

## SAVING LIVES BY TYING HANDS: THE UNEXPECTED EFFECTS OF CONSTRAINING HEALTH CARE PROVIDERS

Jonathan Gruber, Thomas P. Hoe, and George Stoye\*

*Abstract*—We study how emergency department (ED) doctors respond to incentives to reduce wait times. We use bunching techniques to study an English policy that imposed strong incentives to treat patients within four hours. The policy reduced time spent in the ED by 21 minutes for affected patients yet caused doctors to increase treatment intensity and admit more patients. We find a striking 14% reduction in mortality. Analysis of patient severity and hospital crowding strongly suggests it is the wait time reduction that saves lives. We conclude that, despite distorting medical decisions, constraining ED doctors can induce cost-effective reductions in mortality.

### I. Introduction

PERHAPS the most complicated node of health delivery in any modern health care system is the emergency department (ED). Patients arrive with a wide array of different problems. ED nurses and physicians must quickly assess where patients should slot in what can be a very large queue, deciding almost instantly who needs to be treated right away and who can wait. Ultimately these providers need to decide whether those going to the ED are to be admitted to the hospital or sent home—a decision that can, in many instances, have life or death consequences.

Despite its critical role, EDs often face budgetary pressures and a shortfall in resources. These pressures have been especially acute in recent years, with ED performance having been described as an international crisis in several developed economies (Hoot & Aronsky, 2008). Practicing doctors are especially vocal, referring to “battlefield medicine” and “third-world conditions” caused by ED overcrowding in England.<sup>1</sup> Alongside these tensions, EDs are increasingly facing public pressure to advertise and reduce their wait

times. U.S. cities are replete with digital billboards highlighting wait times at local EDs (more often called emergency rooms or ERs there), and other nations use regulatory and financial tools to reward reductions, or penalize increases, in wait times.

Many are concerned that external pressures on wait times could reduce the ability of EDs to maximize care quality. At the same time, however, it is not clear that ED personnel would maximize patient quality in the absence of such pressures. Emergency departments are not directly compensated for shortening wait times. Moreover, although health-maximizing ED personnel may internalize the costs of waiting to the extent that they impact patient outcomes, this may only be partial if physicians have incomplete knowledge or are imperfect agents for their patients. Theoretical ambiguities such as this have motivated a growing number of empirical studies of hospital production in the ED setting (Chan, 2016, 2018; Silver, 2021).

In this paper, we provide new evidence on the impacts of regulating doctors in the ED—and, in particular, putting pressure on them to make decisions more quickly—on treatment decisions and patient outcomes. To do this, we use the “four-hour wait” policy in England. This policy was first announced in 2000 as part of a wide-ranging set of government pledges to decrease wait times for different types of care and came into force in all English public hospitals in 2004.<sup>2</sup> The policy sets arbitrary targets for wait times, with 95% of all patients required to be treated within four hours of arrival.<sup>3</sup> The ability of hospitals to meet this target became an important part of overall hospital evaluation in England, with managers in some cases losing their jobs because of poor wait time performance. In addition, strong financial penalties were associated with breaching the target: hospitals were penalized by an amount that was more than twice the average revenue of an ED patient, and total fines for missing ED and elective wait time targets were equivalent to a third of hospital deficits.<sup>4</sup>

<sup>2</sup>Other targets included maximum limits on wait times for elective surgery.

<sup>3</sup>The initial target stated that 98% of patients should be treated within four hours, but this level was reduced to its current level of 95% in 2010.

<sup>4</sup>English hospitals have no other financial incentive to shorten wait times or monitor the impacts on patient outcomes. Hospitals receive payments that vary by ED diagnosis group but not by wait time or health outcome.

Received for publication November 6, 2019. Revision accepted for publication February 10, 2021. Editor: Brian A. Jacob.

\*Gruber: MIT and NBER; Hoe: Cornell University and Institute for Fiscal Studies; Stoye (corresponding author): Institute for Fiscal Studies.

We thank Richard Blundell, Aureo de Paula, Eric French, Peter Hull, and Henrik Kleven for useful comments, as well as seminar participants at the Institute for Fiscal Studies, the Kellogg Healthcare Markets Conference 2019, MIT, NASMES 2019, and UCL. The authors thank NHS Digital and the Office for National Statistics for access to the Hospital Episode Statistics and official mortality statistics under data sharing agreement CON-205762-B8S7B. Hoe and Stoye gratefully acknowledge financial support from the UK Economic and Social Research Council through the Centre for the Microeconomic Analysis of Public Policy (CPP) at IFS (ES/M010147/1).

A supplemental appendix is available online at [https://doi.org/10.1162/rest\\_a\\_01044](https://doi.org/10.1162/rest_a_01044).

<sup>1</sup>[https://www.nytimes.com/2018/01/03/world/europe/uk-national-health-service.html?smid=tw-share&\\_r=0](https://www.nytimes.com/2018/01/03/world/europe/uk-national-health-service.html?smid=tw-share&_r=0).

