



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Turner, AJ;Anselmi, L;Lau, YS;Sutton, M

Title:

The effects of unexpected changes in demand on the performance of emergency departments

Date:

2020-12-01

Citation:

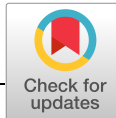
Turner, A. J., Anselmi, L., Lau, Y. S. & Sutton, M. (2020). The effects of unexpected changes in demand on the performance of emergency departments. *Health Economics (United Kingdom)*, 29 (12), pp.1744-1763. <https://doi.org/10.1002/hec.4167>.

Persistent Link:

<https://hdl.handle.net/11343/252597>

License:

CC BY



The effects of unexpected changes in demand on the performance of emergency departments

Alex J. Turner | Laura Anselmi | Yiu-Shing Lau | Matt Sutton

Health Organisation, Policy and Economics (HOPE) Group, Centre for Primary Care & Health Services Research, The University of Manchester, Manchester, UK

Correspondence

Alex J. Turner, Health Organisation, Policy and Economics (HOPE) Group, Centre for Primary Care & Health Services Research, The University of Manchester, Manchester, UK.
Email: alexander.turner@manchester.ac.uk

Funding information

Health Foundation

Abstract

Crowding in emergency departments (EDs) is increasing in many health systems. Previous studies of the relationship between crowding and care quality are limited by the use of data from single hospitals, a focus on particular patient groups, a focus on a narrow set of quality measures, and use of crowding measures which induce bias from unobserved hospital and patient characteristics. Using data from 139 hospitals covering all major EDs in England, we measure crowding using quasi-exogenous variation in the volume of ED attendances and examine its impacts on indicators of performance across the entire EDs care pathway. We exploit variations from expected volume estimated using high-dimensional fixed effects capturing hospital-specific variation in attendances by combinations of month and hour-of-the-week. Unexpected increases in attendance volume result in substantially longer waiting times, lower quantity and complexity of care, more patients choosing to leave without treatment, changes in referral and discharge decisions, but only small increases in reattendances and no increase in mortality. Causal bounds under potential omitted variable bias are narrow and exclude zero for the majority of outcomes. Results suggest that physician and patient responses may largely mitigate the impacts of demand increases on patient outcomes in the short-run.

KEYWORDS

attendance volume, crowding, emergency departments, physician behavior, quality of care

1 | INTRODUCTION

Hospital departments providing emergency care are facing unprecedented pressures. In both the United States and the United Kingdom, the last decade has seen substantial increases in attendances at emergency departments (EDs; Trendwatch, 2015). In England, attendances at major EDs grew 24% between 2003/4 and 2018/19 and by 5.8% between September 2018 and September 2019 (NHS England, 2019b). Coupled with supply side pressures such as reductions in the number of inpatient beds and increased rates of patients facing delayed discharges (NHS England, 2017b; The King's

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Health Economics published by John Wiley & Sons Ltd.

Fund, 2017a), increases in demand have contributed to ED crowding, defined as “when the identified need for emergency services exceeds available resources for patient care” (American College of Emergency Physicians, 2016).

Multiple reviews have established deleterious effects of ED crowding (Bernstein et al., 2009; Carter, Larson, & Pouch, 2014; George & Evridiki, 2015; Higginson, 2011; Hoot & Aronsky, 2008; Morley, Unwin, Peterson, Stankovich, & Kinsman, 2018; Morris, Boyle, Beniuk, & Robinson, 2012; Stang, Crotts, Johnson, Hartling, & Guttmann, 2015). These reviews identify associations between ED crowding and indicators of poor processes of care, such as increased rates of medical errors (Kulstad, Sikka, Sweis, Kelley, & Rzechula, 2010) and delays to treatment (Fee, Weber, Maak, & Bacchetti, 2007; Hwang, Richardson, Sonuyi, & Morrison, 2006; Schull, Vermeulen, Slaughter, Morrison, & Daly, 2004), and worsening patient outcomes, such as increased rates of adverse cardiovascular events (Pines et al., 2009), mortality (Cha et al., 2011), reattendance (Sun, Heng, Tay, & Tan, 2015), and reductions in patient satisfaction (Pines et al., 2008).

However, this literature is limited by the use of crowding measures which are susceptible to significant amounts of unobserved confounding. Crowding has typically been measured using indicators of throughput/output such as ED and inpatient bed occupancy, or measures that could be considered outcomes of crowding, such as ED waiting times. Throughput is partly determined by the actions of EDs, inducing correlation between crowding and unmeasured hospital circumstances that also predict outcomes. For example, high levels of bed occupancy may reflect problems with patient management; a source of unmeasured confounding that could inflate associations between crowding, poor quality of care, and adverse patient outcomes.

Some studies rely solely on input measures of crowding such as attendance volume (Asaro, Lewis, & Boxerman, 2007; Cha et al., 2011; Fee et al., 2007; Hong, Shin, Song, Cha, & Cho, 2013; Miró et al., 1999). Attendance volume is independent of the actions of EDs, reducing the risk of confounding from unmeasured hospital circumstances. However, variations in the actual number of ED attendances have typically been used. The average severity of patients attending ED could differ in periods of high attendance volume, representing confounding from unmeasured patient characteristics. This is supported by evidence of lower levels of average severity in daytime periods when attendance volume is high (Anselmi, Meacock, Kristensen, Doran, & Sutton, 2017), potentially due to the ability of less severe patients to self-select their time of attendance. Using actual variations also implicitly assumes that EDs' capacity to cope with identical changes in attendance volume have the same effect across the day, week, and year. Based on historical patterns in attendance volume (Anselmi et al., 2017; NHS England, 2017a), some variations in attendances are predictable, and EDs can plan for these and adjust capacity (e.g., by setting staff rotas) accordingly. An identical level of attendance volume may have substantially greater effects on ED care in a typically quiet period.

We address these limitations by measuring crowding using quasi-exogenous variation in the volume of ED attendances. We define this as deviations from expected volume, where expected volume is estimated using fixed effects capturing hospital-specific variation in attendances by hour-of-the-day, day-of-the-week, and month-of-the-year. We therefore exploit variation in attendance volume occurring within a single hour. This variation is therefore independent of scheduling decisions made by the hospital and removes the impacts of patient self-selection by time of day. This approach is similar to that employed in Martins and Filipe (2020), who use attendance volume as an instrument to examine physician responses to extended queue lengths in ED. We additionally apply the consistent estimator of bias derived in Oster (2019) to estimate causal bounds of the effect of attendance volume on ED performance.

We also make several additional contributions to the crowding literature. Recent systematic reviews have identified studies mainly conducted in North America and Australasia, with no studies conducted in a UK-setting (Carter et al., 2014; George & Evridiki, 2015). Differences between the UK model of emergency care and the model implemented in North America and Australasia, such as strict performance standards the ability of general practitioners (GPs) to directly admit emergency cases to hospital, may act to change the effects of ED crowding (Higginson, 2011). The generalizability of findings from previous studies is further reduced by the prevalence of time-series studies from single-hospital sites, and a focus on specific patient groups rather than the whole population of attendees. We extend previous studies by using attendance-level data on all attendances (approximately 14 million) from all major EDs in England in the financial year 2016/17.

We further build on the crowding literature by examining the effect of attendance volume on performance indicators covering over the entire ED care pathway. This extends previous studies that have examined effects on single performance measures in isolation, limiting their ability to provide a description of the likely mechanisms through which attendance volume impacts patient outcomes.

Our results suggest higher attendance volume results in a substantial deterioration in some indicators of ED performance. We find that unexpected increases in attendance volume result in substantially higher waiting times, lower quantity and complexity of care, more patients choosing to leave ED without treatment, changes in referral and

discharge decisions, but only small increase in ED reattendances and no increase in mortality. We also show that, unlike raw variation in attendance volume, deviations from expected volume are practically uncorrelated with observed indicators of patient severity. The estimated causal bounds for effects on these outcomes are also narrow and do not include zero for almost all outcomes. The results are also robust to changes in the estimation method used, the definition of attendance volume, and restrictions to the estimation sample.

2 | INSTITUTIONAL BACKGROUND

All emergency care in England are provided in publically funded EDs and care for all residents is free at the point of use. EDs are of three types (The King's Fund, 2017b). We focus on Type 1 EDs, which provide a consultant-led 24-h service with resuscitation facilities, and treat the majority of patients (65% in 2016/17). Type 2 EDs provide emergency care for single specialties and Type 3 EDs are walk-in centers or minor injury units providing ambulatory care only for less urgent and/or less severe cases.

Following treatment, the government reimburses hospitals for the care they provide. Reimbursement is based on 11 healthcare resource groups (HRGs), with reimbursement higher if more resource-intensive investigations (e.g., CT scans) and treatments (e.g., lumbar puncture), are performed (Monitor, 2016; Parkinson, Meacock, & Sutton, 2019). The level of reimbursement is determined by HRG-specific national tariffs, which are updated annually by the Department of Health to account for cost inflation and expected efficiency improvements. 2016/17 tariffs for Type 1 EDs are provided in Table A1 (Supplementary Appendix 1).

A patient's care pathway is broadly consistent across Type 1 EDs, with the first stage depending on the method of arrival. Patients not arriving via ambulance first register at reception and provide personal information and a brief summary of the health condition for which they are attending. These patients then wait to be triaged, where an initial assessment is performed by a medical or nursing professional to determine priority for treatment. This may include taking a brief medical history, a pain assessment and the collection of early warning scores; including data on vital signs (Department of Health, 2010). Patients arriving by ambulance undergo a separate and quicker triage process, aided by information collected on their condition by ambulance staff.

Patients are then added to queues based on their severity where they wait to be assessed by a senior clinician. The senior clinician may arrange or perform a series of investigations to aid diagnosis (e.g., via blood tests and X-rays), followed by definitive treatment where necessary (Department of Health, 2010).

Once any investigations or treatments are performed, the pathway concludes with a senior clinician deciding on the destination of discharge, which includes admission for inpatient care, discharge (either with no follow-up or follow-up by GPs), a transfer to another provider for additional treatment, or a referral for less urgent treatment (e.g., to an outpatient clinic). The discharge destination can also be determined by the patient themselves, who may choose to leave the ED without treatment (after being initially assessed) or after refusing treatment.

