A. V., K. T. Schnepel, and D. G. Steigerwald (2017, October). Asymptotic Behavior of a t -Test Robust to Cluster Heterogeneity. The Review of Economics and Statistics 99(4), 698–709.

Carter, P. (2016). Operational Productivity and Performance in English NHS Acute Hospitals: Unwarranted Variations. An Independent Report for the Department of Health by Lord Carter of Coles. Department of Health.

Centre for Disease Control and Protection (2023). What is Sepsis?

Chan, D. C., M. Gentzkow, and C. Yu (2022, April). Selection with Variation in Diagnostic Skill: Evidence from Radiologists. The Quarterly Journal of Economics 137(2), 729–783.

Chan, David C, J. and Y. Chen (2022, October). The Productivity of Professions: Evidence from the Emergency Department. Working Paper 30608, National Bureau of Economic Research.

Chen, Y. (2021, December). Team-Specific Human Capital and Team Performance: Evidence from Doctors. American Economic Review 111(12), 3923–3962.

Clark, A., P. Moule, A. Topping, and M. Serpell (2015, May). Rescheduling nursing shifts: Scoping the challenge and examining the potential of mathematical model based tools. Journal of Nursing Management 23(4), 411–420.

Cooper, Z., J. Doyle, J. Graves, and J. Gruber (2022, February). Do Higher-Priced Hospitals Deliver Higher-Quality Care? Technical Report w29809, National Bureau of Economic Research, Cambridge, MA.

Cornelissen, T., C. Dustmann, and U. Schönberg (2017, February). Peer Effects in the Workplace. American Economic Review 107(2), 425–456.

Currie, J. and W. B. MacLeod (2017). Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians. Journal of Labor Economics 35(1), 18977.

Currie, J. M. and W. B. MacLeod (2020). Understanding Doctor Decision Making: The Case of Depression Treatment. Econometrica 88(3), 847–878.

Dall'Ora, C., A. Maruotti, and P. Griffiths (2020, March). Temporary Staffing and Patient Death in Acute Care Hospitals: A Retrospective Longitudinal Study. Journal of Nursing Scholarship 52(2), 210–216.

Dierkes, A. M., L. H. Aiken, D. M. Sloane, J. P. Cimiotti, K. A. Riman, and M. D. McHugh (2022). Hospital nurse staffing and sepsis protocol compliance and outcomes among patients with sepsis in the usa: a multistate cross-sectional analysis. BMJ Open 12(3), e056802.

Donaldson, N., L. B. Bolton, C. Aydin, D. Brown, J. D. Elashoff, and M. Sandhu (2005, August). Impact of California's Licensed Nurse-Patient Ratios on Unit-Level Nurse Staffing and Patient Outcomes. Policy, Politics, & Nursing Practice 6(3), 198–210.

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

Dykes, P. C., D. L. Carroll, A. C. Hurley, A. Benoit, and B. Middleton (2009, Jun). Why do patients in acute care hospitals fall? can falls be prevented? Journal of Nursing Administration 39(6), 299–304.

Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey (1998, January). Comorbidity Measures for Use with Administrative Data. Medical Care 36(1), 8–27.

Fischer, K., J. J. Reade, and W. B. Schmal (2021). The Long Shadow of an Infection: COVID-19 and Performance at Work. Working Paper.

Friebel, G., M. Heinz, M. Krueger, and N. Zubanov (2017, August). Team Incentives and Performance: Evidence from a Retail Chain. American Economic Review 107(8), 2168–2203.

Friedrich, B. U. and M. B. Hackmann (2021, October). The Returns to Nursing: Evidence from a Parental-Leave Program. The Review of Economic Studies 88(5), 2308–2343.

Gennaioli, N., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer (2013, February). Human Capital and Regional Development. The Quarterly Journal of Economics 128(1), 105–164.

Griffiths, P., J. Ball, J. Drennan, C. Dall'Ora, J. Jones, A. Maruotti, C. Pope, A. Recio Saucedo, and M. Simon (2016, November). Nurse staffing and patient outcomes: Strengths and limitations of the evidence to inform policy and practice. A review and discussion paper based on evidence reviewed for the National Institute for Health and Care Excellence Safe Staffing guideline development. International Journal of Nursing Studies 63, 213–225.

Griffiths, P., A. Maruotti, A. Recio Saucedo, O. C. Redfern, J. E. Ball, J. Briggs, C. Dall'Ora, P. E. Schmidt, and G. B. Smith (2019, August). Nurse staffing, nursing assistants and hospital mortality: Retrospective longitudinal cohort study. BMJ Quality & Safety 28(8), 609–617.

Griffiths, P., C. Saville, J. Ball, et al. (2020, Mar). The Safer Nursing Care Tool as a guide to nurse staffing requirements on hospital wards: observational and modelling study. Number 8.16 in Health Services and Delivery Research. Southampton, UK: NIHR Journals Library.

Gruber, J. and S. A. Kleiner (2012, February). Do Strikes Kill? Evidence from New York State. American Economic Journal: Economic Policy 4(1), 127–157.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009, September). Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. American Economic Journal: Applied Economics 1(4), 34–68.

Hamilton, B. H., J. A. Nickerson, and H. Owan (2003, June). Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation. Journal of Political Economy 111(3), 465–497.

Health Education England (2023). Sepsis Awareness.

Hegney, D. G., C. S. Rees, R. Osseiran-Moisson, L. Breen, R. Eley, C. Windsor, and C. Harvey (2019, March). Perceptions of nursing workloads and contributing factors, and their impact on implicit care rationing: A Queensland, Australia study. Journal of Nursing Management 27(2), 371–380.

Herrmann, M. A. and J. E. Rockoff (2012, October). Worker Absence and Productivity: Evidence from Teaching. Journal of Labor Economics 30(4), 749–782.

Hoey, S., T. Peeters, and J. van Ours (2022). The Impact of Absent Coworkers on Productivity in Teams. SSRN Electronic Journal.

Huckman, R. S. and G. P. Pisano (2006, April). The Firm Specificity of Individual Performance: Evidence from Cardiac Surgery. Management Science 52(4), 473–488.

Inada-Kim, M. (2022, November). NEWS2 and improving outcomes from sepsis. Clinical Medicine 22(6), 514–517.

Jarosch, G., E. Oberfield, and E. Rossi-Hansberg (2021). Learning From Coworkers. Econometrica 89(2), 647–676.

Jones, B. F. (2021, May). The Rise of Research Teams: Benefits and Costs in Economics. Journal of Economic Perspectives 35(2), 191–216.

Jones, T. L., P. Hamilton, and N. Murry (2015, June). Unfinished nursing care, missed care, and implicitly rationed care: State of the science review. International Journal of Nursing Studies 52(6), 1121–1137.

Kalisch, B. J., D. Tschannen, and K. H. Lee (2012, January). Missed Nursing Care, Staffing, and Patient Falls. Journal of Nursing Care Quality 27(1), 6–12.

Kleinpell, R. (2017). Promoting early identification of sepsis in hospitalized patients with nurse-led protocols. Critical Care 21(10), 10.

Kleinpell, R., L. Aitken, and C. A. Schorr (2013, May). Implications of the New International Sepsis Guidelines for Nursing Care. American Journal of Critical Care 22(3), 212–222.

Lasater, K. B., D. M. Sloane, M. D. McHugh, J. P. Cimiotti, K. A. Riman, B. Martin, M. Alexander, and L. H. Aiken (2021). Evaluation of hospital nurse-to-patient staffing ratios and sepsis bundles on patient outcomes. American Journal of Infection Control 49(7), 868–873.

Lee, C. H. and D. G. Steigerwald (2018, June). Inference for Clustered Data. The Stata Journal: Promoting communications on statistics and Stata 18(2), 447–460.