Despite this focus on the target, little consistent evidence is available from either the United Kingdom or other nations that have introduced wait time targets on their impact on patient costs and health outcomes. This is because the policies are generally introduced nationwide, with no “hold-out” or control populations, making it impossible to apply quasi-experimental methods such as difference-in-difference estimation. An additional challenge in the case of the English policy is that no systematic data on wait times are available before the policy was introduced in 2004.

We therefore take a different approach to estimate the effects of the policy on treatments, costs, and patient outcomes. We apply the bunching techniques that have been used widely in other contexts (see Kleven, 2016) to analyze wait times and outcomes using administrative hospital data from 2011 to 2013, a period when the policy was already in place. This approach allows us to model how the four-hour target impacts wait times, costs, and outcomes, conditional on the underlying hospital technology in place to monitor patient wait times without using prepolicy data. That is, we estimate here the short-term impact of changing wait times but hold constant the underlying technological changes that might be associated with the introduction or removal of a wait time target and the prioritization of patients treated in the ED. This counterfactual focuses attention on the impact of incentives rather than technology adoption or on broader changes to the way that hospitals treat ED patients.

We initially examine the distribution of wait times around the four-hour target, where we define “wait times” as total time spent in the ED (including time being examined and treated) consistent with the definition of the policy. We find a very large spike at four hours. We then estimate counterfactual distributions of wait times to measure the effect of the four-hour policy. We estimate that, relative to the counterfactual, the target led wait times to be 21 minutes (8%) lower for patients affected by the policy, and for those patients that move from after to before the four-hour position, the wait time reductions are large and average 59 minutes.

The regulations may also change the treatments provided by doctors and the outcomes of their patients. For example, doctors may order fewer tests or treat patients less intensely as a result of the policy, which could have negative effects on health outcomes. On the other hand, the associated reduction in wait times may allow patients to receive hospital care more quickly, which could be beneficial for patient health. We therefore use the data to also study the impact of the policy on patient treatment and outcomes. Without preperiod data and exogenous variation in policy effects across hospitals, we cannot directly use data on treatments and outcomes to identify policy effects, but we argue that under a set of testable assumptions we can directly identify policy effects from bunching at the four-hour target.

Plotting these treatments and health outcomes conditional on wait times reveals spikes just before the four-hour wait time. We then decompose these spikes into two separate channels. First, we have a “composition effect.” If the target causes

patients to be moved from later to earlier in the wait time distribution, and patient characteristics also vary across this distribution, then the observed change in outcomes before the four-hour target will in part reflect this movement of patients. For example, admission probability is increasing with wait time (as more severe patients undergo more testing and treatment in the ED before admission). Moving patients from just after four hours to just before will increase the average admission probability of patients seen before the target, even if the target has no impact on the admission probability of each individual patient. Second, there may be an additional “treatment change effect” if the target itself leads to direct changes in treatment received by patients or in their outcomes.

To separately identify the treatment change effect, we estimate a “composition-adjusted counterfactual outcome” by imposing a “no-selection” assumption on the distribution of patients that obtain shorter wait times because of the policy. This assumes that patients who are moved forward as a result of the wait time target are representative of those who are not. Under this assumption, we can use observed outcomes of patients treated just after four hours to adjust the observed outcomes of patients treated just before four hours for these compositional changes. Comparing these “composition-adjusted counterfactual outcomes” with observed outcomes therefore provides an estimate of the “treatment change effects” of the policy.

We can test this “no-selection” assumption directly using patient characteristics such as age, sex, and past health status. The hospital cannot change these variables at the time of the ED visit, and so by definition, any observed spikes in these outcomes are due solely to a composition effect (i.e., the treatment change effect is zero). Consistent with our assumption, we show along multiple dimensions little meaningful difference between patients who are moved forward and not. In the rare cases where there does appear to be non-random movement of patients, the evidence suggests patients who experience wait time reductions are slightly more severe than those who do not. However, these differences are very small in magnitude: we show the observed distribution of past comorbidities looks very similar to simulated data with random selection, whereas selection would create very large and obvious distortions.

Our analysis also relies on a “local effects” assumption. This assumes that the wait time and treatments of patients outside of an “exclusion window” around the four-hour mark are unaffected by the target. This would be violated if doctors substitute resources away from patients in the early part of the wait time distribution to reduce waits for patients in danger of breaching the target. We argue that institutional factors make such behavior unlikely, and we present a range of empirical tests to support this. We also show that although the exact magnitude of our estimates is sensitive to some choices of parameters used in the estimation, our overarching conclusions are extremely robust.

We estimate a significant treatment change effect of the English policy. We find more intensive testing of patients in

the ED, leading to a modest rise in ED costs. We also find a significant increase in hospital admissions as a means of meeting the target, with corresponding reductions in those discharged to home. Among those marginal admits, inpatient resource use is insignificant, suggesting that such admissions were just placeholders to meet the four-hour target. These admissions were not costless, however, and we estimate that inpatient payments from the government to hospitals rose by roughly 5% because of the target.