3 | MECHANISMS

Demand-induced crowding has the potential to impact all stages of the ED care pathway. However, physician and patient decisions (and subsequent patient outcomes) are inherently inter-linked, making generating theoretical predictions of the effects of crowding complex. Figure 1 presents one possible set of mechanisms.

Physician decision-making over the supply of medical care is typically assumed to be determined by a profit motive as well as an altruistic motive governing physician preferences relating to improving patient health (Allard, Jelovac, & Léger, 2011; Chalkley & Malcomson, 1998; Ellis & McGuire, 1986; Liu & Ma, 2013). In re-imbursement systems where physician pay is invariant to the volume and type of patient care, such as in England, the lack of financial incentives may mean physician behavior more closely aligns with health maximization (Godager & Wiesen, 2013).

In the event of unexpected increases in demand, physician responses (under fixed ED capacity) are therefore likely guided by a time-quality trade-off (Martins & Filipe, 2020). Patient health, as well as other important outcomes such as lower treatment costs, has been found to be increasing with greater time spent with patients (Chen, Farwell, & Jha, 2009) and reducing in waiting times for both elective care (Nikolova, Harrison, & Sutton, 2016) and ED care (Gruber, Hoe, & Stoye, 2018; Guttman, Schull, Vermeulen, & Stukel, 2011; Woodworth & Holmes, 2020). Thus, in trying to maximize patient health in response to a demand increase, physicians must choose to either maintain waiting times

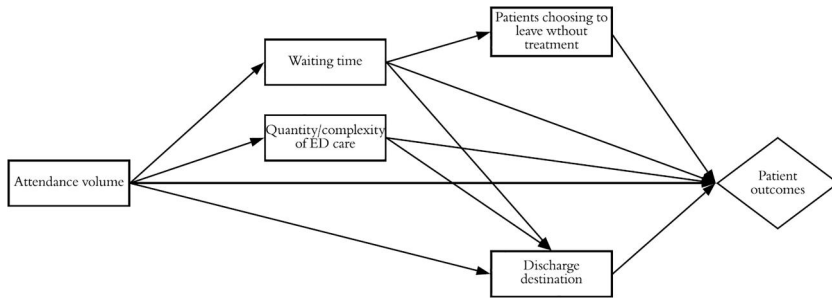


FIGURE 1 Mechanisms through which attendance volume may impact the emergency department care pathway

through reducing the amount of time spent with patients (likely by rationing the volume and/or complexity of treatment); or maintaining the same volume/complexity of care and allowing waits to increase.

ED physicians may also attempt to relieve pressure through their choice of discharge destination by substituting care within the ED for care provided by alternative health care services. This may be for “step-up” care via inpatient admission, or for “step-down” care via referral to outpatient clinics or primary care practitioners, with more risk-averse physicians likely to substitute for “step-up” services (Pearson et al., 1995; Pines et al., 2010). The level of substitution to external services may be more limited in “out-of-hours” periods (weekends and weekday-nights), when these services are not available.

Via increases in waiting time, demand increases may also lead to increases in patients choosing to leave ED without treatment. In the absence of price, waiting times act as the rationing mechanism, with increases in waits deterring patients from demanding treatment (Lindsay & Feigenbaum, 1984; Martin & Smith, 1999). In the period of study, patients did not have access to real-time information on waiting times at alternative EDs. Therefore, unlike for elective care, patients could not observe ED waits at the point of demanding care and so actual waits could not determine the choice of whether to attend ED. Patients’ waiting time expectations are therefore set on arrival, and the reduction in quantity of demand is observed through whether patients choose to remain in ED once expected waits are revealed. The increase in leaving without treatment may also be more limited in out-of-hours periods, when the availability of substitutes is lower. Sivey (2018) finds that demand-induced increases in ED waits result in a large increase in the volume of patients leaving without treatment, with lower elasticity of demand on weekends.

Finally, through all of these mechanisms, an increase in attendance volume may result in a worsening of patient outcomes (such as mortality and reattendance rates). This could be driven by an increase in waiting times for all stages of treatment, reductions in the quantity and complexity of care, inappropriate changes to discharge destinations, and/or patients incorrectly leaving ED without treatment when ED care was required. Crowding-induced reductions in quality of care, detectable for example through increases in medication errors (Kulstad et al., 2010), may also have direct impact on patient outcomes. However, impacts on patient outcomes will depend on whether physician and patient responses are efficient. If increases in demand cause an increase in leaving without treatment only for patients who will benefit little from treatment, and physicians ration care or allow waits to increase only for patients where the consequences for health are minimal, effects on patient outcomes could be small.

4 | DATA

4.1 | Data sources

We use data from Hospital Episode Statistics (HES), which provides information on all attendances at English EDs and all admissions to English hospitals (NHS Digital, 2017a). The ED record contains the exact date and time of attendance, demographic characteristics, diagnoses made in the ED, arrival method, attendance category, accident location, accident type, and the source of referral (NHS Digital, 2016). The inpatient record contains information on date of admission and discharge, discharge method (including death), admission type (emergency or elective), and primary and secondary diagnoses based on International Classification of Diseases-10 codes (Herbert, Wijlaars, Zylbersztejn, Cromwell, & Hardelid, 2017). HES records also contain a pseudo-anonymized patient identifier which enables the tracking of ED attendances and inpatient stays for the same patient over time.

We additionally use the proportion of residents who are reliant on means-tested state benefits measured for a patient's small geographical area of residence (containing approximately 1500 individuals), captured by the income component of the 2015 Index of Multiple Deprivation (Department for Communities and Local Government, 2015).

The study population comprises ED attendances in the financial year 2016/17 occurring between 1st April 2016 and 31st March 2017. Data on attendances in the 2015/16 financial year (1st April 2015 to 31st March 2016) and inpatient admissions in the 2015/16 and 2016/17 financial years are used in the derivation of covariates.

4.2 | Outcomes

We use HES to construct outcomes covering the entire care pathway (Figure 2).

Following NHS methodology, we construct binary attendance-level indicators for nonadherence to NHS waiting time targets (NHS Digital, 2017b): total ED waiting time (between arrival and conclusion) > 4 h; waiting time for initial assessment for the most severe patients (identified by those arriving via ambulance) > 15 min; and waiting time for treatment > 60 min. We also examine impacts on continuous measures of these waiting times reported for all patients, as well as on the time between assessment and treatment and the time between treatment and conclusion.

We measure the quantity of ED care using separate per-attendance counts of the number of investigations performed and treatments received. To proxy complexity of care, we construct a measure of per-attendance ED reimbursement using data from national tariffs and the reported HRG.

We construct a binary indicator for patients refusing or leaving without treatment, and several indicators of patient's discharge destination including indicators for inpatient admission, discharge with no follow-up, discharge with follow-up from a GP in a primary care setting, referral to an outpatient/emergency clinic, and transfer for care at another hospital provider or with another health care professional.

Finally, we create three patient outcomes. First is a binary indicator of an unplanned follow-up attendance within 7 days classified at the same ED as being for the same incident. The NHS defines an unplanned follow-up as indicative of poor quality care and NHS targets require EDs to keep the rate below 5%. Second, we construct a more general measure of re-attendance, defined as any subsequent attendance, at any ED, within 7 days of the original attendance. This captures patients who reattend ED as a result of poor quality of care, but do not attend the same ED in which the poor care was provided; and patients for which it was unclear on attendance that a patient was reattending for a condition treated in a previous attendance. Finally, by linking ED and inpatient records, we create an indicator for in-hospital mortality, defined as the existence of a subsequent admission, at any hospital, with a discharge method of death, and a discharge date within 30 days of the original attendance.

4.3 | Sample restrictions

Due to differences in missing outcome data as well as differences in the relevant populations, we allow estimation samples to differ across outcomes (Figure A1, Supplementary Appendix 2).

In total, there were 20,886,412 attendances at English EDs in the 2016/17 financial year. We first restrict the analysis to the Type 1 departments due to the case-mix of attendances at these departments being more comparable across hospitals than Type 3 or Type 2 EDs. We then: (1) exclude individuals who were dead upon arrival at ED; (2) exclude planned follow-up attendances; and (3) where patients presented at the same ED on multiple occasions within the same day, restrict the sample to only the first attendance (see Supplementary Appendix 2 for further details).

The final sample contained data from 139 hospitals, with estimation samples ranging from 4,129,771 to 14,830,920.

5 | EMPIRICAL STRATEGY

5.1 | Measuring attendance volume

We measure attendance volume as the number of other patients arriving at the same hospital in the 3 h before and 3 h after the hour of attendance of an index patient. Given NHS targets require EDs to conclude 95% of attendances within