Lin, H. (2014). Revisiting the relationship between nurse staffing and quality of care in nursing homes: An instrumental variables approach. Journal of Health Economics, 12.

Marquet, Kristel RN, M. P. s., P. e. Claes, Neree MD, E. M. De Troy, G. M. Kox, M. M. Droogmans, P. Schrooten, Ward MD, P. Weekers, Frank MD, P. Vlayen, Annemie MSc, P. Vandersteen, Marjan MD, and P. Vleugels, Arthur MD (2015, May). One fourth of unplanned transfers to a higher level of care are associated with a highly preventable adverse event: A patient record review in six belgian hospitals. Critical Care Medicine 43(5), 1053–1061.

Mas, A. and E. Moretti (2009, February). Peers at Work. American Economic Review 99(1), 112–145.

McKinsey (2014). Optimizing the nursing skill mix: A win for nurses, patients, and hospitals. https://www.mckinsey.com/industries/healthcare/our-insights/optimizing-the-nursing-skill-mix–a-win-for-nurses-patients-and-hospitals.

Menendez, M. E., V. Neuhaus, C. N. van Dijk, and D. Ring (2014, September). The Elixhauser comorbidity method outperforms the Charlson index in predicting inpatient death after orthopaedic surgery. Clinical Orthopaedics and Related Research 472(9), 2878–2886.

Mincer, J. A. (1974). Schooling, Experience, and Earnings. NBER.

Moretti, E. (2004, May). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. American Economic Review 94(3), 656–690.

Morris, M. E., K. Webster, C. Jones, A.-M. Hill, T. Haines, S. McPhail, D. Kiegaldie, S. Slade, D. Jazayeri, H. Heng, R. Shorr, L. Carey, A. Barker, and I. Cameron (2022, May). Interventions to reduce falls in hospitals: a systematic review and meta-analysis. Age and Ageing 51(5), afac077.

Morris, R. and S. O'Riordan (2017, Jul). Prevention of falls in hospital. Clinical Medicine 17(4), 360–362.

Needleman, J., C. L. Leibson, and M. Harris (2011). Nurse Staffing and Inpatient Hospital Mortality. The New England Journal of Medicine, 10.

NHS Digital (2023). NHS Workforce Statistics - September 2023.

NHS England (2023). NHS England - National Early Warning Score.

NICE (2014a, July). Safe staffing for nursing in adult inpatient wards in acute hospitals: example scenario. National Institute for Health and Care Excellence.

NICE (2014b, July). Safe staffing for nursing in adult inpatient wards in acute hospitals: safe staffing guideline. National Institute for Health and Care Excellence.

NICE (2016). Sepsis: Recognition, Diagnosis and Early Management. National Institute for Health and Care Excellence.

Patrician, P. A., L. Loan, M. McCarthy, M. Fridman, N. Donaldson, M. Bingham, and L. R. Brosch (2011, February). The Association of Shift-Level Nurse Staffing With Adverse Patient Events. JONA: The Journal of Nursing Administration 41(2), 64–70.

Peutere, L., J. Pentti, A. Ropponen, M. Kivimäki, M. Härmä, O. Krutova, J. Ervasti, A. Koskinen, and M. Virtanen (2024, February). Association of nurse understaffing and limited nursing work experience with in-hospital mortality among patients: A longitudinal register-based study. International Journal of Nursing Studies 150, 104628.

Pinnock, D. (2012, October). The role of the ward manager in promoting patient safety. British Journal of Nursing 21(19), 1144–1149.

Propper, C. and J. Van Reenen (2010, April). Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labor Markets on Hospital Performance. Journal of Political Economy 118(2), 222–273.

Quan, H., V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J.-C. Luthi, L. D. Saunders, C. A. Beck, T. E. Feasby, and W. A. Ghali (2005, November). Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data:. Medical Care 43(11), 1130–1139.

Redfern, O. C., P. Griffiths, A. Maruotti, A. Recio Saucedo, and G. B. Smith (2019, September). The association between nurse staffing levels and the timeliness of vital signs monitoring: A retrospective observational study in the UK. BMJ Open 9(9), e032157.

Royal College of Nursing (2009). Breaking down barriers, driving up standards: the role of the ward sister and charge nurse. Technical report, Royal College of Nursing.

Royal College of Nursing (2022). The importance of being supernumerary. Technical report, Royal College of Nursing.

Royal College of Nursing (2023, November). Sepsis Subject Guide. Technical report, Royal College of Nursing.

Scheffler, R. M. and D. R. Arnold (2019, April). Projecting shortages and surpluses of doctors and nurses in the OECD: What looms ahead. Health Economics, Policy and Law 14(2), 274–290.

Schorr, C. (2016). Nurses can help improve outcomes in severe sepsis: from early screening to timely interventions, nurses can save lives. American Nurse Today 11(3), 20–25.

Schubert, M., T. R. Glass, S. P. Clarke, L. H. Aiken, B. Schaffert-Witvliet, D. M. Sloane, and S. De Geest (2008, April). Rationing of nursing care and its relationship to patient outcomes: The Swiss extension of the International Hospital Outcomes Study. International Journal for Quality in Health Care 20(4), 227–237.

Shelford Group (2015). Safer nursing care tool implementation resource pack. Technical report, The Shelford Group.

Silver, D. (2021, May). Haste or Waste? Peer Pressure and Productivity in the Emergency Department. The Review of Economic Studies 88(3), 1385–1417.

Smith, G. B., O. Redfern, A. Maruotti, A. Recio-Saucedo, and P. Griffiths (2020, April). The association between nurse staffing levels and a failure to respond to patients with deranged physiology: A retrospective observational study in the UK. Resuscitation 149, 202–208.

Southern, D. A., H. Quan, and W. A. Ghali (2004, April). Comparison of the Elixhauser and Charlson/Deyo Methods of Comorbidity Measurement in Administrative Data. Medical Care 42(4), 355–360.

Stuart, H. C. (2017, April). Structural Disruption, Relational Experimentation, and Performance in Professional Hockey Teams: A Network Perspective on Member Change. Organization Science 28(2), 283–300.

Sutton, E., J. Bion, C. Aldridge, A. Boyal, J. Willars, and C. Tarrant (2018, December). Quality and safety of in-hospital care for acute medical patients at weekends: A qualitative study. BMC Health Services Research 18(1).

Torsvik, M., L. T. Gustad, A. Mehl, I. L. Bangstad, L. J. Vinje, J. K. Damås, and E. Solligård (2016, December). Early identification of sepsis in hospital inpatients by ward nurses increases 30-day survival. Critical Care 20(1).

Van Walraven, C., P. C. Austin, A. Jennings, H. Quan, and A. J. Forster (2009, June). A Modification of the Elixhauser Comorbidity Measures Into a Point System for Hospital Death Using Administrative Data. Medical Care 47(6), 626–633.

WHO (2020). State of the world's nursing 2020: investing in education, jobs and leadership. World Health Organisation.