Most interestingly, we find significant improvements in patient outcomes associated with the four-hour policy. We estimate that 30-day patient mortality falls by 14% among patients who are impacted by the wait time change, a very sizeable positive effect. This effect falls slightly over time while baseline mortality rises, so that by one year after ED admission this amounts to a 3% mortality reduction, which is still quite large.

We then turn to understanding the mechanism behind the improvements in patient outcomes that we observe. To do so we exploit heterogeneity across patient groups affected along different margins. The first is patients of different severity: across severity groups, the four-hour policy is associated with differential impacts on wait times but not admission probabilities. The second is patients facing different levels of crowding of the inpatient department when they arrive at the ED: across different levels of crowding, the target is associated with differential impacts on admission probabilities but little variation in the wait time impacts. We show that the estimated mortality effect varies strongly across patient severity but not across inpatient crowding. Taken together, this evidence suggests that the wait time mechanism, and not the admissions mechanism, is driving our mortality effect. As a final check, we examine whether mortality reductions occur among patients with potentially time-sensitive conditions, and we find that the majority of reductions are found among conditions that are known to benefit from rapid treatment.

We contribute to two literatures. First, a growing literature has begun documenting features of hospital production relevant for incentive setting (Chan, 2016, 2018; Silver, 2021). Chan (2016) and Chan (2018), for example, study how ED physicians respond to team environments and work schedules, and Silver (2021) studies peer effects in the ED. A medical literature has also documented robust correlations between mortality rates and measures of ED crowding and wait times (Hoot & Aronsky, 2008). Our contribution is to show how ED production is affected when doctors are put under pressure to make decisions quicker. We find the wait time policy generated cost-effective mortality improvements through reduced wait times but at the expense of distorting medical decisions. These findings are consistent with the medical literature and highlight that ED wait times are an important input to the health production process. The findings also illustrate how constraining health care providers through regulatory interventions can improve health outcomes even in the presence of significant distortions.

The second contribution we make is to the literature using bunching estimators. From its origins in the tax setting

(Saez, 2010; Chetty et al., 2013; Kleven & Waseem, 2013), these estimators have now been deployed in other settings such as health insurance (Einav, Finkelstein, & Schrimpf, 2015, 2017; Einav, Finkelstein, & Polyakova, 2018), mortgage markets (Best & Kleven, 2018), and education (Diamond & Persson, 2016). We apply these estimators in a health care provision setting, adapting them to study outcomes indirectly affected by a discontinuity in the incentives associated with the running variable, and devise new empirical tests to evaluate the credibility of the bunching assumptions required in our context.

Our paper proceeds as follows. Section II provides background information on emergency care in England and on the four-hour target policy. Section III describes the data. Section IV sets out our methodology, and section V examines our identifying assumptions. Section VI describes our results for wait times, treatment decisions, and health outcomes. Section VII explores heterogeneity and mechanisms. Section VIII concludes.

## II. Background

### A. Emergency Care in England

Emergency care in England is publicly funded and is free for all residents. No private market exists for emergency care. The majority of care is provided at EDs attached to large, publicly owned hospitals. These major EDs are physician-led providers of 24-hour services, based in specifically built facilities to treat emergency patients that contain full resuscitation facilities. In 2011/2012, 9.2 million patients made 13.6 million visits to 174 EDs. In addition, 2.1 million patients made an additional 2.7 million visits to specialist emergency clinics and “walk in” or minor injury centers where simple treatment is provided for less serious diagnoses; as discussed below, we exclude patients from these centers because of the minor nature of their injuries and our results are unaffected if they are included.

EDs provide immediate care to patients. Hospitals are reimbursed by the government for the care they provide, receiving a nationally fixed payment for providing certain types of treatment.<sup>5</sup> In 2015/2016, 11 tariffs were used for ED treatment depending on the severity of the patient and the type of treatments administered.<sup>6</sup> These tariffs ranged from \$77 to \$272 (£57 to £200) per visit.<sup>7</sup> Revenue from the ED accounted for 5.3% of total hospital income in 2015/2016.<sup>8</sup>

<sup>5</sup>Treatments are assigned to a Healthcare Resource Group (HRG), similar to diagnosis-related groups (DRGs) in the United States, with national tariffs for each HRG announced each year by the Department of Health.

<sup>6</sup><https://www.gov.uk/government/publications/confirmation-of-payment-by-results-pbr-arrangements-for-2012-13>.

<sup>7</sup>Figures in 2017/2018 U.S. dollars. Figures are deflated using the U.K. GDP deflator and then converted from sterling to dollars using an exchange rate of 1 GBP:1.35 USD (U.S. Treasury, data last accessed on Dec. 31, 2017, <https://www.fiscal.treasury.gov/fsreports/rpt/treasRptRateExch/currentRates.htm>).