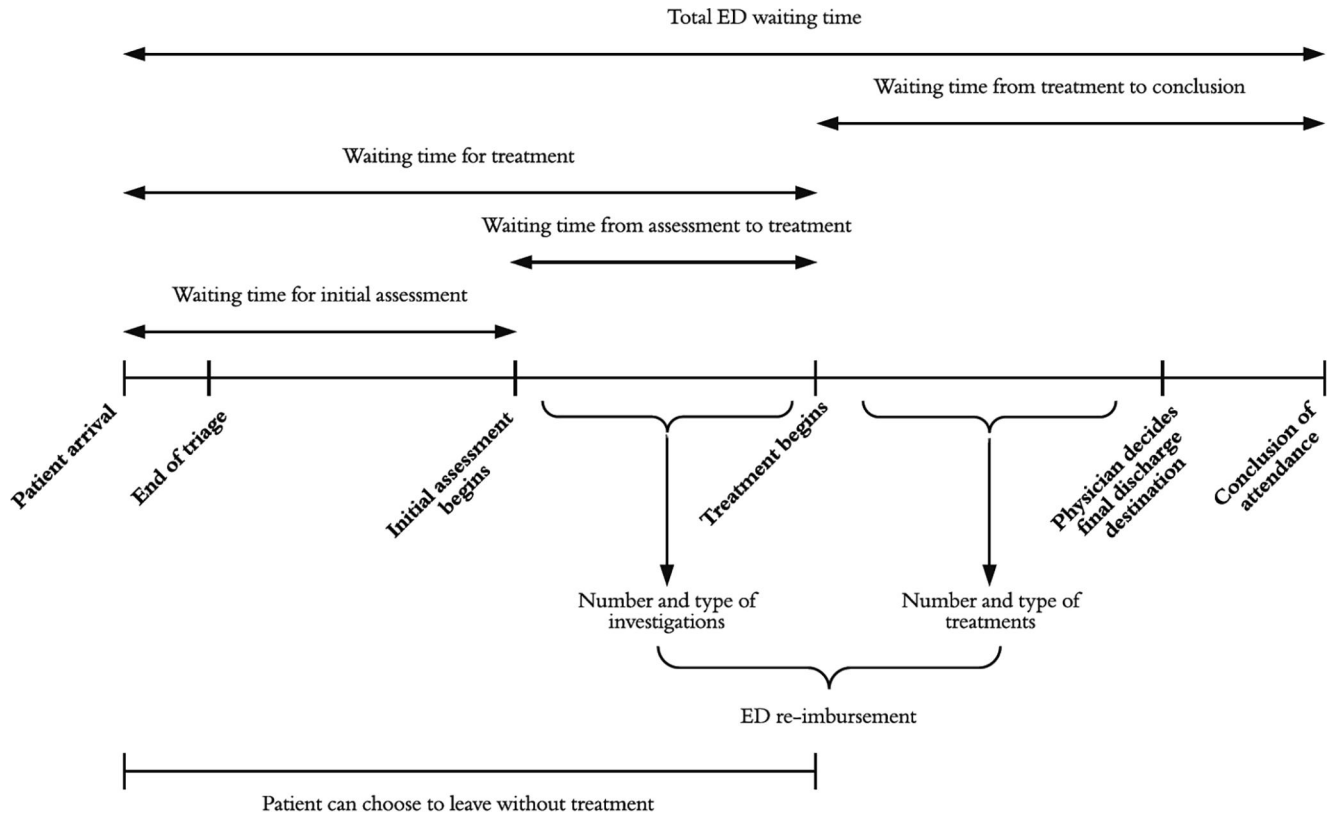


FIGURE 2 Performance indicators and the emergency department care pathway

4 h of arrival, this 7-h window captures the volume of other attendances that could potentially be competing for the same ED resources. The use of this same window for all attendances ensures the volume measure does not reflect hospital management of patient flow.

To explain further, consider a patient, *i*, who arrives at a hospital in hour *h*. If all patients had to wait 4 h between arrival and conclusion, then a patient arriving 3 h prior to patient *i* (*h*-3) would still be in the ED when patient *i* arrives in hour *h*. Similarly, patient *i* would still be in ED at the time of arrival of another patient arriving at ED 3 h following the arrival of patient *i* (in hour *h*+3). Thus, all other patients attending ED within the same 7-h window of patient *i*'s arrival are potentially competing for the same ED resources, and their arrival may therefore impact the level or quality of patient *i*'s care.

5.2 | Estimating volume effects

We estimate the effects of attendance volume on each indicator of ED performance using the following baseline specification:

$$y_{ijmh} = \beta v_{jmh} + \mathbf{x}_i \boldsymbol{\gamma} + \theta_{jmh} + u_{ijmh} \tag{1}$$

Here y_{ijhw} represents the outcomes of patient *i* attending hospital *j* in month-of-the-year *m* at hour-of-the-week *h*, v_{jhw} denotes attendance volume, \mathbf{x}_i is a vector of patient characteristics, and u_{ijhw} is an error term.

θ_{jmh} represents fixed effects for the 2,80,224 combinations of hospital, month, and hour-of-the-week. These fixed effects capture the expected volume of attendances at each patient's hour of attendance. After conditioning on these fixed effects, v_{jhw} can be interpreted as the variation in attendance volume causing volume to deviate from its expected levels. β can then be interpreted as the effect of an unexpected increase in attendance volume relative to hours in which attendances are equal to their expected levels.

The identification of causal effects requires unexpected variation in attendance volume to be independent of unobserved patient and hospital characteristics. After conditioning on fixed effects, we are exploiting variation in the volume of attendances across the 4–5 instances in which a given hour-of-the-week, in a given month, in a given hospital, is observed. Assuming that shift patterns at the same hospital are constant within the same month and hour-of-the-week, this variation can be assumed independent of scheduling decisions made by the hospital. Unexpected attendance volume is also less likely to be correlated with patient severity, as the inclusion of these fixed effects removes the potential impact of patient self-selection by time of day. These fixed effects also control for time-invariant unobserved hospital characteristics (e.g., hospital size and location) and, by interacting hospital-fixed effects with month-fixed effects, also control for time-varying unobserved hospital characteristics. By interacting time (month)-fixed effects with hospital-fixed effects, we also control for time shocks common to all hospitals (e.g., weather episodes and infectious illness) and allow these time shocks to have effects which vary by hospital.

However, to mitigate the risk of residual omitted variable bias, we control for a detailed set of patient socio-demographics and indicators of the severity of attendees.

We measure patient sociodemographics using gender-specific 5-year age bands, ethnicity (grouped according to census definitions), and quintiles of income deprivation in the individual's place of residence.

To control for patient severity, we include indicators for each of 41 primary ED diagnoses (such as cardiac conditions, respiratory conditions, and cerebro-vascular conditions; NHS Digital, 2016), mode of arrival (ambulance, other, and unknown), attendance category (first or unplanned follow-up attendance), cause of attendance (including road traffic accident and assault), referral source (including GP referral and self-referral), and incident location (including home and work). We use inpatient records to construct the number of emergency admissions to any hospital in the year prior to attendance (integers between 0 and 10, 11–15, 16–20, 21–30, >30). We also include separate indicators for the presence of 30 Elixhauser morbidities recorded during inpatient stays in the year prior to attendance, and a count of these morbidities (integers between 2 and 9, 10+) to control for multimorbidity (Gutacker, Bloor, & Cookson, 2015). We also include the number of ED attendances the individual had in the year prior to attendance (integers between 0 and 10, 2-integer groups between 11 and 20, 21–25, 26–30, 31–40, 41–50, >50).

We additionally control for quintiles of inpatient bed occupancy in the day prior to the attendance, as this may impact physician's choice of discharge destination (particularly admission) and be correlated with unexpected changes in attendance volume if variations in attendance volume are serially correlated. This was constructed from inpatient records as the number of inpatients in the hospital of attendance in the day prior to attendance, relative to the maximum daily number of inpatients staying in the same hospital over all days in that year. The maximum served as a proxy for the full capacity of the hospital.

We estimate the baseline specification using a linear high-dimensional fixed effects model (Correia, 2016). Standard errors were corrected for heteroscedasticity, and clustered at a hospital-level. Prior to estimation, we standardize the volume measure by dividing attendance volume by the standard deviation of unexpected attendance volume, such that β is interpreted on the standard deviation scale. To examine nonlinearity in the effects of attendance volume, we reestimate Equation (1) replacing v_{ijhw} with a deciles of attendance volume.

We also estimate all effects separately for “out-of-hours” periods (weekends and weekday nights [6 p.m.–6 a.m.]) and “in-hour” periods, to test whether effects on leaving without treatment and referral to other health care services are dependent on the availability of substitutes and whether this subsequently impacts patient outcomes.

Finally, we estimate effects on discharge destination excluding patients who leave without treatment. This is because these patients are included in the denominator for all discharge destination indicators and so an increase in leaving without treatment will reduce the likelihood of alternative discharge destinations irrespective of changes in physician behavior. We do this despite recognizing that conditioning on leaving without treatment will result in a selected sample.

6 | ROBUSTNESS CHECKS

We implement numerous tests to examine the robustness of results to confounding from unobserved patient severity under the assumption that bias from observed covariates is informative about bias from unobserved covariates (Altonji, Elder, & Taber, 2005). In addition to traditional coefficient comparison tests and balancing regressions

(see Supplementary Appendix 3), we implement the partial identification approach of Oster (2019), which estimates causal bounds by scaling changes in coefficients following the inclusion/removal of covariates by the incremental explained outcome variance from these covariates.

Bias-adjusted volume effects, β^* , are estimated as:

$$\beta^*(R_{max}, \delta) = \tilde{\beta} - \left[\delta(\tilde{\beta} - \check{\beta}) \frac{R_{max} - \check{R}}{\check{R} - \tilde{R}} \right] \quad (2)$$

where $\tilde{\beta}$ and \tilde{R} are volume effects and within-R-squared values estimated from the baseline specification, and $\check{\beta}$ and \check{R} are these same estimates when patient characteristics are removed. R_{max} is the within-R-squared from an (impossible) regression including all observed and unobserved covariates. δ is the degree of selection on observables relative to unobservables.

Assuming $R_{max} = 1.3\tilde{R}$ and less-than-proportional selection on observables and unobservables ($\delta < 1$), we estimate causal bounds as:

$$[\tilde{\beta}, \beta^*(\min[1.3\tilde{R}, 1], 1)] \quad (3)$$

We also estimate, assuming $R_{max} = 1.3\tilde{R}$, the value of δ required for bias-adjusted volume effects to be zero.

We also examine robustness to different estimation methods, by reestimating effects on binary outcomes using two-step logistic regression, effects on waiting times and the number of investigations and treatments using conditional fixed-effects Poisson regression, and effects on total ED waiting time after applying the inverse hyperbolic sine transformation (Supplementary Appendix 4).

Furthermore, we examine robustness to two alternative measures of attendance volume: (1) attendances occurring in the 1 h before and 1 h following a patient's hour of arrival, and (2) attendances occurring in the 3 h preceding the patient's hour of arrival (Supplementary Appendix 5).

Finally, we examine robustness to restricting samples to only the first attendances for each individual during the study period (Supplementary Appendix 6).

7 | MAIN RESULTS

7.1 | Descriptive statistics

Figure 3 shows variation in expected attendance volume by hour-of-the-week and week-of-the-year, generated as the prediction from a regression of attendance volume on hospital-month-hour-of-the-week fixed effects. Expected attendance volume is lowest on average between 04:00 and 05:00 and highest in the early afternoon, with smaller troughs on weekends. Attendances were highest at midday on a Monday. Expected volume was generally highest in summer and winter months, with troughs in May, September, and October.