A Appendix

A.1 Summary statistics

Table A1: Summary statistics: within-ward and between-ward standard deviation (SD)

	Mean	SD	Within SD	Between SD
Staffing quantity and seniority:				
Overall fill-rate	95.7	5.4	4.9	2.2
NA fill-rate	94.8	10.9	10.3	3.8
RN fill-rate	96.4	5.8	5.3	2.3
Band 5 RN fill-rate	47.1	11.5	8.2	8.1
Band 6 RN fill-rate	17.0	12.5	6.3	10.8
Band 7 and 8 RN fill-rate	3.8	4.5	3.3	3.1
Shortfall relative to planned RN staffing (hours)	4.8	7.6	6.8	3.5
Firm-specific experience:				
Mean years of experience in the hospital group (NAs)	5.6	3.2	2.1	2.4
Mean years of experience in the hospital group (RNs)	6.7	2.5	1.5	2.0
Ward-specific experience:				
Mean shifts in the ward in the past 90 days (all)	30.3	5.2	4.2	3.2
Mean shifts in the ward in the past 90 days (NAs)	27.8	9.7	8.1	7.1
Mean shifts in the ward in the past 90 days (RNs)	31.5	5.9	4.7	3.5
Team-specific experience:				
Mean % of team members worked with 10+ times before (NAs)	0.53	0.25	0.21	0.17
Mean % of team members worked with 10+ times before (RNs)	0.59	0.20	0.16	0.12
Patient outcomes:				
Patient death (binary)	0.06	0.24	0.24	0.06
Death rate per 1,000 patients	3.58	16.20	15.67	4.15

Figure A1: Distribution of time of recorded patient death

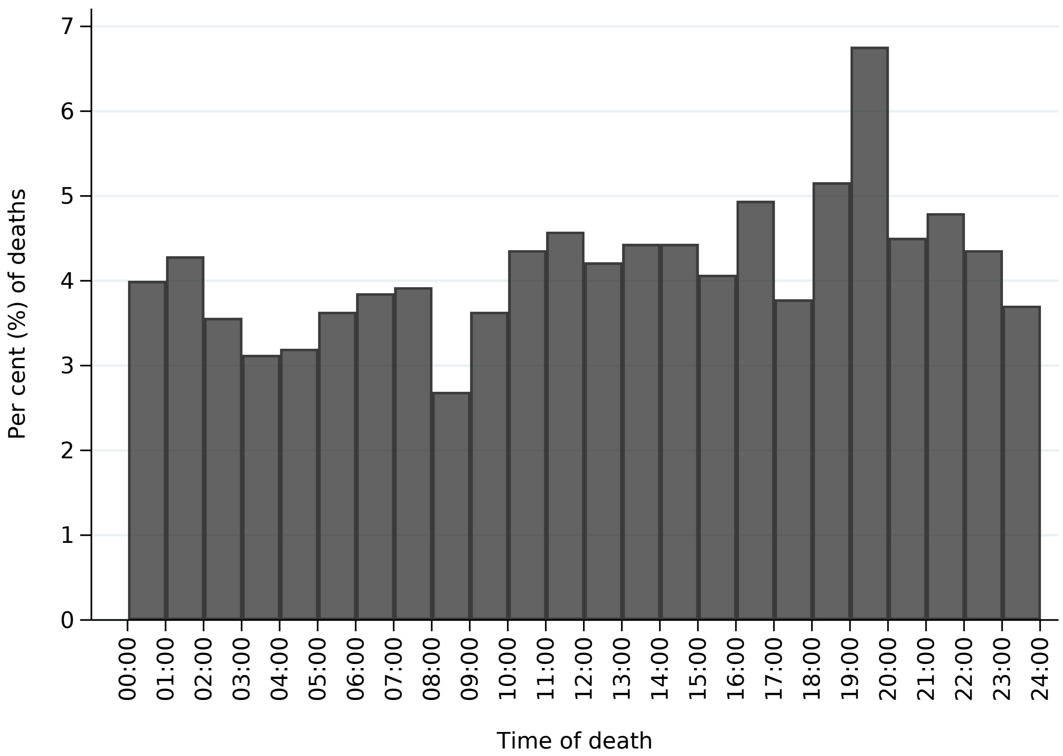
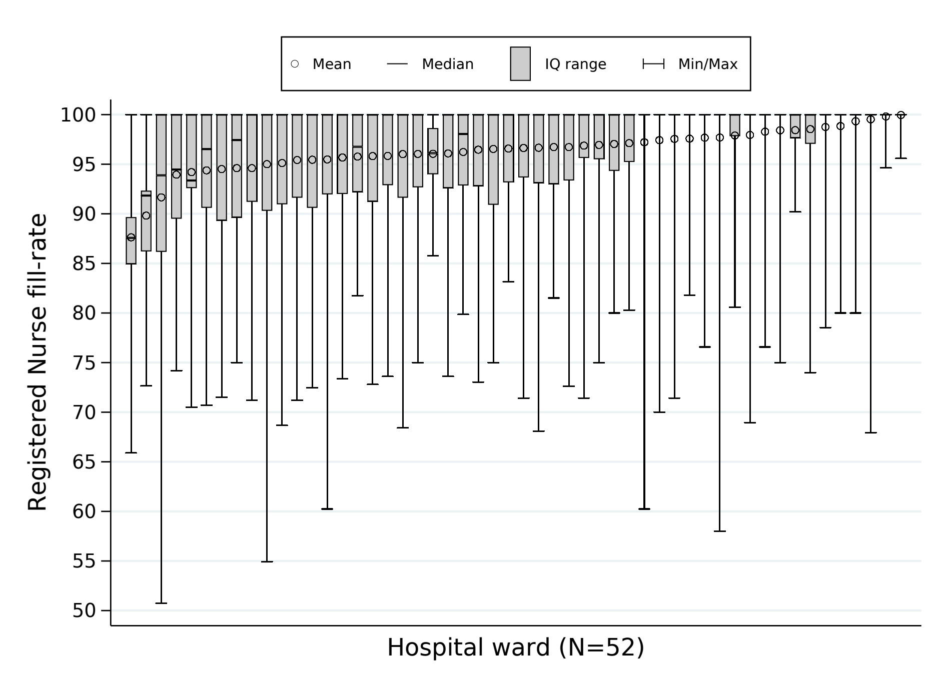
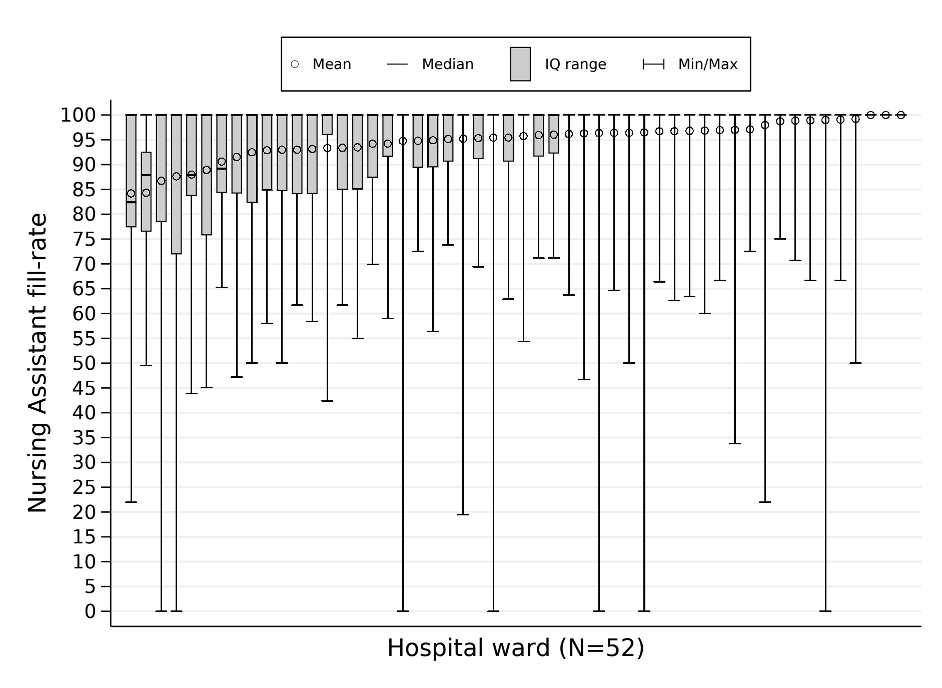