<sup>8</sup>Figures calculated from the 2015/2016 U.K. Department of Health Reference Costs. See <https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016>.

Treatment in the ED follows one of two pathways depending upon the method of arrival. Nonambulance patients register at reception upon arrival, where they must identify themselves and provide basic details of their condition. Patients then undergo an initial assessment to establish the seriousness of their condition. This triage process is carried out by a specialist triage nurse or doctor and includes taking a medical history and, where appropriate, conducting a basic physical examination of the patient. Patients are then prioritized according to severity.

Alternatively, patients can arrive at the ED by ambulance following an emergency callout. In 2011/2012, 29.4% of ED patients arrived by ambulance. For these patients, ambulance staff collect medical details en route and report these details to hospital staff upon arrival.<sup>9</sup> These details feed into a separate triage process, where patients are categorized by their severity.

These triage processes sort patients into “minor” and “major” cases. Minor cases require relatively simple treatment and are often treated quickly. Major cases are often those who arrive by ambulance, although some exceptions to this are found (e.g., a patient with chest pain may arrive independently at the hospital). Major cases will receive treatment more quickly, as they often present with more severe symptoms but usually require more treatment and examinations within the ED and are therefore likely to spend longer in the ED. Treatment of the two types often requires the use of different resources (including staff and machines), and in most large hospitals, treatment for minor conditions will take place in a separate part of the ED (e.g., in the hospital’s “urgent care center”). In particular, senior staff spend most of their time treating major cases, with little interaction with minor cases (except to sign off admission or discharge decisions by junior staff).

Following triage, patients are placed into a queue on the basis of their severity and time of arrival. Patients are not aware of their position in the queue. Patients are assigned to individual doctors as they become available. These doctors will carry out a series of further examinations and tests. The nature of these depends on the symptoms presented by the patients and ranges from physical examinations to tests such as x-rays or MRI scans. Patients can also receive treatment in the ED, ranging from sutures to resuscitation, before being admitted for further treatment in an inpatient ward or discharged from the hospital.<sup>10</sup>

### B. *The Four-Hour Target*

All public hospitals with EDs in England are subject to a wait time target. This target specifies that 95% of ED patients must be admitted for further inpatient treatment, discharged, or transferred to another hospital within four hours of their

arrival. Although the target is officially a “wait time” target, the definition employed—which includes the time being examined and treated in the ED—corresponds more precisely to the total time a patient spends physically in the ED. We use the terminology and definition of “wait times” consistent with the policy throughout this paper. The target level was initially set at 98% when it was first introduced in December 2004, before being relaxed to its current level (95%) in November 2010.<sup>11</sup>

This target is important to hospitals in two ways. First, the target is widely used by policy makers and the media as a measure for the wider performance of the public health service in England.<sup>12</sup> Hospital managers who consistently fail to meet this target are likely to be fired and therefore have a strong incentive to organize emergency care in a way that minimizes the number of patients who take more than four hours to treat.

Second, hospitals face significant financial incentives to meet the target. As the target came into force between March 2004 and March 2005, hospitals were offered payments (to be used only for hospital investment) if they met the target level early (National Audit Office, 2004). In recent years, significant financial penalties have been imposed for missing the target. In 2011/2012, hospitals were fined \$300 (£220) for every patient who failed to be treated within four hours if the hospital missed the overall 95% target during that week.<sup>13</sup> This compares to an average payment of just over \$140 (£100) per patient in the same year. In 2015, a report commissioned by a number of hospitals indicated that public hospitals paid \$325 million (£250 million) in fines because of missed performance targets (including the four-hour target), with total penalties equal to around a third of the average deficit of public hospitals in that year.<sup>14</sup>

Hospital staff therefore face pressure from hospital management to meet the target. As a result, the organization of EDs has changed significantly since the target was introduced.<sup>15</sup> This includes the use of new IT systems tracking patient wait times in real time. The exact systems vary by hospital but will indicate when patients reach particular waiting thresholds (e.g., three hours) and alert physicians (e.g., through changing the color of something on the computer screen).

<sup>11</sup>Interviews with hospital staff and regulators suggest that it is the “four-hour” component that matters rather than the absolute level of the target. Hospitals attempt to meet the target on a daily basis, aiming to achieve the highest proportion possible. This suggests certain behaviors (e.g., relaxing or improving performance in later parts of the reporting period) are unlikely.

<sup>12</sup>See <http://www.mirror.co.uk/news/uk-news/ae-crisis-exposed-only-three-9801509>.

<sup>13</sup>This penalty was decreased to \$170 (£125) in 2015.

<sup>14</sup><https://www.theguardian.com/society/2016/mar/29/nhs-bosses-slam-600m-hospital-fines-over-patient-targets>.