Unexpected variations in attendance volume were approximately normally distributed around its mean of zero (Figure 4). This variation ranged from 121 fewer attendances than expected in a 7-hour arrival window to 101 additional attendances than expected. One standard deviation is equivalent to a deviation of 11.5 attendances from expectations within a 7-h period.

Table A10 (Supplementary Appendix 7) presents differences in average covariates for: (1) the full sample; (2) arrival hours where actual attendances were above and below the average; and (3) arrival hours where unexpected attendance volume was above and below the average (i.e., where attendances were above and below expected levels). These statistics indicate high actual attendance volume is correlated with a large reduction in patient severity. On average, patients attending ED in periods where actual attendance volume was above average were younger and a lower proportion arrived by ambulance, were referred by emergency services, and were recorded as having a severe ED diagnosis. These patients also had a lower proportion of each of the 30 Elixhauser comorbidities recorded on admissions in the year prior to attendance, had lower levels of multimorbidity, and had fewer emergency admissions and ED attendances in the year prior to attendance.

FIGURE 3 Expected attendance volume by day-of-the-week and hour of arrival and week of the year of arrival. Attendance volume is defined as the number of attendance in the 7-h window surrounding the hour of attendance. Expected attendance volume is calculated as the prediction from a regression of attendance volume on hospital-month-hour-of-the-week fixed effects [Colour figure can be viewed at wileyonlinelibrary.com]

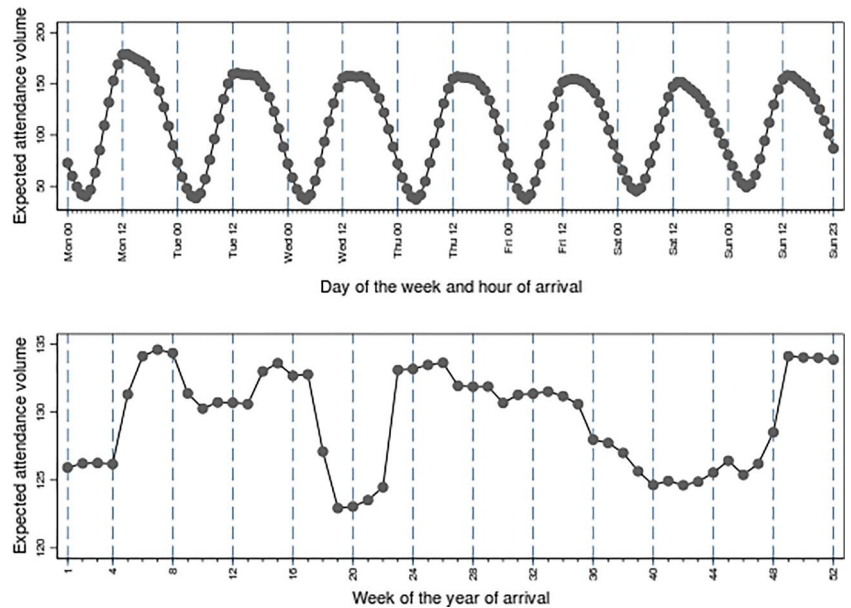
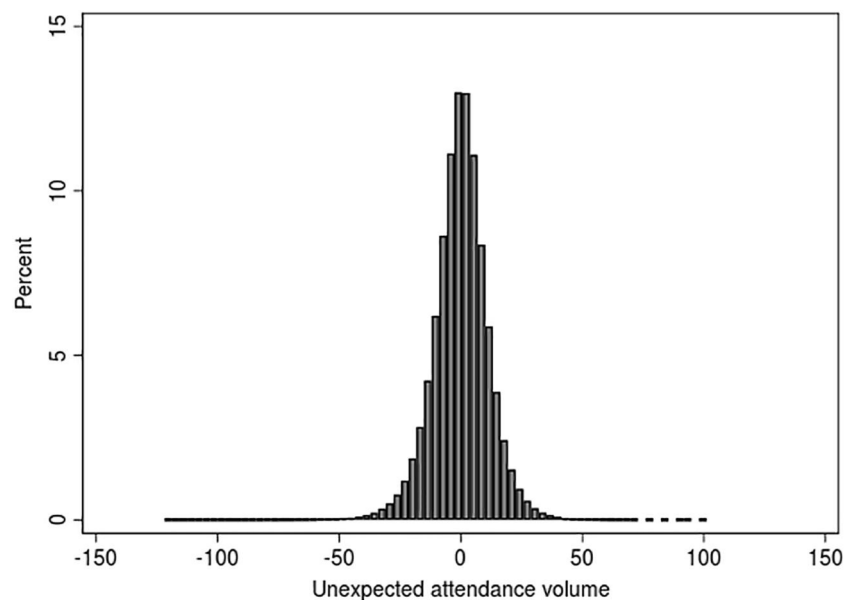


FIGURE 4 Distribution of unexpected attendance volume. Mean = 0.00, standard deviation = 11.49, minimum = -121.94, maximum = 101.11. Attendance volume is defined as the number of attendance in the 7-hour window surrounding the hour of attendance. Unexpected attendance volume is calculated as the residual from a regression of attendance volume on hospital-month-hour-of-the-week fixed effects



However, unexpected variation in attendance volume is considerably less correlated with patient severity, with patients attending in periods of high and low unexpected attendance appearing more similar. This provides supportive evidence for the relative exogeneity of unexpected attendance volume.

Unadjusted differences in average outcomes between patients attending ED in periods above and below mean unexpected volume suggest that higher unexpected attendance volume is associated with higher waiting times across multiple stages of the care pathway, and small reductions in ED reimbursement and the number of investigations and treatments (Table 1). There is also evidence of a large increase in patients choosing to leave ED without treatment. For patients not leaving without treatment, there is evidence of small differences in discharge destination, with a reduction in the rate of admission the most pronounced. Above mean unexpected attendance volume is also associated with a small increase in 7-day reattendance rates and a small reduction in 30-day mortality rates.

TABLE 1 Outcome descriptive statistics for the full sample and for periods below and above average unexpected attendance volume

Variable (Mean or %)	Full sample (<i>n</i> = 14,830,920)	Unexpected attendance volume		Difference
		Below mean (<i>n</i> = 7,240,116)	Above mean (<i>n</i> = 7,590,804)	
Waiting times targets				
>4 h total waiting time	17.07%	15.69%	18.39%	2.70%***
>15 min waiting time for initial assessment	27.73%	26.32%	29.12%	2.80%***
>60 min waiting time for treatment	53.72%	50.03%	57.28%	7.25%***
Continuous waiting times				
Total waiting time (minutes)	196.59	191.14	201.8	10.66***
Waiting time for initial assessment (minutes)	22.88	21.31	24.38	3.09***
Waiting time for treatment (minutes)	83.01	77.05	88.74	11.69***
Waiting time from assessment to treatment (minutes)	63.96	59.51	68.24	8.73***
Waiting time from treatment to departure (minutes)	119.3	119.35	119.25	−0.09
Quantity and complexity of ED care				
ED re-imburement (£)	114.05	114.65	113.48	−1.17***
Number of investigations	2.48	2.52	2.44	−0.08***
Number of treatments	2.60	2.62	2.57	−0.05***
Left without treatment	3.84%	3.35%	4.31%	0.95%***
Discharge destination				
Admitted	27.04%	27.41%	26.68%	−0.73%***
Discharged with no follow-up	33.86%	33.94%	33.79%	−0.15%**
Discharged with GP follow-up	19.65%	19.69%	19.61%	−0.08%**
Referred to clinic	9.73%	9.75%	9.72%	−0.03%
Referred to other healthcare professional or provider	4.22%	4.21%	4.22%	0.01%
Discharge destination (patients not choosing to leave without treatment)				
Admitted	28.12%	28.36%	27.88%	−0.48%***
Discharged with no follow-up	35.22%	35.12%	35.31%	0.20%***
Discharged with GP follow-up	20.44%	20.38%	20.49%	0.11%**
Referred to clinic	10.12%	10.09%	10.16%	0.07%***
Referred to other healthcare professional or provider	4.39%	4.36%	4.41%	0.05%**
Patient outcomes				
Unplanned follow-up attendance within 7 days	1.66%	1.65%	1.67%	0.03***
Any ED re-attendance within 7 days	8.68%	8.63%	8.73%	0.11%***
Mortality within 30 days	1.25%	1.27%	1.22%	−0.04%***

Notes: Statistics based on the estimation samples used for each outcome. Further details how these estimation samples are derived are provided in Supplementary Appendix 2.

Abbreviations: ED, emergency department; GP, general practitioner.

****p* < 0.001, ***p* < 0.01, **p* < 0.05. *p*-values calculated clustering standard errors at a hospital level.

7.2 | Main regression results

Table 2 presents the effects of attendance volume on each indicator of ED performance. Unexpected increases in attendance volume result in large and statistically significant increases in waiting times. A one standard deviation increase in attendance volume increases total ED waiting time by 6.9 min (3.5% relative to the baseline mean), driven by a 1.9 min (8.2%) increase in waiting time for initial assessment, a 4.9 min (7.6%) increase in waiting time between the initiation of assessment and treatment and a smaller 0.7 min (0.6%) increase in waiting time between treatment initiation and the conclusion of the attendance. This translates into an increase in the probability of total ED waiting time exceeding 4 h of 1.6 percentage points (9.6%), an increase in the probability of waiting over 15 min for initial assessment of 1.7 percentage points (6.1%), and an increase in the probability of waiting over 60 min for treatment of 4.2 percentage points (7.9%).

This same increase in attendance volume also leads to small but statistically significant reductions in the number of investigations per attendance (0.024; 1%), the number of treatments per attendance (0.021; 0.8%), and in ED reimbursement per attendance (£0.41; 0.4%).