Figure A2: Distribution of and variation in fill-rate across wards

(a) Registered Nurses



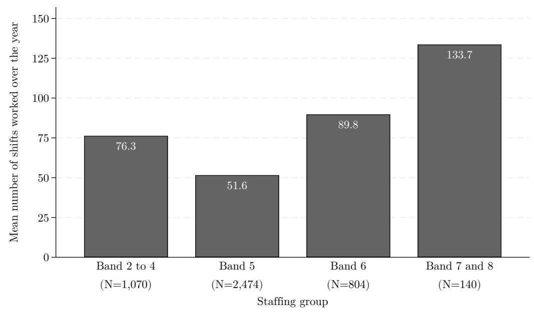
(b) Nursing Assistants



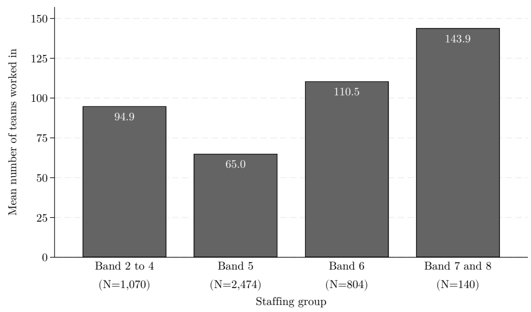
Note: Figures are calculated with respect to the 2017 calendar year (the period covered by our sample). The ordering of wards differs between panel a) and panel b).

Figure A3: Staff movements across teams, wards and hospitals

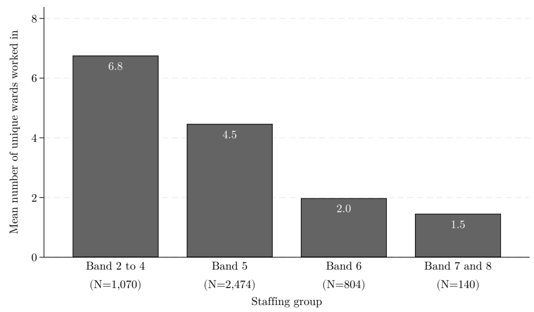
(a) Number of shifts (mean)



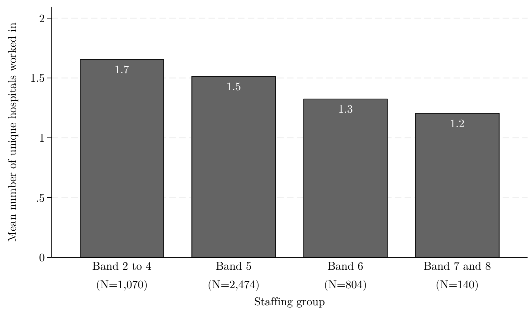
(b) Unique teams (mean)



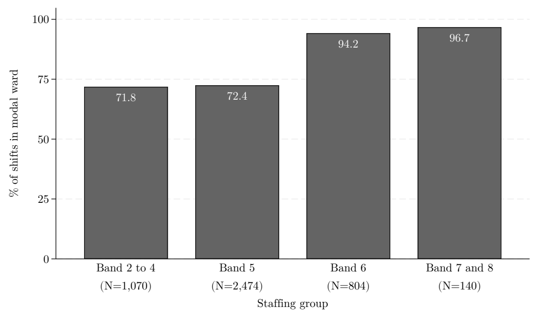
(c) Unique wards (mean)



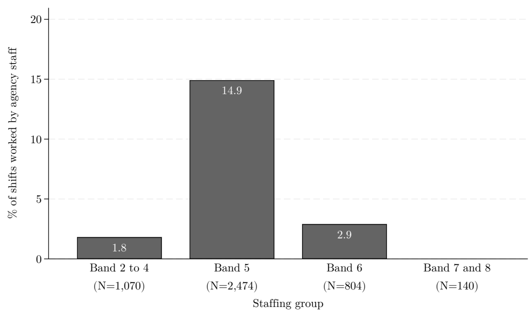
(d) Unique hospitals (mean)



(e) Share (%) of shifts in usual (modal) ward



(f) Share (%) of shifts worked by agency staff



Note: Figures are calculated with respect to the 2017 calendar year (the period covered by our sample).

Table A2: Summary statistics at the patient-day level

	Mean	SD
Staffing		
Fill-rate (all)	95.6	5.1
Nursing Assistant (NA) fill-rate	94.6	10.1
Registered Nurse (RN) fill-rate	96.2	5.6
Planned NA hours	59.9	29.7
Shortfall relative to planned NA staffing (hours)	3.6	6.3
Planned RN hours	145.2	83.8
Shortfall relative to planned RN staffing (hours)	5.4	7.7
RN hourly shortfall $\geq$ 4 hours (binary)	0.41	0.49
RN hourly shortfall $\geq$ 8 hours (binary)	0.38	0.49
RN hourly shortfall $\geq$ 12 hours (binary)	0.26	0.44
RN hourly shortfall $\geq$ 16 hours (binary)	0.12	0.33
RN hourly shortfall $\geq$ 20 hours (binary)	0.08	0.28
RN hourly shortfall $\geq$ 24 hours (binary)	0.04	0.20
Patient characteristics		
Age	64.49	19.09
Female (%)	0.47	0.50
Non-White (%) †	0.51	0.50
Elixhauser Comorbidity Index score	3.49	2.51
Diagnosed with sepsis (%)*	0.10	0.30
Length of hospital stay (days)‡	21.6	29.8
Patient outcomes		
Death	0.0040	0.0630
Transferred to Intensive Care Unit	0.0041	0.0639
Fall	0.0038	0.0618
Observations	276,237	

Note: Unit of analysis is the patient-day. Staffing variables are weighted by the hours spent in the ward on the day in question.
† Non-white includes those patients for whom ethnicity was recorded as unknown.
‡ Calculated as the average across all patient days, rather than the average for all patients.
* A patient is coded as having sepsis if they have an ICD-10 diagnosis code of A400, A403, A409, A410, A411, A414, A415, A418, or A419. A patient is coded as having pneumonia if they have a ICD-10 diagnosis of J180, J181, J188 or J189.

A.2 Definitions

Fill rates

$$FillRateNA_{it} = 100 * \frac{\text{Hours worked by NAs in ward } i \text{ on day } t}{\text{Planned NA hours in ward } i \text{ on day } t}$$

$$FillRateRN_{it} = 100 * \frac{\text{Hours worked by RNs in ward } i \text{ on day } t}{\text{Planned RN hours in ward } i \text{ on day } t}$$

$$FillRateRN_{it}^{Band5} = 100 * \frac{\text{Hours worked by Band 5 RNs in ward } i \text{ on day } t}{\text{Planned RN hours in ward } i \text{ on day } t}$$

$$FillRateRN_{it}^{Band6} = 100 * \frac{\text{Hours worked by Band 6 RNs in ward } i \text{ on day } t}{\text{Planned RN hours in ward } i \text{ on day } t}$$

$$FillRateRN_{it}^{Band7and8} = 100 * \frac{\text{Hours worked by Band 7 and 8 RNs in ward } i \text{ on day } t}{\text{Planned RN hours in ward } i \text{ on day } t}$$

Firm-specific experience

Let T_{kt} be a team consisting of n members working in location (hospital ward) k on day t. $|T_{kt}|$ is the number of staff in the team, which is not fixed and varies across both wards and time. X_{iht} is the amount of time (in years) that individual i has been employed in the NHS Trust at date t. We then define:

$$\text{Average experience in the hospital group} \frac{1}{|T_{kt}|} \sum_{i \in T_{kt}} X_{iht}$$

$$\text{Average experience in the hospital group for RNs} = \frac{1}{|T_{kt}^{RN}|} \sum_{i \in T_{kt}^{RN}} X_{iht}$$

$$\text{Average experience in the hospital group for NAs} = \frac{1}{|T_{kt}^{NA}|} \sum_{i \in T_{kt}^{NA}} X_{iht}$$

where T_{kt} is the team working in ward k on day t, T_{kt}^{RN} is the set of RNs working in ward k on day t, and T_{kt}^{NA} is the set of NAs working in ward k on day t.