<sup>15</sup>Interviews with senior members of the Emergency Care Improvement Programme, a clinically led program to improve the performance of EDs, clearly describe significant changes to ED technology since the target was introduced. One manager claimed that the target “is the most monitored part of the entire health care system with software specifically designed for it.”

<sup>9</sup>Ambulance staff also provide some emergency treatment in the ambulance when required.

<sup>10</sup>See appendix B for further details on the range of ED treatments and investigations.

### C. *How Do EDs Respond to the Target in the Short Run?*

Setting aside these longer-run changes to ED organizations, the target incentivizes doctors to speed up treatment for at least some patients. This could be achieved in two different ways, which will have important implications for the validity of our identification strategy.

First, doctors may proceed with treatment as they would do in the absence of the target and change their behavior only if patients wait long enough so as to begin to approach the target. In this case, as wait times exceed a certain point, EDs may speed up treatment by reducing the number of examinations or tests conducted in the ED (either pushing these into inpatient treatment after admission or discharging patients with less information), reducing the waiting time between receiving results and implementing treatment decisions, or reallocating senior doctors to make clinical decisions quickly about long-wait patients. This approach would alter only the wait times of those approaching the four-hour mark. We assume that this is the case in our empirical methodology and discuss the implications of this in detail in section IV.

Alternatively, EDs may more fundamentally change the way that patients are treated and prioritized. For example, doctors may substitute resources (such as time) away from minor patients with otherwise short waits to concentrate on major patients who are more likely to breach the target. More broadly, doctors may try to speed up the treatment of all patients in anticipation of the target. In both cases, the target would have implications for the wait times of patients across the wait time distribution (not just those approaching the target).

In section VA, we provide three pieces of empirical evidence that suggest little wholesale change to the way in which all patients are treated as a result of the target. In addition, a number of institutional details suggest that such dynamic behavior by doctors and managers in the ED is unlikely.

Substitution of resources between patients is likely to occur only if this makes it easier for hospitals to meet the target. Two factors limit the extent to which this is the case. First, little scope is at hand for substitution between patients because no incentive exists to substitute between patients that are not within the same four-hour window. Our analysis aggregates patients across many hospitals, days, and time periods. However, on average only 33 patients arrive at a hospital within a given four-hour window. This limits the potential extent to substitute effort or other resources across patients.

Second, ED staff are generally separately assigned to minor or major units within the ED, and this physical separation limits the prospect of substitution between early and late exit patients. It is, of course, possible that as a major unit becomes busy, staff could be diverted from the minor unit to assist. In this case the presence of the target may incentivize more staff to be moved to treat major cases than would otherwise occur. We test for this directly in section VA and find no obvious evidence of substitution in these cases.

More broadly, changes in behavior will be limited by the willingness of doctors to alter the way in which they treat patients and the necessity of this behavioral change in meeting the target. Two further factors are likely to restrict this willingness. First, hospitals are already attempting to maximize an objective function that, at least in part, contains patient mortality. This will naturally place limits on their willingness to alter the way these patients are treated. For example, patients with clear and life-threatening injuries (e.g., knife wounds) will always be treated immediately, and for a similar length of time, irrespective of the target. Similarly, patients with very minor injuries will always be sent home shortly after initial assessment. These unambiguously high- and low-severity patients are likely to account for a significant proportion of exits from the ED in the early part of the wait time distribution.

Second, it is important to note that the definition of wait times in our setting (total time in the ED) allows for the possibility that physicians can shorten a patient's wait time by simply admitting them as an inpatient. Shortening wait times therefore does not require the physician to alter the way they treat a patient until they begin to approach the target, and so it provides little incentive to shorten or lengthen waits for patients who are unlikely to approach the target.<sup>16</sup>

Taken together, these factors suggest that hospitals are unlikely to change the way they treat all patients in light of the target, with any changes restricted to patients treated further up the wait time distribution.

## III. Data

### A. *Hospital Episodes Statistics*

Our primary source of data are the Hospital Episode Statistics (HES). These contain the administrative records of all visits to public hospitals between April 2011 and March 2013 and include information on both ED visits and inpatient admissions.

The ED data record treatment at the visit level and include information on the precise time of arrival, initial treatment, and the admission decision. We define ED "wait times" as total time spent in the ED, consistent with the policy definition. This includes time being examined and treated. We calculate ED wait times as the time elapsed between arrival and the admission decision, where the arrival time is recorded as patients enter the ED.<sup>17</sup> The data also include a hospital identifier, whether the patient is admitted or discharged, details of

<sup>16</sup>Admitting a patient is not costless: it takes time to prepare patients for admission and provide necessary information for inpatient staff. However, two institutional factors reduce these costs. First, minor patients can be admitted to wards attached to the ED (known as acute medical units) for further observation and minor tests before discharge. These units allow hospitals to meet the target while reducing the preparation required to admit patients. Second, more complex tests (e.g., MRI scans) can be carried out after admission. As a result, some of the effort required to treat patients may be pushed from the ED to inpatient departments.