We also find that a one standard deviation increase in volume results in a large increase in the probability of patients leaving without treatment of 0.5 percentage points (12.5%). Consistent with this large effect, we find reductions, although small, in all discharge destinations of patients, with statistically significant reductions in the probabilities of admission, discharge with no-follow-up, and referral to outpatient clinics of 0.17 percentage points (0.63%), 0.19 percentage points (0.56%), and 0.06 percentage points (0.62%), respectively. Reestimating these effects using only the sample of patients who did not leave without treatment (thus more closely reflecting effects on discharge destination due to the explicit decisions of ED physicians), we find that in response to increases in attendance volume physicians chose to admit fewer patients (0.1 percentage points, 0.36%) and refer fewer patients to outpatient clinics (0.02 percentage points, 0.20%) in favor of discharging more patients with follow-up in primary care (0.07 percentage points, 0.34%) and transferring more patients to other healthcare providers (0.02 percentage points, 0.46%).

For patient outcomes, we find no statistically significant change in the probability of mortality within 30 days of attendance, but small statistically significant increases in unplanned follow-up attendances and any reattendances within 7-days of 0.01 percentage points (0.7%) and 0.08 percentage points (1%), respectively.

Figures A12–A15 (Supplementary Appendix 8) depict the non-linearity in these effects. For the majority of outcomes, outcomes increased/decreased monotonically across the deciles of unexpected volume, although reductions in the number of treatments per attendance and increases in the rate of 7-day unplanned follow-up were more pronounced between deciles 9 and 10.

These Figures also show that the effects of attending ED in the highest volume periods are substantial compared to the lowest volume periods, for example, resulting in a 26 percentage point (152%) increase in the probability of total ED waiting time exceeding the 4-h target, and a 8 percentage point (200%) increase in the probability of patients choosing to leave ED without treatment. However, effects on discharge destinations and patient outcomes are smaller, with this same shift in volume causing a 0.2 percentage point (12%) and a 1.4 percentage point (16%) increase in the probabilities of unplanned follow-up and reattendance within 7 days, respectively.

7.3 | Heterogeneity by “in-hours” and “out-of-hours”

Patients who arrive in out-of-hours periods (weekends and weekday nights) wait longer for treatment, are more likely to choose to leave without treatment, receive similar quantity and complexity of care, and have larger reattendance rates (Table A11, Supplementary Appendix 9). In out-of-hours periods physicians are slightly more likely to admit patients, refer patients to other health care provides, and discharge patients (either with no or primary care follow-up), and considerably less likely to refer patients to outpatient clinics.

Estimating volume effects in these subsamples, we find that increases in attendance volume have greater impacts in out-of-hours periods (Table A12, Supplementary Appendix 9). Increases in waiting times for all stages of the ED care pathway are higher in these periods, as are the subsequent reductions in adherence to waiting times targets. The exception is for waiting time for treatment, where the smaller reduction in adherence to the target (despite larger increases in continuous waits) suggests effects on these waits in out-of-hours periods are more concentrated in the lower parts of the distribution.

TABLE 2 Regression estimates of the effect of a standard deviation increase in unexpected attendance volume on ED performance indicators

Outcome	Observations	AVE	AVE (% of baseline mean)
Waiting times targets			
>4 h total waiting time	14,822,381	0.016*** [0.014, 0.019]	9.61
>15 min waiting time for initial assessment	4,129,771	0.017*** [0.014, 0.020]	6.09
>60 min waiting time for treatment	13,180,807	0.042*** [0.037, 0.048]	7.89
Continuous waiting times			
Total waiting time	14,822,381	6.92*** [6.19, 7.64]	3.52
Waiting time for initial assessment	13,391,649	1.88*** [1.64, 2.11]	8.21
Waiting time for treatment	13,552,467	6.68*** [5.91, 7.45]	8.04
Waiting time from assessment to treatment	12,335,911	4.85*** [4.17, 5.53]	7.58
Waiting time from treatment to conclusion	13,548,661	0.71*** [0.47, 0.95]	0.59
Quantity and complexity of ED care			
ED re-imburement (£)	14,793,350	-0.409*** [-0.459, -0.359]	-0.36
Number of investigations	14,830,920	-0.024*** [-0.028, -0.021]	-0.98
Number of treatments	14,830,920	-0.021*** [-0.024, -0.019]	-0.82
Leaving without treatment	14,797,634	0.0048*** [0.0042, 0.0054]	12.50
Discharge destination			
Admitted	14,797,634	-0.0017*** [-0.0021, -0.0013]	-0.63
Discharged with no follow-up	14,797,634	-0.0019*** [-0.0023, -0.0014]	-0.56
Discharged with GP follow-up	14,797,634	-0.0005 [-0.0011, 0.0001]	-0.25
Referred to outpatient clinic	14,797,634	-0.0006*** [-0.0008, -0.0004]	-0.62
Referred to other healthcare professional or provider	14,797,634	-0.0002 [-0.0004, 0.0001]	-0.47
Discharge destination (for patients not choosing to leave without treatment)			
Admitted	14,229,264	-0.0007*** [-0.0011, -0.0003]	-0.25
Discharged with no follow-up	14,229,264	-0.0002 [-0.0006, 0.0003]	-0.06
Discharged with GP follow-up	14,229,264	0.0007* [0.0001, 0.0012]	0.34
Referred to outpatient clinic	14,229,264	-0.0002* [-0.0003, -0.0000]	-0.20
Referred to other healthcare professional or provider	14,229,264	0.0002* [0.0000, 0.0005]	0.46
Patient outcomes			
Unplanned follow-up attendance within 7 days	14,830,920	0.00011** [0.00004, 0.00018]	0.65
Any ED re-attendance within 7 days	14,830,920	0.0008*** [0.0005, 0.0010]	0.92
Mortality within 30 days	14,830,920	-0.00004 [0.00010, 0.00002]	-0.33

Notes: Coefficients represent marginal effects of a one standard deviation increase in attendance volume (~11.5 additional attendances within a 7-h window of a patient's arrival); 95% confidence intervals in parenthesis generated from standard errors clustered at a hospital-level. All models controlled for measures of age; gender; ethnicity; primary ED diagnosis; mode of arrival; attendance category; cause of attendance; referral source; incident location; Elixhauser morbidities; number of emergency admissions, morbidities and ED attendances in the year prior to attendance; bed occupancy; and hospital-month-hour-of-the-week fixed effects. Regressions were estimated using the high-dimensional fixed effects model introduced in Correia (2016).

Abbreviations: AVE, average volume effect; ED, emergency department; GP, general practitioner.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

We also find greater reductions in the quantity and complexity of care in out-of-hours periods, as well as larger increases in patients choosing to leave ED without treatment.

In addition, for patients who choose to remain in ED for treatment, we find different effects on physician's choice of discharge destination. During in-hours periods, physicians respond to increases in attendance volume by substituting inpatient admission and referral to outpatient clinics with discharges with primary care follow-up. However, in out-of-hours periods (when clinics and primary care facilities are closed), we find smaller reductions in admission rates and evidence of substitution between inpatient admission and referral to other healthcare providers (which are still open in these periods).

We also find that increases in attendance volume led to a greater worsening of patient outcomes in out-of-hours periods. During in-hours periods, we find a smaller increase in the probability of any reattendance within 7 days of arrival, and that the increase in unplanned follow-ups is not statistically significant.

8 | RESULTS OF THE ROBUSTNESS CHECKS

Oster causal bounds, assuming proportional selection on unobservables and observables ($\delta = 1$) and assuming either $R_{max} = 1.3R$ (Table 3) or $R_{max} = 1.5R$ (Table A3, Supplementary Appendix 3), exclude zero for all but two outcomes. The exceptions are the probability of admission, where bounds including zero reduces confidence in the statistically significant positive volume effect found in the main analysis, and for mortality within 30 days where the causal bounds including zero is consistent with the statistically insignificant effect found in the main analysis. The causal bounds are also narrow for the majority of outcomes, providing support for a causal interpretation of volume effects. Table 3 also shows that for all outcomes except the probabilities of admission and mortality, assuming $R_{max} = 1.3R$, the degree of selection on unobservables must be significantly greater than that on observables for bias-adjusted volume effects to be zero, with δs ranging from 3.62 (for the number of investigations) to 219.7 (for waiting time for initial assessment > 15 min). For outcomes where the inclusion of covariates increased the magnitude of volume effects, δs are negative, suggesting effects in the main analysis are conservative estimates of the true causal effects.

These results are consistent with the more traditional tests for robustness to omitted variable bias (Supplementary Appendix 3), with results from the balancing regressions indicating that, for the majority of covariates, associations between attendance volume and covariates are not statistically significant and effects relative to the baseline mean of these covariates are ~0% (Table A2). Distributions of volume effects when iteratively removing each severity indicator from the covariate set are also very tightly clustered around the effects from the baseline specification (Figures A2–A5). Finally, effects of attendance volume on the majority of outcomes are either invariant or increase in magnitude when indicators of patient severity are incrementally added as covariates (Figures A6–A9). Controlling for patient sociodemographics dampens the negative effect of volume on the probability of admission. It also dampens the initially small positive effects in the probability of discharge with no follow-up and referral to outpatient clinics, which become small and negative when all covariates are included. For the remaining outcomes, controlling for patient sociodemographics dampens volume effects, but controlling for further severity indicators leads to little change.

These tests indicate that omitted variable bias is unlikely to be having large impacts on our results.

Results are also robust to reestimating volume effects using two-step logistic regression (Table A4, Supplementary Appendix 4), conditional fixed-effect Poisson models (Table A5, Supplementary Appendix 4), and to transforming total ED waiting time using the inverse hyperbolic sine transformation (Table A6, Supplementary Appendix 4).

Results are also similar to baseline estimates when restricting the attendance window from 7 to 3 hours surrounding the hour of arrival (Table A7, Supplementary Appendix 5). For the majority of outcomes, results also remain qualitatively similar when restricting the attendance window to the 3-h period prior to the hour of arrival (Table A8, Supplementary Appendix 5). However, although effects all discharge destinations do not change in sign, they do change in magnitude and statistical significance.

These same changes in effects on discharge destinations are found when restricting estimation samples to the first attendance for each individual within the study period, although effects on other outcomes are virtually identical (Table A9, Supplementary Appendix 6).