Ward-specific experience and familiarity

Let T_{kt} be the team working in ward k on day t, T_{kt}^{RN} be the set of RNs working in ward k on day t, and T_{kt}^{NA} be the set of NAs working in ward k on day t. Let $K(h)$ be the set of wards in hospital h. Then for all $k \in K(h)$ we define Y_{ihkt} as the ward-specific experience for worker i in ward k on day t as:

$$Y_{ihkt} = \sum_{\tau=t-90}^{t-1} \mathbb{1}(i \in T_{k\tau}) \quad \forall k \in K(h)$$

where $\mathbb{1}(i \in T_{k\tau})$ is an indicator equal to one if individual i was part of team $T_{k\tau}$ and zero otherwise. Y_{ihkt} therefore represents the number of times the individual has worked in ward k in the previous 90 days. We then further define:

$$\text{Mean shifts in ward in the past 90 days} = \frac{1}{|T_{kt}|} \sum_{i \in T_{kt}} Y_{ihkt}$$

$$\text{Mean NA shifts in ward in the past 90 days} = \frac{1}{|T_{kt}^{NA}|} \sum_{i \in T_{kt}^{NA}} Y_{ihkt}$$

$$\text{Mean RN shifts in ward in the past 90 days} = \frac{1}{|T_{kt}^{RN}|} \sum_{i \in T_{kt}^{RN}} Y_{ihkt}$$

Team-specific experience and familiarity

Let T_{kt} be the team working in ward k on day t, T_{kt}^{RN} be the set of RNs working in ward k on day t, and T_{kt}^{NA} be the set of NAs working in ward k on day t. Let $K(h)$ be the set of wards in hospital h. We define team-specific experience, Z_{-ihkt} as the average number of shifts that individual i has worked together with their team mates, in any ward, over the past 90 days prior to day t. Let $E(i, j, t)$ represent the (recent) shared work experience between workers i and j, as of date t, where:

$$E(i, j, t) = \sum_{k \in K(h)} \sum_{\tau = t-90}^{t-1} \mathbb{1}(i \in T_{k\tau} \wedge j \in T_{k\tau})$$

where $\mathbb{1}(i \in T_{k\tau} \wedge j \in T_{k\tau})$ is an indicator that equals one if both workers i and j worked in ward k on day $\tau \in [t-90, t-1]$ (i.e. equals one if both i and j were in team $T_{k\tau}$). $E(i, j, t)$ therefore measures the number of shifts i and j have worked together. We then define an individual's average shared work experience with fellow team members as:

$$Z_{-ihkt} = \frac{1}{|T_{kt}| - 1} \cdot \sum_{j \in T_{kt}, j \neq i} E(i, j, t)$$

We then define:

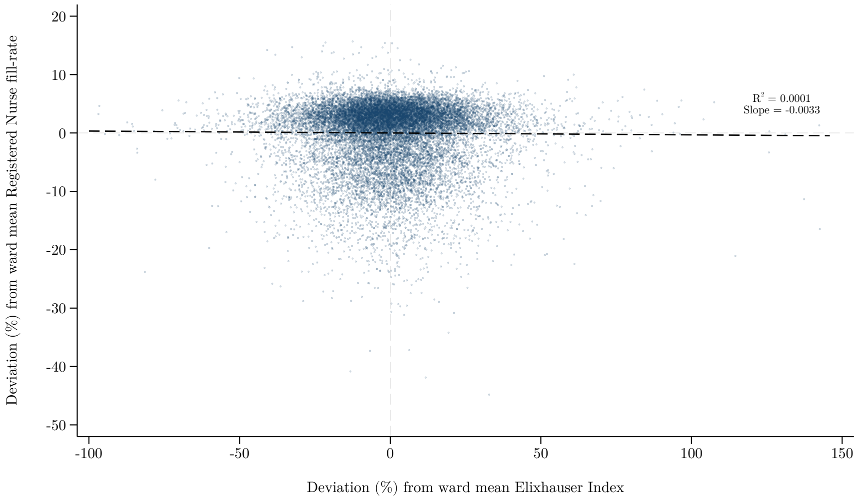
$$\text{Mean NA shifts with other team members in past 90 days} = \frac{1}{|T_{kt}^{NA}|} \sum_{i \in T_{kt}^{NA}} Z_{-ihkt}$$

$$\text{Mean RN shifts with other team members in past 90 days} = \frac{1}{|T_{kt}^{RN}|} \sum_{i \in T_{kt}^{RN}} Z_{-ihkt}$$

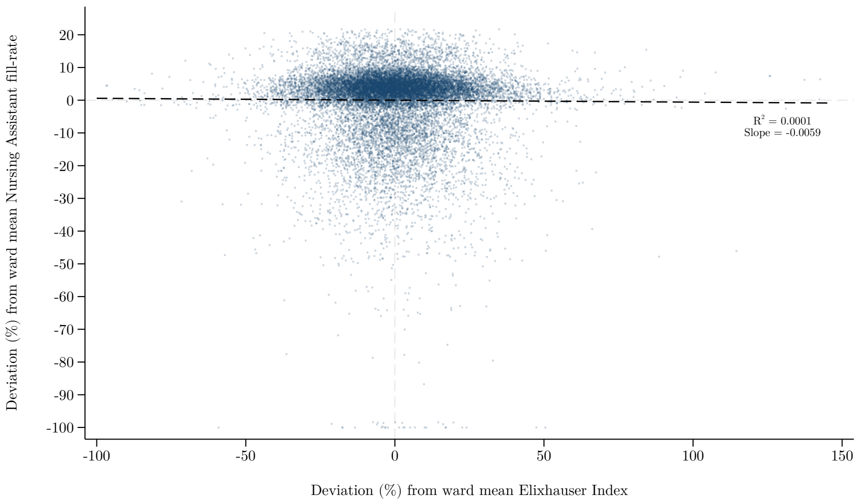
A.3 Additional results and figures

Figure A4: Relationship between staffing levels and observable patient comorbidity, as measured by Elixhauser Cormorbidity Index