<sup>17</sup>For nonambulance patients, this time is recorded when they first speak with the receptionist. Hospitals may attempt to manipulate wait times to meet the target. We evaluated one possibility in this regard, namely, that

basic diagnoses, the number and types of ED examinations and treatments, whether the patient arrived by ambulance, and some basic patient characteristics such as age, sex, and local area of residence.

Patients are identified by a pseudo-anonymized identifier that allows patients to be followed over time and across hospitals, and enable linkage between ED and inpatient records. Inpatient records contain detailed information on treatment undergone in the hospital, including dates of admission and discharge, and information on up to twenty diagnoses and procedures undertaken. Treatment is recorded at the episode level, defined as a period of treatment under the care of a single senior doctor.<sup>18</sup> We combine information across all episodes within the same admission to create visit-level variables for total length of stay (in days) and number of inpatient procedures. Each episode also contains a health care resource group (HRG) code, similar to diagnosis-related groups (DRGs). Hospitals are compensated by the government through a system of national tariffs for each HRG.<sup>19</sup> We calculate “costs” for each episode by matching tariffs to the appropriate HRG, which gives us a measure of the cost to the government, and revenue received by the hospital, associated with each visit. We then sum all treatment costs over a thirty-day period to estimate the cost associated with each ED visit and any follow-up treatment.

Mortality outcomes are recorded in records made available by the U.K. Office for National Statistics (ONS). These records are linked to HES through anonymized identifiers based on patient National Insurance (social security) numbers. The data include date of death for all U.K. citizens and any other individuals who died in the United Kingdom between April 2010 and March 2014. We create indicators of whether a patient dies within 30, 90, and 365 days of an ED visit.

*Sample construction.* Our analysis focuses on a sample of emergency patients treated in “major” EDs.<sup>20</sup> We keep all patients with full information relating to the timing of treatment and their exit route from the ED, in addition to their age, gender, and whether they arrived by ambulance.

hospitals simply miscode the timing of the admission decision, such that the total wait time is four hours or less. Following Locker and Mason (2006), we analyzed the distribution of “final digits” in wait times (e.g., the digits 0 to 9 at the end of each wait time value), which in the absence of manipulation should be uniformly distributed. Relative to this benchmark, we found that fewer than 1% of records were likely to be miscoded and that this would have a negligible impact on our analysis.

<sup>18</sup>Senior doctors—equivalent to U.S. attending physicians—are known as “consultants.”

<sup>19</sup>National tariffs are calculated for each HRG on the basis of annual cost reports submitted by hospitals to the U.K. Department of Health. These tariffs reflect the average cost of providing the procedure. Payments are then adjusted for unavoidable regional differences in providing care and unusually long hospital stays.

<sup>20</sup>Major EDs are defined as consultant-led providers of 24-hour services, based in specifically built facilities to treat emergency patients that contain full resuscitation facilities. We exclude patients treated at specialist clinics that treat only particular diagnoses (e.g., dental) and minor injury (“walk-in”) centers, where wait times are not typically long enough to be affected by the target. This excludes 18% of emergency visits.

Dropping patients with some missing information reduces the number of visits in the sample by 14.5%.<sup>21</sup> This yields an analysis sample of 14.7 million patients, who made 24.7 million visits to 184 EDs between April 2011 and March 2013.

*Summary statistics.* Table 1 reports summary statistics. The first two columns present the mean and standard deviation for a range of patient characteristics, treatments, and outcomes for all ED patients in the sample. Mean ED patient age was 39 years, and 51% of patients were male; 29% of patients arrived by ambulance; 5.8 million visits, or 24% of all ED episodes, resulted in an inpatient admission at the same hospital; and 58% of visits did not require further hospital treatment and led to a patient being discharged. The remaining visits resulted in a transfer to an outpatient clinic or another hospital for further treatment. Mean thirty-day treatment costs were \$1,676 (£1,240), of which 89% was accounted for by subsequent inpatient treatment. In the short term, mortality among ED patients is relatively rare. Two percent of patients died within thirty days of visiting the ED, which increases to 3% over a ninety-day period and 5% during the following year.

Table 1 also shows summary statistics separately for visits that led to an inpatient admission. As expected, these cases are typically more severe, with an older average age (55 years) and twice the likelihood of arriving in an ambulance (60%). Mortality rates (5% over thirty days, 16% over a year) are substantially higher than in the main sample. ED treatment is also more intense, with a higher mean number of treatments and examinations. Their treatment is also more expensive, with an average total cost over a thirty-day period of \$4,762 (£3,530).

Inpatients also experienced longer mean wait times in the ED than those who are not admitted. Mean wait times were 223 minutes for patients who were eventually admitted as inpatients, compared to a mean of 155 minutes for all ED patients. This demonstrates that the level of patient complexity, and the intensity of treatment for these patients, is likely to vary by wait time. This variation is important to account for when analyzing the impact of the target.