TABLE 3 Oster causal bounds for the effects of attendance volume on all indicators of ED performance

Outcome	$\hat{\beta}$	\tilde{R}	$\tilde{\beta}$	\tilde{R}	$R_{max} = 1.3\tilde{R}$	Oster causal bounds	δ where $\beta^* = 0$
Waiting times targets							
>4 h total waiting time	0.0147	0.002	0.0164	0.124	0.161	[0.0164, 0.0170]	<0
>15 min waiting time for initial assessment	0.0169	0.002	0.0169	0.007	0.010	[0.0169, 0.0169]	219.71
>60 min waiting time for treatment	0.0421	0.009	0.0424	0.023	0.030	[0.0424, 0.0425]	<0
Continuous waiting times							
Total waiting time	6.03	0.002	6.91	0.190	0.246	[6.91, 7.18]	<0
Waiting time for initial assessment	1.90	0.003	1.88	0.022	0.029	[1.87, 1.88]	151.83
Waiting time for treatment	6.63	0.011	6.68	0.034	0.044	[6.68, 6.70]	<0
Waiting time from assessment to treatment	4.77	0.007	4.85	0.026	0.033	[4.85, 4.88]	<0
Waiting time from treatment to departure	-0.015	<0.001	0.709	0.191	0.248	[0.709, 0.927]	<0
Quantity and complexity of ED care							
ED re-impursement (£)	-0.674	<0.001	-0.408	0.243	0.315	[-0.408, -0.328]	5.06
Number of investigations	-0.047	<0.001	-0.024	0.316	0.411	[-0.024, -0.018]	3.62
Number of treatments	-0.030	<0.001	-0.021	0.107	0.140	[-0.021, -0.019]	8.18
Left without being seen	0.0053	0.001	0.0048	0.066	0.086	[0.0047, 0.0048]	31.60
Discharge destination							
Admitted	-0.0033	<0.001	-0.0007	0.312	0.406	[-0.0007, 0.00006]	0.93
Discharged with no follow-up	0.0016	<0.001	-0.0002	0.147	0.191	[-0.0007, -0.0002]	<0
Discharged with GP follow-up	0.0007	<0.001	0.0007	0.044	0.057	[0.0007, 0.0007]	28.88
Referred to outpatient clinic	0.0004	<0.001	-0.0002	0.146	0.190	[-0.0003, -0.0002]	<0
Referred to other health care professional or provider	0.0004	<0.001	0.00024	0.029	0.038	[0.00020, 0.00024]	6.79
Patient outcomes							
Unplanned follow-up attendance within 7 days	0.00012	<0.001	0.00011	0.005	0.007	[0.00010, 0.00011]	35.69
Any ED reattendance within 7 days	0.00062	<0.001	0.00079	0.043	0.056	[0.00079, 0.00084]	<0
Mortality within 30 days	-0.00030	<0.001	-0.00004	0.064	0.083	[-0.00004, 0.00004]	0.54

Notes: R^2 values represent the within R^2 : the outcome-variation (after accounting for hospital-month-hour-of-the-week fixed effects) explained by attendance volume and covariates. $\hat{\beta}$ and \tilde{R} represent marginal effects of a one standard deviation increase in attendance volume (~11.5 additional attendances within a 7-h window of a patient's arrival) and within- R^2 from "uncontrolled" regressions of outcomes on attendance volume, and hospital-month-hour-of-the-week fixed effects. $\tilde{\beta}$ and \tilde{R} represent these same estimates from regressions of outcomes on attendance volume, hospital-month-hour-of-the-week fixed effects, and the full set of observed covariates (the baseline specification). β^* represent bias-adjusted volume effects. Causal bounds are estimated assuming $R_{max} = 1.3\tilde{R}$ and proportional selection on observables and unobservables ($\delta=1$). R_{max} is the within- R^2 from a regression of outcomes on attendance volume, fixed effects, and all observed and unobserved covariates.

Abbreviations: ED, emergency department; GP, general practitioner.

9 | DISCUSSION

ED crowding has increased considerably in many health systems. Using data on all attendances at all major EDs in England, this study examines the effects of a demand-side measure of crowding, attendance volume, on indicators of ED performance covering the entire care pathway. Through the use of hospital-month-hour-of-the-week fixed effects, we exploit "unexpected" quasi-exogenous variation in the volume of ED attendances, which is independent

of both the scheduling decisions of hospitals and variations in patient severity caused by patient self-selection by time-of-day.

We find that increases in attendance volume result in substantial increases in waiting times at all key stages of the care pathway. We also find a substantial increase in patients choosing to leave ED without treatment. Even after extended delays, we find that patients attending in high-volume periods face reductions in the quantity and complexity of ED care. Higher attendance volume also leads to small changes in physician choice of discharge destination, with physicians admitting less patients and referring less patients to outpatient clinics in favor of discharging more patients with primary care follow-up and transferring more patients to other health care providers. However, we find little impacts on patient outcomes, with only small increases in ED re-attendances and no increases in mortality.

For almost all outcomes, estimated causal bounds of attendance volume effects are narrow and do not include zero, providing support for a causal interpretation of results. The majority of results were also robust to changes in the estimation method used, the definition of attendance volume, and restrictions to the estimation sample.

Substantial effects on waiting times are consistent with previous studies measuring crowding using variation in actual attendance volumes, which have established delays to prescribing antibiotics (Fee et al., 2007), delays to resuscitation (Hong et al., 2013), and extended total waiting time and waiting time to be seen (Asaro et al., 2007) in periods of high volume. Our effects are however inconsistent with Martins and Filipe (2020), who use similar time-variation in the volume of attendances at a single hospital in Lisbon, Portugal, to instrument crowding measured by ED queue-length, and find that doctors respond to volume-induced queue increases by discharging patients more quickly. This is perhaps surprising given our approach is the reduced form equivalent of their analysis. Cross-country variation in physician preferences may explain these conflicting findings. However, consistent with our findings, Martins and Filipe (2020) also find that physicians respond to crowding by reducing the complexity of diagnoses and treatments.

Increases in waiting times and reductions in the quantity and complexity of care are consistent with physicians' actions being guided by a time-quality trade-off. Our results suggest that physicians allow patient waiting times to rise in response to a positive demand shock, but also ration care in an attempt to avoid substantial increases in waiting times. However, the smaller reduction in quantity/complexity of care relative to the much larger increase in waiting times suggests physicians may be wary of rationing care and less wary of the potential health impacts of extended waits.

We also extend Martins and Filipe (2020) by examining a richer set of discharge destinations. Martins and Filipe (2020) find that increases in demand reduce the rate of admission and increase the rate of discharge with primary care follow-up. We find similar changes in these indicators, but also that demand increases lead to increases in transfers to other hospital providers and a reduction in referral to outpatient clinics; suggesting that physicians do not necessarily favor either "step-up" care or "step-down" care when selecting substitutes for within-ED care. However, a general substitution away from inpatient admission toward out-of-hospital services may reflect an attempt to ration in-hospital services and avoid ED crowding translating into inpatient crowding.

This substitution toward external services may also explain our finding that effects on waiting times are much larger in the earlier stages of the care pathway. We find that increases in waiting time for initial assessment and treatment are large, but that increases in time between treatment initiation and attendance conclusion are relatively small in magnitude. Although this may partially reflect the reduction in the number and complexity of treatments being performed, it may also be due to a lower proportion of patients facing lengthy waits for admission. "Exit block," where patients are unable to be admitted due to a lack of inpatient capacity (Mason, Knowles, & Boyle, 2017), may be a driver of these results. Although the lack of a large increase in waits between treatment and conclusion is evidence against exit block (as exit block is typically associated with lengthy delays prior to admission), our results may be consistent with physicians responding to exit block by substituting inpatient services with care from external services in order to avoid patients facing lengthy waits for admission.

The substantial increase in patients choosing to leave ED without treatment, coupled with substantial increases in waiting times, is consistent with waiting times acting as a rationing mechanism deterring some patients from demanding treatment (Lindsay & Feigenbaum, 1984; Martin & Smith, 1999). These findings contribute to previous studies which have found that volume-induced increases in ED waits lead to reductions in the quantity of treatment demanded (Sivey, 2018), and are consistent with studies finding positive associations between patients leaving without treatment and attendance volume (Mohsin, Young, Ieraci, & Bauman, 2005).

Finally, effects on reattendance rates indicate that physicians are not able to prevent increases in attendance volume translating into a worsening of patient outcomes. This result is consistent with previous studies which have established positive associations between attendance volume and reattendance rates (Miró et al., 1999). Although specific mechanisms are difficult to isolate, higher reattendance rates in periods of high volume could be driven by higher waiting

times for all stages of treatment, lower quantity and complexity of care, inappropriate changes to discharge destination, and/or higher levels of patients incorrectly leaving ED without treatment when ED care was required. Demand-induced reductions in quality of care, due to for example increases in medication errors (Kulstad et al., 2010), may also have also had a direct impact on patient outcomes.

However, despite large increases in both waiting times and patients leaving without treatment, effects on reattendance rates are relatively small in magnitude. This, coupled with the lack of effects on mortality, may suggest that physician (and patient) responses to unexpected demand increases may be sufficient to avoid the most serious of outcomes. Our results may therefore suggest that the patients leaving without treatment are those who would have benefitted little from ED treatment (and for which more suitable care could be provided outside of an ED setting), that physicians ration care or allow waits to increase only for patients where the consequences for health are minimal, and that physicians do not substitute care within the ED for care at alternative healthcare providers inappropriately. The lack of mortality effects found in this study is however generally inconsistent with previous studies (Cha et al., 2011; Hong et al., 2013; Miró et al., 1999), as well as those measuring crowding using throughput measures (Carter et al., 2014; George & Evridiki, 2015). However, these studies have generally focused only on admitted and/or high severity patients, so lack of mortality effects in our study may reflect the rarity of mortality in a whole-population sample and the inability of mortality measures to detect small changes in patient health for low-severity patients.