(a) Registered Nurses

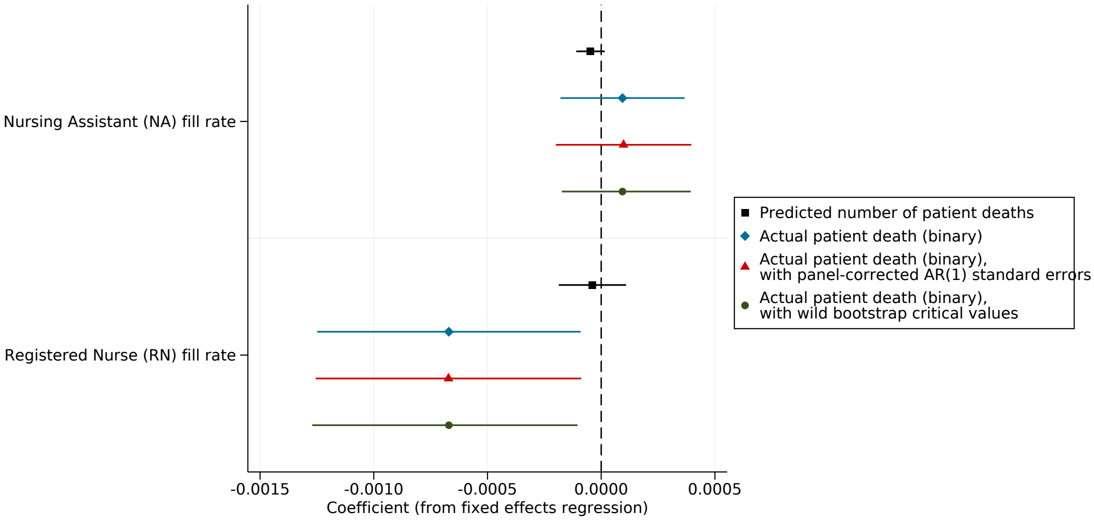


(b) Nursing Assistants



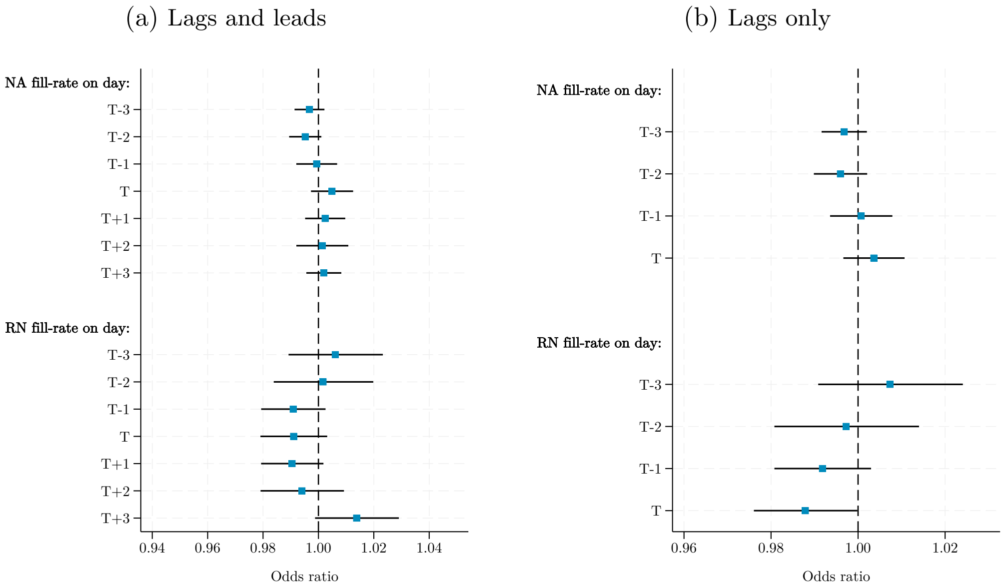
Note: each point denotes an observation (i.e. a team, defined as a hospital ward-day). Values are calculated relative to the unit ward, as our baseline specification includes ward fixed effects, and so all results are identified from within-ward variation.

Figure A5: Impact of RN staffing on predicted and actual mortality



Note: Coefficients shown are from a fixed-effects OLS regression with hospital ward fixed effects (rather than a logit model, to allow for a continuous predicted mortality measure). In the first two specifications, robust standard errors are clustered at the hospital ward level. In the third, panel-corrected AR(1) standard errors are estimated. In the fourth, wild bootstrap critical values and confidence intervals (generated with 999 replications) are calculated. Predicted mortality is calculated at the patient level and then aggregated up to the team (hospital ward-day) level. Values are between 0 and 1, and are predicted using patient age; age squared; age cubed; sex; ethnicity; Elixhauser Cormorbidity Index; pairwise interactions between each of these; dummy variables for 17 separate diseases; how long the patient had been in hospital up to that point; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday. All regressions include the same set of controls as in the baseline specification in Table 5.

Figure A6: Dynamic effects on patient mortality



Note: Both panels display the results from a logit model with ward fixed effects. Both regressions include the same set of controls as detailed in the footnote to Table 5.

Table A3: Relationship between patient severity and team composition

	Outcome: mean patient Elixhauser Cormorbidity Index score				
	(1)	(2)	(3)	(4)	(5)
Nursing Assistant (NA) fill-rate	-0.001 (0.001) [0.270]	-0.001 (0.001) [0.278]	-0.001 (0.001) [0.146]	-0.002* (0.001) [0.060]	-0.002* (0.001) [0.060]
Registered Nurse (RN) fill-rate	-0.000 (0.001) [0.784]		-0.001 (0.002) [0.684]	-0.000 (0.002) [0.933]	-0.000 (0.002) [0.921]
RN fill-rate, of which:					
Band 5 nurses		-0.000 (0.001) [0.732]			
Band 6 nurses		-0.001 (0.002) [0.497]			
Band 7 and 8 nurses		0.002 (0.003) [0.418]			
Average firm-specific experience among NAs (years)			-0.003 (0.004) [0.355]		
Average firm-specific experience among RNs (years)			-0.002 (0.004) [0.605]		
Average shifts in the ward in past 90 days (NAs)				-0.000 (0.001) [0.853]	
Average shifts in the ward in past 90 days (RNs)				0.002 (0.002) [0.473]	
Average shifts with team members in past 90 days (NAs)					-0.001 (0.002) [0.640]
Average shifts with team members in past 90 days (RNs)					0.003 (0.003) [0.349]
Patient volume and characteristics	✓	✓	✓	✓	✓
Time controls	✓	✓	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓	✓	✓
N	18,922	18,922	18,001	13,598	13,598
Clusters	52	52	52	52	52
Mean of outcome variable	3.69	3.69	3.69	3.69	3.69

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Coefficients shown are from a fixed-effects regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses; p-values are shown in square brackets. The unit of analysis is the hospital unit-day. Time controls include a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

Table A4: Impact of deaths on previous days on levels of nurse staffing

	Outcome:	
	NA new sickness absence	RN new sickness absence
	(1)	(2)
Death in the ward on day T-1	0.880 (0.082) [0.171]	0.959 (0.070) [0.570]
Death in the ward on day T-2	1.364*** (0.140) [0.002]	1.025 (0.076) [0.734]
Death in the ward on day T-3	0.925 (0.152) [0.635]	1.057 (0.122) [0.630]
Patient and time controls	✓	✓
Ward fixed effects	✓	✓

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Coefficients shown are odds ratios from a fixed-effect logit regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses; p-values are shown in square brackets. The unit of analysis is the hospital ward-day. All regressions control for the mean patient age; the mean patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise interaction terms between age, female share, non-white share and Elixhauser Cormorbidity Index; the mean hospital length of stay of patients treated in the unit that day; the number of patients treated in the unit that day; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

Table A5: Serial correlation in patient deaths

	Outcome: patient death on day T		
	(1)	(2)	(3)
Patient death on day T-1	1.003 (0.157) [0.986]	1.004 (0.150) [0.980]	1.000 (0.149) [0.999]
Time controls	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓
Patient volume and characteristics	×	✓	✓
Staffing levels and composition	×	×	✓
N	18,870	18,870	18,870
Clusters	52	52	52

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Coefficients shown are odds ratios from a fixed-effect logit regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses; p-values are shown in square brackets. The unit of analysis is the hospital unit-day. Time controls include a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday. Patient controls include the mean patient age; the mean patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise interaction terms between age, female share, non-white share and Elixhauser Cormorbidity Index; the mean hospital length of stay of patients treated in the unit that day; and the number of patients treated in the unit that day. Staffing controls include the fraction of planned NA hours actually worked (the NA fill-rate); the fraction of planned RN hours worked by band 5s; the fraction of planned RN hours worked by band 6s; and the fraction of planned RN hours worked by band 7 and 8s.