Figure 1 shows the distribution of ED wait times. A noticeable discontinuity is shown in the proportion of patients who exit the ED in the period immediately prior to four hours. This spike is unlikely to naturally occur and is instead induced by the target. We cannot illustrate the absence of this spike before the wait times target, since we do not have systematic data available from that period. But it is worth noting, as we do in appendix figure A1, that such a spike is not present in data on ED wait times from a major U.S. hospital.<sup>22</sup>

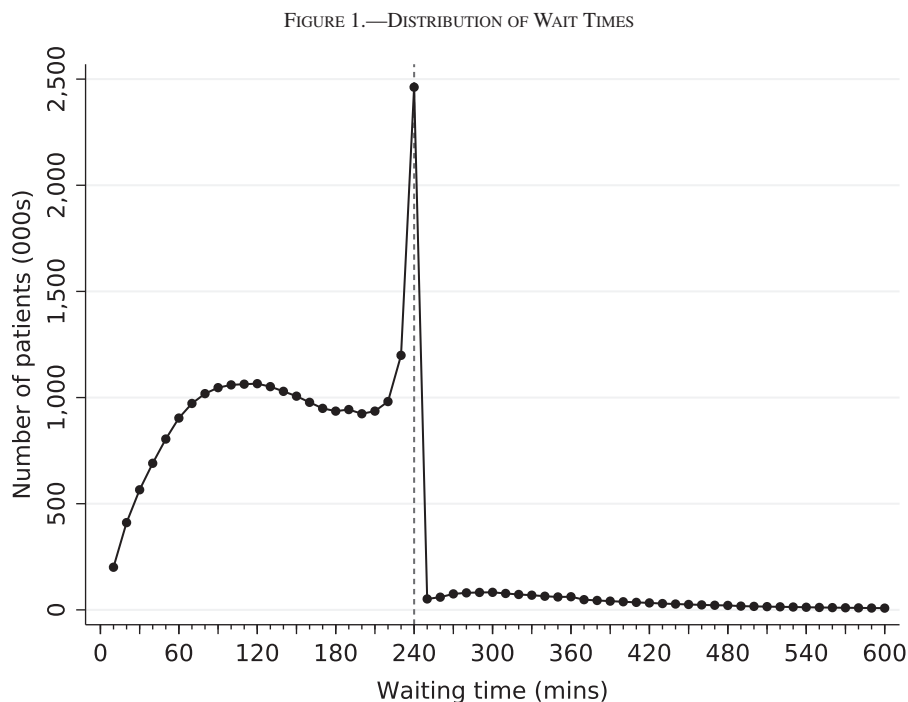
<sup>21</sup>Results are unaffected by the inclusion of patients with full information relating to treatment times and decisions but who are missing demographic information.

<sup>22</sup>Of course, different ED objectives and technologies across countries means that the U.S. data do not provide a natural comparison group, but the lack of any spike confirms our conclusion that the large spike here is particular to the wait time policy.

TABLE 1.—SUMMARY STATISTICS

	All patients		Admitted inpatients	
	Mean	Std. dev.	Mean	Std. dev.
Patient characteristics				
Age	38.99	26.22	54.64	27.84
Male	0.51	0.50	0.48	0.50
Ambulance arrival	0.29	0.45	0.60	0.49
Past-CCI	0.20	0.78	0.47	1.20
Treatment decisions				
Inpatient admission	0.24	0.42	1.00	0.00
ED discharge	0.58	0.49	0.00	0.00
ED referral	0.19	0.39	0.00	0.00
Wait time (minutes)	154.56	100.20	222.50	120.46
ED treatment count	1.81	1.38	2.22	1.68
ED investigation count	1.54	2.03	3.18	2.50
Inpatient length of stay (days)	1.28	5.63	5.41	10.58
Inpatient procedure count	0.16	0.64	0.69	1.18
Costs				
30-day ED cost	172.35	117.21	203.98	114.98
30-day inpatient cost	1,503.58	5,321.99	4,558.00	8,524.53
30-day total cost	1,675.93	5,358.37	4,761.98	8,559.73
Mortality outcomes				
30-day mortality	0.02	0.13	0.05	0.23
60-day mortality	0.03	0.16	0.09	0.29
365-day mortality	0.05	0.22	0.16	0.37

(1) Costs reported in 2018 USD and refer to payments from the government to hospitals based on the prospective payment system; (2) all inpatient variables (e.g., length of stay, costs) take on the value zero for patients that are not admitted.



(1) Wait time intervals are ten-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) wait times over 600 minutes not shown; (3) 240 minutes are the four-hour threshold specified in the policy.