Our results also provide some evidence that the impacts of attendance volume on physicians' choice of discharge destination may be sensitive to the availability of substitutes. During in-hours periods, physicians respond to increases in attendance volume by substituting inpatient admission and referral to outpatient clinics with discharges with primary care follow-up. However, in out-of-hours periods (when clinics and primary care facilities are closed), we find smaller reductions in admission rates and evidence of substitution between inpatient admission and referral to other health care providers (which are still open in these periods). However, we find no evidence that effects on patients leaving without treatment are higher when substitute services are unavailable. In fact, we find volume-induced increases in leaving without treatment are higher in out-of-hours periods, potentially reflecting a higher opportunity cost of time spent waiting at weekends and in night-time periods. This is somewhat inconsistent with Sivey (2018) who finds demand-induced increases in ED waiting times cause larger reductions in the number of patients treated in weekend periods. However, differences in magnitudes between weekday and weekend periods identified in Sivey (2018) were small, and, consistent with our findings, the largest effects were found in an out-of-hours period (6 p.m.-midnight).

Effects on waiting times, quantity and complexity of care, and patient outcomes are also larger in out-of-hours periods, suggesting that effects of unexpected demand increases are greater on weekends and weekday nights. This may reflect a lower ability of EDs to mitigate capacity shortfalls due to unexpected demand increases by bringing in on-call or temporary staff.

Despite the range of robustness checks conducted in this study, some limitations remain. First, effects on discharge destination, although robust to changing the estimation method, were sensitive to the definitions of attendance volume and the estimation sample, and should therefore be interpreted with more caution. Also, poor coding could partially explain the effects of attendance volume on the quantity of ED care. If EDs code fewer investigations and treatments when volume is unexpectedly high, this would inflate negative effects on these outcomes. Coding is less of a concern for the complexity of care as, due to ED reimbursement being dependent on the most severe investigation and treatment reported, there is a financial incentive for EDs to code correctly even in busy periods.

Our findings add to the debate regarding the potential solutions to the recent decline in performance at English EDs. There has been a sharp decline in adherence to the flagship 4-h total waiting time target, with the proportion of patients meeting the target falling from 94.9% in 2011/12 to 81.4% in 2018/19, well below the 95% standard (NHS England, 2019a). Inadequate capacity and staffing in EDs and exit block, caused by a lack of acute inpatient beds and increases in delayed discharges with a social care cause, have been suggested as reasons (Boyle & Higginson, 2018; Iacobucci, 2018; Torjesen, 2018). In an attempt to improve performance, reforms have committed to an expansion and upskilling of the ED workforce (NHS Improvement, 2017) as well as improved patient management through the introduction of GP-led triage, the development of specialist ED mental health teams, and policies to reduce delayed discharges (NHS England, 2017b, 2017c).

However, much of the policy focus has been on demand reduction, with the introduction of reforms to expand urgent care helplines to direct patients away from EDs, the roll-out of weekend and evening primary care services to provide alternatives to ED (NHS England, 2017), and the integration of primary, secondary and acute care in small geographies, with the aim of actively managing patients to treat illness before urgent care is required (Ham, Alderwick, Dunn, & McKenna, 2017). Recent commentary has questioned the focus on demand reduction (Boyle & Higginson, 2018).

Although this study examines the effects of short-term unexpected fluctuations in attendance volume holding capacity (supply) constant, our results provide predictions about the likely effects of long-run increases in demand for ED care if volume increases are not met by similar increases in available capacity. Our results suggest that if investments are not made to expand ED capacity in line with future increases in attendance volume (e.g., if supply remains constant in the long run), performance on ED waiting times targets are likely to worsen. Policies which either reduce attendance volume (holding capacity fixed), or increase capacity in line with increases in volume, are therefore likely to improve adherence to waiting times targets. However, the ultimate aim of designing policy is to improve patient outcomes. Our results suggest that physician and patient responses are currently able to mitigate most of the impact of demand increases, so although these policies may reduce waits, substantial improvements in patient outcomes may not occur. However, this assumes that physician and patient responses to higher-than-expected demand in the long-run are equivalent to that observed in the short-run. Although physicians may be able to respond “efficiently” if attendances temporarily exceed capacity, it is unclear whether physicians could continue to respond efficiently if demand exceeded capacity permanently. Therefore, in the long-run, investments to reduce attendances (or increase ED capacity in line with increases in attendances) may also be required to avoid a worsening of patient outcomes.

ACKNOWLEDGMENTS

The authors would like to thank the two anonymous reviewers for their very helpful and constructive feedback. This work was supported by funding from The Health Foundation.

CONFLICT OF INTEREST

Other than the funding highlighted above, all authors have no conflicts of interest to declare.

ETHICS STATEMENT

This study used secondary data accessed from NHS Digital. No further ethical approval was required.

ORCID

Alex J. Turner  <https://orcid.org/0000-0003-4139-941X>

Laura Anselmi  <https://orcid.org/0000-0002-2499-7656>

REFERENCES

- Allard, M., Jelovac, I., & Léger, P. T. (2011). Treatment and referral decisions under different physician payment mechanisms. *Journal of Health Economics*, 30(5), 880–893. <https://doi.org/10.1016/j.jhealeco.2011.05.016>
- Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1), 151–184. <https://doi.org/10.1086/426036>
- American College of Emergency Physicians. (2016). Crowding. Retrieved from <https://www.acep.org/Clinical—Practice-Management/Crowding/>
- Anselmi, L., Meacock, R., Kristensen, S. R., Doran, T., & Sutton, M. (2017). Arrival by ambulance explains variation in mortality by time of admission: Retrospective study of admissions to hospital following emergency department attendance in England. *BMJ Quality & Safety*, 26(8), 613–621. <https://doi.org/10.1136/bmjqs-2016-005680>
- Asaro, P. V., Lewis, L. M., & Boxerman, S. B. (2007). The impact of input and output factors on emergency department throughput. *Academic Emergency Medicine*, 14(3), 235–242. <https://doi.org/10.1197/j.aem.2006.10.104>
- Bernstein, S. L., Aronsky, D., Duseja, R., Epstein, S., Handel, D., Hwang, U., & Asplin, B. R. (2009). The effect of emergency department crowding on clinically oriented outcomes. *Academic Emergency Medicine*, 16(1), 1–10. <https://doi.org/10.1111/j.1553-2712.2008.00295.x>
- Boyle, A., & Higginson, I. (2018). This emergency department crisis was predictable—and partly preventable. *British Medical Journal*, 360, k64. <https://doi.org/10.1136/bmj.k64>
- Carter, E. J., Larson, E. L., & Pouch, S. M. (2014). The relationship between emergency department crowding and patient outcomes: A systematic review. *Journal of Nursing Scholarship*, 46(2), 106–115. <https://doi.org/10.1111/jnu.12055>
- Cha, W. C., Shin, S. Do, Cho, J. S., Song, K. J., Singer, A. J., & Kwak, Y. H. (2011). The association between crowding and mortality in admitted pediatric patients from mixed adult-pediatric emergency departments in Korea. *Pediatric Emergency Care*, 27(12), 1136–1141. <https://doi.org/10.1097/PEC.0b013e31823ab90b>
- Chalkley, M., & Malcomson, J. M. (1998). Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics*, 17(1), 1–19. [https://doi.org/10.1016/S0167-6296\(97\)00019-2](https://doi.org/10.1016/S0167-6296(97)00019-2)
- Chen, L. M., Farwell, W. R., & Jha, A. K. (2009). Primary care visit duration and quality: Does good care take longer? *Archives of Internal Medicine*, 169(20), 1866–1872. <https://doi.org/10.1001/archinternmed.2009.341>