Table A6: Nurse staffing, team spillovers and subsequent patient deaths

	Outcome: patient treated by the team dying under the care of another team within:		
	1 day	2 days	3 days
	(1)	(2)	(3)
Nursing Assistant (NA) fill-rate	1.001 (0.003) [0.618]	0.998 (0.002) [0.498]	0.998 (0.002) [0.284]
Registered Nurse (RN) fill-rate	0.994 (0.005) [0.210]	0.998 (0.005) [0.715]	1.000 (0.005) [0.960]
Patient volume and characteristics	✓	✓	✓
Time controls	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓
N	18,818	18,766	18,714
Clusters	52	52	52

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Coefficients shown are odds ratios from a fixed-effect logit regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital ward level, are shown in parentheses; p-values are shown in square brackets. The outcome variable is defined as being equal to 1 if a patient under the care of a team subsequently dies under the care of another team within the specified number of days. The unit of analysis is the hospital ward-day. All regressions control for the mean patient age; the mean patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise interaction terms between age, female share, non-white share and Elixhauser Cormorbidity Index; the mean hospital length of stay of patients treated in the unit that day; the number of patients treated in the unit that day; a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

Table A7: Serial correlation in Registered Nurse staffing

	RN staffing outcome on day T:			
	Fill rate		Staffing shortfall (hours)	
	(1)	(2)	(3)	(4)
RN fill rate on day T-1	0.265***	0.263***		
	(0.016)	(0.015)		
RN hourly staffing shortfall on day T-1			0.338***	0.337***
			(0.033)	(0.033)
Time controls	✓	✓	✓	✓
Patient volume and characteristics	×	✓	×	✓
Hospital ward fixed effects	✓	✓	✓	✓
N	18,870	18,870	18,870	18,870
Adjusted R-squared	0.105	0.107	0.149	0.150

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Coefficients shown are from a fixed-effect regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital ward level, are shown in parentheses. The unit of analysis is the hospital ward-day. Time controls include a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday. Patient controls include the mean patient age; the mean patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise interaction terms between age, female share, non-white share and Elixhauser Cormorbidity Index; the mean hospital length of stay of patients treated in the unit that day; and the number of patients treated in the unit that day. RN = Registered Nurse. Staffing shortfall is defined as the difference between the planned number of RN hours in a ward on a given day and the actual number of hours worked by RNs.

Table A8: Impact of firm-specific experience on patient mortality

	Outcome: patient death		
	(1)	(2)	(3)
Share (ppt) of planned Nursing Assistant (NA) hours filled	1.003 (0.003) [0.386]	1.003 (0.003) [0.331]	1.003 (0.003) [0.365]
Share (ppt) of planned Registered Nurse (RN) hours filled by:			
Band 5 nurses	0.989* (0.006) [0.082]	0.989* (0.006) [0.062]	0.988* (0.006) [0.067]
Band 6 nurses	0.986** (0.007) [0.037]	0.986** (0.007) [0.036]	0.987** (0.007) [0.049]
Band 7 and 8 nurses	0.979* (0.011) [0.066]	0.980* (0.011) [0.078]	0.980* (0.011) [0.088]
Average experience in the Trust (years)	0.939* (0.034) [0.079]		
Average experience in the Trust for NAs (years)		1.017 (0.015) [0.246]	
Average experience in the Trust for RNs (years)		0.928*** (0.021) [0.001]	
Percent of Registered Nurses (RNs) with: †			
1-3 years Trust experience			0.991** (0.004) [0.047]
3-7 years Trust experience			0.990** (0.005) [0.034]
7-11 years Trust experience			0.987*** (0.004) [0.003]
11-15 years Trust experience			0.988** (0.005) [0.016]
15+ years Trust experience			0.985*** (0.005) [0.005]
Patient volume and characteristics	✓	✓	✓
Time controls	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓
N	18,922	18,922	18,922
Clusters	52	52	52

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Coefficients shown are odds ratios from a fixed-effect logit regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses; p-values are shown in square brackets. The unit of analysis is the hospital unit-day. Nursing assistants (NAs) are personnel without formal training or registration requirements; Registered Nurses (RNs) are fully qualified nurses on the Nursing and Midwifery Council register, who have completed formal training and hold a university diploma or degree-level qualification. See notes to Table ?? for a full list of controls.

† Reference group is those with less than one year of experience in the Trust.

Table A9: Robustness of main results to alternative specification of staffing shortfalls

	Outcome: patient death		
	OR	Std. Err.	p value
RN staffing shortfall relative to target (hours)	1.007	(0.005)	[0.126]
Binary indicator of RN staffing shortfall:			
Above ward-specific mean	1.053	(0.065)	[0.401]
$\geq$ 1 S.D. above ward-specific mean	1.210**	(0.091)	[0.011]
Binary indicator of RN staffing below target by:			
4 hours or more	1.048	(0.062)	[0.434]
8 hours or more	1.065	(0.065)	[0.302]
12 hours or more	1.099	(0.078)	[0.185]
16 hours or more	1.241**	(0.131)	[0.040]
20 hours or more	1.277*	(0.163)	[0.054]
24 hours or more	1.390**	(0.216)	[0.035]
NA staffing hours		✓	
Planned RN staffing hours		✓	
Patient volume and characteristics		✓	
Time controls		✓	
Hospital ward fixed effects		✓	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Note: Results listed are from nine separate logistic regression models, each with ward fixed effects. Unit of analysis is the hospital ward-day (N=18,922). ORs denote odds ratios. Std. Err. denotes standard errors clustered at the ward level (52 clusters). RNs = registered nurses. NAs = nursing assistants. S.D. = standard deviation. 'Patient volume and characteristics' includes: mean patient age; the patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise-interaction terms between mean age, female share, non-white share, and Elixhauser Comorbidity Index; the mean hospital length of stay of patients treated in the unit that day; and the number of patients treated in the unit that day. 'Time controls' includes: a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday. Target RN staffing is defined as the total number of rostered (planned) hours for registered nurses (those in band 5 or above), which is equal to the sum of filled and unfilled hours.

Table A10: Patient-day level analysis: impact of staffing fill-rate and firm-specific experience on mortality

	Outcome: death (binary)		
	(1)	(2)	(3)
NA fill-rate	1.004 (0.004) [0.280]	1.004 (0.004) [0.262]	1.004 (0.004) [0.284]
RN fill-rate	0.993 (0.007) [0.306]		0.993 (0.007) [0.274]
Band 5 RN fill-rate		0.995 (0.007) [0.453]	
Band 6 RN fill-rate		0.989 (0.007) [0.138]	
Band 7 and 8 RN fill-rate		0.981* (0.011) [0.088]	
Average experience in the Trust for NAs (years)			1.022 (0.017) [0.182]
Average experience in the Trust for RNs (years)			0.945** (0.024) [0.026]
Patient characteristics	✓	✓	✓
Time controls	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓
N	270,553	270,553	270,356
Clusters (teams)	18,564	18,564	18,544

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Coefficients shown are odds ratios from a logit regression. Robust standard errors clustered at the hospital unit-day level (i.e. team level) are shown in parentheses, and p-values are shown in square brackets. The unit of analysis is the patient-day. Staffing variables are calculated as the average across the wards the patient was treated in that day, weighted by the number of hours spent in each. All regressions control for patient age; patient age^2; patient age^3; sex; a dummy for being non-white; the patient's Elixhauser Cormorbidity Index; pairwise interaction terms between age, sex, non-white dummy and Elixhauser Cormorbidity Index; a set of dummies for 17 separate diseases, derived from diagnosis codes; the day within the patient's hospital spell; a dummy for each day of the week; a dummy for month of the year; a dummy indicating whether the day was a bank holiday; and a set of dummies for each hospital ward the patient was treated in that day.