- Correia, S. (2016). *Linear models with high-dimensional fixed effects: An efficient and feasible estimator*. Retrieved from <http://scoreia.com/research/hdfe.pdf>
- Department for Communities and Local Government. (2015). *English indices of deprivation 2015*. Retrieved from <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>
- Department of Health. (2010). *A&E clinical quality indicators*. Retrieved from http://webarchive.nationalarchives.gov.uk/20130105030902/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_122868
- Ellis, R. P., & McGuire, T. G. (1986). Provider behavior under prospective reimbursement. *Journal of Health Economics*, 5(2), 129–151. [https://doi.org/10.1016/0167-6296\(86\)90002-0](https://doi.org/10.1016/0167-6296(86)90002-0)
- Fee, C., Weber, E. J., Maak, C. A., & Bacchetti, P. (2007). Effect of emergency department crowding on time to antibiotics in patients admitted with community-acquired pneumonia. *Annals of Emergency Medicine*, 50(5), 501–509. <https://doi.org/10.1016/j.annemergmed.2007.08.003>
- George, F., & Evridiki, K. (2015). The effect of emergency department crowding on patient outcomes. *Health Science Journal*, 9(1), 1–6.
- Godager, G., & Wiesen, D. (2013). Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. *Journal of Health Economics*, 32(6), 1105–1116. <https://doi.org/10.1016/j.jhealeco.2013.08.008>
- Gruber, J., Hoe, T., & Stoye, G. (2018). Saving lives by tying hands: The unexpected effects of constraining health care providers (No. 24445) NBER Working Paper Series. <https://doi.org/10.3386/w24445>
- Gutacker, N., Bloor, K., & Cookson, R. (2015). Comparing the performance of the Charlson/Deyo and Elixhauser comorbidity measures across five European countries and three conditions. *European Journal of Public Health*, 25, 15–20. <https://doi.org/10.1093/eurpub/cku221>
- Guttmann, A., Schull, M., Vermeulen, M., & Stukel, T. (2011). Association between waiting times and short term mortality and hospital admission after departure from emergency department: Population based cohort study from Ontario, Canada. *British Medical Journal*, 342(1), d2983.
- Ham, C., Alderwick, H., Dunn, P., & McKenna, H. (2017). *Delivering sustainability and transformation plans: From ambitious proposals to credible plans*. Retrieved from https://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/STPs_proposals_to_plans_Kings_Fund_Feb_2017_0.pdf
- Health Education England, NHS England, NHS Improvement, & Royal College of Emergency Medicine. (2017). *Securing the future workforce for emergency departments in England*. Retrieved from https://improvement.nhs.uk/documents/1826/Emergency_department_workforce_plan_-_111017_Final.3.pdf
- Herbert, A., Wijlaars, L., Zylbersztejn, A., Cromwell, D., & Hardelid, P. (2017). Data resource profile: Hospital episode statistics admitted patient care (HES APC). *International Journal of Epidemiology*, 46(4), 1093–1093i. <https://doi.org/10.1093/ije/dyx015>
- Higginson, I. (2011). Emergency department crowding. *Emergency Medicine Journal*, 29, 437–443. <https://doi.org/10.1136/emered-2011-200532>
- Hong, K. J., Shin, S. D., Song, K. J., Cha, W. C., & Cho, J. S. (2013). Association between ED crowding and delay in resuscitation effort. *American Journal of Emergency Medicine*, 31(3), 509–515. <https://doi.org/10.1016/j.ajem.2012.09.029>
- Hoot, N. R., & Aronsky, D. (2008). Systematic review of emergency department crowding: Causes, effects, and solutions. *Health Policy and Clinical Practice*, 52(2), 126–136. <https://doi.org/10.1016/j.annemergmed.2008.03.014>
- Hwang, U., Richardson, L. D., Sonuyi, T. O., & Morrison, R. S. (2006). The effect of emergency department crowding on the management of pain in older adults with hip fracture. *Journal of the American Geriatrics Society*, 54(2), 270–275. <https://doi.org/10.1111/j.1532-5415.2005.00587.x>
- Iacobucci, G. (2018). Patients urged to lobby MPs over emergency department crisis. *British Medical Journal*, 360, k1129. Retrieved from <https://doi.org/10.1136/bmj.k1129>
- Kulstad, E., Sikka, R., Sweis, R. T., Kelley, K. M., & Rzechula, K. H. (2010). ED overcrowding is associated with an increased frequency of medication errors. *American Journal of Emergency Medicine*, 28(3), 304–309. <https://doi.org/10.1016/j.ajem.2008.12.014>
- Lindsay, C. M., & Feigenbaum, B. (1984). Rationing by waiting lists. *American Economic Review*, 74(3), 404–417. <https://doi.org/10.2307/1804016>
- Liu, T., Ma, C., & to, A. (2013). Health insurance, treatment plan, and delegation to altruistic physician. *Journal of Economic Behavior and Organization*, 85(1), 79–96. <https://doi.org/10.1016/j.jebo.2012.11.002>
- Martin, S., & Smith, P. C. (1999). Rationing by waiting lists: An empirical investigation. *Journal of Public Economics*, 71(1), 141–164. [https://doi.org/10.1016/S0047-2727\(98\)00067-X](https://doi.org/10.1016/S0047-2727(98)00067-X)
- Martins, B., & Filipe, L. (2020). Doctors' response to queues: Evidence from a Portuguese emergency department. *Health Economics*, 29(2), 123–127.
- Mason, S., Knowles, E., & Boyle, A. (2017). Exit block in emergency departments: A rapid evidence review. *Emergency Medicine Journal*, 34(1), 46–51. <https://doi.org/10.1136/emered-2015-205201>
- Miró, O., Antonio, M. T., Jiménez, S., De Dios, A., Sánchez, M., Borrás, A., & Millá, J. (1999). Decreased health care quality associated with emergency department overcrowding. *European Journal of Emergency Medicine*, 6(2), 105–107.
- Mohsin, M., Young, L., Ieraci, S., & Bauman, A. E. (2005). Factors associated with walkout of patients from New South Wales hospital emergency departments, Australia. *Emergency Medicine Australasia*, 17(5–6), 434–442.
- Monitor (2016). *National tariff payment system*. 2016/17. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/509697/2016-17_National_Tariff_Payment_System.pdf

- Morley, C., Unwin, M., Peterson, G. M., Stankovich, J., & Kinsman, L. (2018). Emergency department crowding: A systematic review of causes, consequences and solutions. *PLoS ONE*, *13*(8), e0203316. <https://doi.org/10.1371/journal.pone.0203316>
- Morris, Z., Boyle, A., Beniuk, K., & Robinson, S. (2012). Emergency department crowding: Towards an agenda for evidence-based intervention. *Emergency Medicine Journal*, *29*, 460–466. <https://doi.org/10.1136/emj.2010.107078>
- NHS Digital. (2016). *HES data dictionary: Accident and emergency - accident and emergency (A&E) hospital episode statistics (HES) data dictionary*. Retrieved from <http://content.digital.nhs.uk/media/19424/AE-DD-Final-Doc/pdf/DD-AE-V7.pdf>
- NHS Digital. (2017a). *Hospital episode statistics*. Retrieved from <http://content.digital.nhs.uk/hes>
- NHS Digital. (2017b). *Methodology for producing the A&E clinical quality indicators from provisional Hospital Episode Statistics (HES) data*. Retrieved from [https://files.digital.nhs.uk/91/B89F13/Provisional Accident Emergency Quality Indicators for England - Indicator Generation Methodology - NHS Digital.pdf](https://files.digital.nhs.uk/91/B89F13/Provisional%20Accident%20Emergency%20Quality%20Indicators%20for%20England%20-%20Indicator%20Generation%20Methodology%20-%20NHS%20Digital.pdf)
- NHS England. (2017a). *A&E attendances and emergency admissions*. Retrieved from <https://www.england.nhs.uk/statistics/statistical-work-areas/ae-waiting-times-and-activity/>
- NHS England. (2017b). *Delayed transfers of care*. Retrieved from <https://www.england.nhs.uk/statistics/statistical-work-areas/delayed-transfers-of-care/>
- NHS England. (2017c). *Next steps on the NHS five year forward view*. Retrieved from <https://www.england.nhs.uk/wp-content/uploads/2017/03/NEXT-STEPS-ON-THE-NHS-FIVE-YEAR-FORWARD-VIEW.pdf>
- NHS England. (2019a). *A&E attendances and emergency admissions 2018-19*. Retrieved from <https://www.england.nhs.uk/statistics/statistical-work-areas/ae-waiting-times-and-activity/ae-attendances-and-emergency-admissions-2018-19/>
- NHS England. (2019b). *A&E attendances and emergency admissions 2019-20*. Retrieved from <https://www.england.nhs.uk/statistics/statistical-work-areas/ae-waiting-times-and-activity/ae-attendances-and-emergency-admissions-2019-20/>
- Nikolova, S., Harrison, M., & Sutton, M. (2016). The impact of waiting time on health gains from surgery: Evidence from a national patient-reported outcome dataset. *Health Economics*, *25*(8), 955–968. <https://doi.org/10.1002/hec.3195>
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204. <https://doi.org/10.1080/07350015.2016.1227711>
- Parkinson, B., Meacock, R., & Sutton, M. (2019). How do hospitals respond to price changes in emergency departments? *Health Economics*, *28*(7), 830–842. <https://doi.org/10.1002/hec.3890>
- Pearson, S. D., Goldman, L., Orav, E. J., Guadagnoli, E., Garcia, T. B., Johnson, P. A., & Lee, T. H. (1995). Triage decisions for emergency department patients with chest pain - do physicians' risk attitudes make the difference? *Journal of General Internal Medicine*, *10*(10), 557–564. <https://doi.org/10.1007/BF02640365>
- Pines, J. M., Iyer, S., Disbot, M., Hollander, J. E., Shofer, F. S., & Datner, E. M. (2008). The effect of emergency department crowding on patient satisfaction for admitted patients. *Academic Emergency Medicine*, *15*(9), 825–831. <https://doi.org/10.1111/j.1553-2712.2008.00200.x>
- Pines, J. M., Joshua, A., Isserman, J., Demian Szyld, D., Dean, A., McCusker, C., & Hollander, J. (2010). The effect of physician risk tolerance and the presence of an Observation unit on decision making for ED patients with chest pain. *American Journal of Emergency Medicine*, *28*, 771–779.
- Pines, J. M., Pollack, C. V., Jr, Diercks, D. B., Chang, A. M., Shofer, F. S., & Hollander, J. E. (2009). The association between emergency department crowding and adverse cardiovascular outcomes in patients with chest pain. *Academic Emergency Medicine*, *16*(7), 617–625. <https://doi.org/10.1111/j.1553-2712.2009.00456.x>
- Schull, M. J., Vermeulen, M., Slaughter, G., Morrison, L., & Daly, P. (2004). Emergency department crowding and thrombolysis delays in acute myocardial infarction. *Annals Emergency Medicine*, *44*(6), 577–585. <https://doi.org/10.1016/j.annemergmed.2004.05.004>
- Sivey, P. M. (2018). Should I stay or should I go? Hospital emergency department waiting times and demand. *Health Economics*, *27*(3), e30–e42. <https://doi.org/10.1002/hec.3610>
- Stang, A. S., Crotts, J., Johnson, D. W., Hartling, L., & Guttman, A. (2015). Crowding measures associated with the quality of emergency department care: A systematic review. *Academic Emergency Medicine*, *22*(6), 643–656. <https://doi.org/10.1111/acem.12682>
- Sun, Y., Heng, B. H., Tay, S. Y., & Tan, K. B. (2015). Unplanned 3-day re-attendance rate at Emergency Department (ED) and hospital's bed occupancy rate (BOR). *International Journal of Emergency Medicine*, *8*(1). Retrieved from <https://intjem.biomedcentral.com/articles/10.1186/s12245-015-0082-3>
- The King's Fund. (2017a). *NHS hospital bed numbers: Past, present, future*. Retrieved from <https://www.kingsfund.org.uk/publications/nhs-hospital-bed-numbers>
- The King's Fund. (2017b). *What's going on in A&E? The key questions answered*. Retrieved from <https://www.kingsfund.org.uk/projects/urgent-emergency-care/urgent-and-emergency-care-mythbusters>
- Torjesen, I. (2018). Latest waiting time figures for emergency departments in England are worst on record. *British Medical Journal*, *361*, k1658. <https://doi.org/10.1136/bmj.k1658>
- Trendwatch. (2015). *Trends affecting hospitals and health systems*. Retrieved from <http://www.aha.org/research/reports/tw/chartbook/2015/table3-3.pdf>
- Woodworth, L., & Holmes, J. F. (2020). Just a minute: The effect OF emergency department wait time ON the cost OF care. *Economic Inquiry*, *58*(2), 698–716. <https://doi.org/10.1111/ecin.12849>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Turner AJ, Anselmi L, Lau YS, Sutton M. The effects of unexpected changes in demand on the performance of emergency departments. *Health Economics*. 2020;29:1744–1763. <https://doi.org/10.1002/hec.4167>