Table A11: Patient-day level analysis: heterogeneity by patient diagnosis

	Outcome: patient death			
	Without sepsis†		With sepsis†	
	(1)	(2)	(3)	(4)
NA fill-rate	1.003 (0.004) [0.498]	1.003 (0.004) [0.483]	1.007 (0.007) [0.347]	1.007 (0.007) [0.320]
RN fill-rate	1.003 (0.008) [0.740]		0.969*** (0.012) [0.009]	
Band 5 RN fill-rate		1.005 (0.008) [0.591]		0.971** (0.012) [0.019]
Band 6 RN fill-rate		1.000 (0.009) [0.958]		0.958*** (0.013) [0.003]
Band 7 & 8 RN fill-rate		0.987 (0.013) [0.326]		0.961** (0.019) [0.048]
Patient characteristics	✓	✓	✓	✓
Time controls	✓	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓	✓
N	232,324	232,324	24,715	24,715
Clusters	17,070	17,070	11,201	11,201

† A patient is coded as having sepsis if they have a primary or secondary ICD-10 diagnosis code of A400, A403, A409, A410, A411, A414, A415, A418, or A419.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Coefficients shown are odds ratios from a fixed-effect logit regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses; p-values are shown in square brackets. The unit of analysis is the patient-day. Staffing variables are calculated as the average across the wards the patient was treated in that day, weighted by the number of hours spent in each. All regressions control for patient age; patient age^2; patient age^3; sex; a dummy for being non-white; the patient's Elixhauser Cormorbidity Index; pairwise interaction terms between age, sex, non-white dummy and Elixhauser Cormorbidity Index; a set of dummies for 17 separate diseases, derived from diagnosis codes; the day within the patient's hospital spell; a dummy for each day of the week; a dummy for month of the year; a dummy indicating whether the day was a bank holiday; and a set of dummies for each hospital ward the patient was treated in that day.

Table A12: Patient-day level analysis: heterogeneity by patient severity

	Outcome: patient death			
	Low-severity patients†		High-severity patients†	
	(1)	(2)	(3)	(4)
NA fill-rate	0.998 (0.005) [0.661]	0.998 (0.005) [0.648]	1.008 (0.005) [0.124]	1.008 (0.005) [0.115]
RN fill-rate	0.984 (0.011) [0.166]		0.998 (0.008) [0.827]	
Band 5 RN fill-rate		0.986 (0.012) [0.233]		1.000 (0.009) [0.999]
Band 6 RN fill-rate		0.984 (0.012) [0.167]		0.993 (0.009) [0.422]
Band 7 & 8 RN fill-rate		0.961** (0.017) [0.024]		0.990 (0.014) [0.497]
Patient characteristics	✓	✓	✓	✓
Time controls	✓	✓	✓	✓
Hospital ward fixed effects	✓	✓	✓	✓
N	138,649	138,649	120,771	120,771
Clusters	16,658	16,658	17,853	17,853

† The median Elixhauser Cormorbidity Index among patients who died in our sample was 5. A low-severity patient is defined as one with an Elixhauser score of less than 5; a high-severity patient as one with an Elixhauser score of 5 or more.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Coefficients shown are odds ratios from a fixed-effect logit regression, with hospital ward fixed effects. Robust standard errors, clustered at the hospital unit level, are shown in parentheses; p-values are shown in square brackets. The unit of analysis is the patient-day. Staffing variables are calculated as the average across the wards the patient was treated in that day, weighted by the number of hours spent in each. All regressions control for patient age; patient age^2; patient age^3; sex; a dummy for being non-white; the patient's Elixhauser Cormorbidity Index; pairwise interaction terms between age, sex, non-white dummy and Elixhauser Cormorbidity Index; a set of dummies for 17 separate diseases, derived from diagnosis codes; the day within the patient's hospital spell; a dummy for each day of the week; a dummy for month of the year; a dummy indicating whether the day was a bank holiday; and a set of dummies for each hospital ward the patient was treated in that day.

Table A13: Interactions between NA and RN staffing

	Outcome: patient death		
	OR	Std. Err.	p value
NA fill-rate	1.043	(0.062)	[0.483]
RN fill-rate	1.025	(0.060)	[0.666]
NA fill-rate * RN fill-rate	1.000	(0.001)	[0.521]
Patient volume and characteristics		✓	
Time controls		✓	
Hospital ward fixed effects		✓	
N		18,922	
Clusters (wards)		52	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Note: ORs denote odds ratios. Std. Err. denotes standard errors clustered at the ward level (52 clusters). RNs = registered nurses. NAs = nursing assistants. S.D. = standard deviation. 'Patient volume and characteristics' includes: mean patient age; the patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise-interaction terms between mean age, female share, non-white share, and Elixhauser Comorbidity Index; the mean hospital length of stay of patients treated in the unit that day; and the number of patients treated in the unit that day. 'Time controls' includes: a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.

Table A14: Interactions between presence of a senior nurse and band 5 and 6 RN staffing

	Outcome: patient death		
	OR	Std. Err.	p value
NA fill-rate	1.003	(0.004)	[0.371]
Estimated Band 5 & 6 fill-rate†	0.982*	(0.009)	[0.058]
Dummy for presence of any Band 7 or 8 staff	0.262	(0.311)	[0.259]
Estimated Band 5 & 6 fill-rate * Band 7 or 8 dummy	1.013	(0.013)	[0.310]
Patient volume and characteristics		✓	
Time controls		✓	
Hospital ward fixed effects		✓	
N		18,922	
Clusters (wards)		52	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Note: ORs denote odds ratios. Std. Err. denotes standard errors clustered at the ward level (52 clusters). RNs = registered nurses. NAs = nursing assistants. S.D. = standard deviation. 'Patient volume and characteristics' includes: mean patient age; the patient age^2; the female share of patients; the non-white share of patients; the mean Elixhauser Comorbidity Index; pairwise-interaction terms between mean age, female share, non-white share, and Elixhauser Comorbidity Index; the mean hospital length of stay of patients treated in the unit that day; and the number of patients treated in the unit that day. 'Time controls' includes: a dummy for each day of the week; a dummy for month of the year; and a dummy indicating whether the day was a bank holiday.
† This is estimated fraction of planned Band 5 & 6 RN hours worked by Band 5 & 6 staff, defined as (Actual Band 5 & 6 hours) / (Total planned RN hours, minus actual band 7 & 8 hours).

A.4 Back of the envelope calculation of cost of staffing shortages

Our estimates suggest that adding a 12-hour shift from a degree-qualified nurse to the average team would reduce the odds of experiencing a patient death by around 10% (as per Section 5). This implies a reduction in the probability of a death of around 0.6 percentage points.[35] That suggests that it would require roughly 170 additional shifts in order to save one life. Based on the hourly rates for agency nurses – the marginal unit of nursing available in the short-term – used by the hospital in 2022/23, each additional shift would cost around £440.[36] A back-of-the-envelope calculation therefore suggests that each life saved would cost approximately £75,000. The cost per quality-adjusted life year (QALY) threshold used by the National Institute for Health and Care Excellence (NICE) for England and Wales is between £20,000 and £30,000. We are not able to estimate the number of QALYs saved by each additional nurse and so cannot conduct a direct cost-utility analysis of the effects of higher nurse staffing. Our estimates suggest, though, that each patient who died in our sample – among whom the median age was 77 – who might have otherwise survived with higher levels of nursing staff, would need to have otherwise lived for a further three or four years for this to be a cost-effective intervention.

[35] A patient death occurs in 6.3% of teams in our sample, equivalent to baseline odds of $0.063/(1 - 0.063) \approx 0.067$. A 10% reduction in those odds would take it to ≈ 0.060, equivalent to a probability of $0.06/1.06 \approx 5.7\%$, around 0.6 percentage points lower than the sample mean probability of 6.3%.

[36] This is the mean hourly price cap for agency nurses, using the highest spine point within each AfC band and weighting by the proportion of hours worked by staff of different grades, on weekends, and at night. This suggests a mean hourly rate of around £36.70.