One possibility is that this spike in wait times simply reflects recoding and is not a real change in patient wait times. Two features suggest this is not the case. First, a sizeable share of hospitals pay large penalties and are publicly criticized as a result. Indeed, a substantial number of hospitals only just miss the target, with 23% of hospitals missing the

target by less than two percentage points in 2011/2012. If recoding explained the spike, then those hospitals should do more recoding to avoid the penalty altogether. Second, we show below comparable spikes in a number of real outcomes, such as hospital admissions, costs, and mortality, that are inconsistent with this simply being a coding response.

#### IV. Empirical Methodology

A key challenge when analyzing the four-hour target is that without prepolicy data or a control sample, quasi-experimental methods cannot be used to construct counterfactual outcomes. To address this issue we use and extend bunching estimators that were developed in the tax literature (Saez, 2010; Chetty et al., 2013). We argue these methods can be used to estimate counterfactual outcomes that would occur if the target was removed but other aspects of hospital production were held constant, allowing us to quantify the short-run impact of the policy.<sup>23</sup>

We now set out our empirical methodology. We begin by setting out a bunching estimator for waiting times before giving an overview of our analysis of treatment decisions and health outcomes. More details on this methodology are set out in appendix C.

##### A. Wait Times

We first apply a bunching estimator to the distribution of wait time outcomes. Let  $w$  be the wait time in minutes, where  $w^* = 240$  (the target threshold). Denote the density function of  $w$  in the targeted regime as  $f_t(w)$ , where  $t = \{0, 1\}$  signifies whether the function relates to the targeted or nontargeted regime. We observe data on  $f_1(w)$  and use a bunching estimator to obtain  $f_0(w)$ .

To implement the bunching estimator we aggregate the data to ten-minute wait time bins and then interpolate parts of the distribution using a polynomial regression. Following Kleven (2016) we define  $\hat{f}_0(w) \equiv \sum_{i=0}^p \hat{\beta}_i w^i$  and obtain the estimates  $\hat{\beta}_i$  from the following regression:

$$c_j = \sum_{i=0}^p \beta_i (w_j)^i + \sum_{k=w^-}^{w^+} \gamma_k \mathbf{1}[w_j = k] + u_j, \quad (1)$$

where  $c_j$  is the number of individuals in wait time bin  $j$ ,  $w_j$  is the maximum wait time in bin  $j$  (e.g.,  $w_j = 10$  for the 1–10 minute wait time bin,  $w_j = 20$  for the 11–20 minute wait time bin, etc),  $p$  is the order of the polynomial, and  $[w^-, w^+]$  is an “exclusion window” that contains  $w^*$  and is the period during which we assume that the target may have had local effects on the wait time. This regression fits a polynomial to the wait time distribution in periods outside of the exclusion window, where the window is captured by the indicator variables that then do not feature in  $\hat{f}_0(w)$ .

<sup>23</sup>The counterfactual that the bunching estimator delivers in our context holds constant other aspects of hospital production, such as patient prioritization, capital and labor inputs, and government funding. As a benchmark, the counterfactual focuses attention on the role of incentives in determining outcomes rather than the specifics of the production function in our setting. We see it as a logical benchmark for understanding how wait time incentives affect outcomes, but our counterfactual will differ from the prepolicy or long-run outcomes. For example, it rules out wholesale changes in production technology or the prioritization of patients in earlier parts of the wait time distribution. The full policy impact relative to the prepolicy situation would potentially include the impact of these changes as well as the discontinuity in incentives introduced by the target.

Equation (1) makes the following assumption in relation to the exclusion window.

**Assumption 1 (Local wait time effects).** *Wait times of patients outside of an “exclusion window,” defined locally around the threshold  $w^*$ , are unaffected by the target:*

$$f_0(w) = f_1(w), \quad \forall w \notin [w^-, w^+]. \quad (2)$$

This assumption will hold if hospitals do not respond to the target by substituting resources between patients that are inside and outside of the exclusion window.<sup>24</sup> We discuss this assumption at length in the next section.

To establish the bounds of the exclusion window, we follow Kleven and Waseem (2013) and set  $w^-$  visually by examining when the distribution changes sharply and determine  $w^+$  using an iterative procedure that equates the excess mass in the period  $[w^-, w^*]$  with the missing mass in the period  $(w^*, w^+]$ .<sup>25</sup> An advantage of this iterative approach is that we make no assumption about  $w^+$  and let the data determine where the effects on the wait time distribution end. In the baseline analysis we use a polynomial of order 10 and set  $w^- = 180$ . After applying the iterative procedure this produces an upper cutoff of  $w^+ = 400$ . We show below that although the exact magnitude of the results is somewhat sensitive to the choice of parameters, our conclusions are qualitatively robust to variations in the choice of polynomial and  $w^-$  (see appendix tables A1 and A2).

The observed data and our estimated counterfactual distribution are shown in figure 2, which indicates that the target moves a number of patients from the postthreshold period to the prethreshold period (“postthreshold movers”). We later use these distributions to estimate the impact of the target on wait times.

##### B. Treatment Decisions and Mortality Outcomes

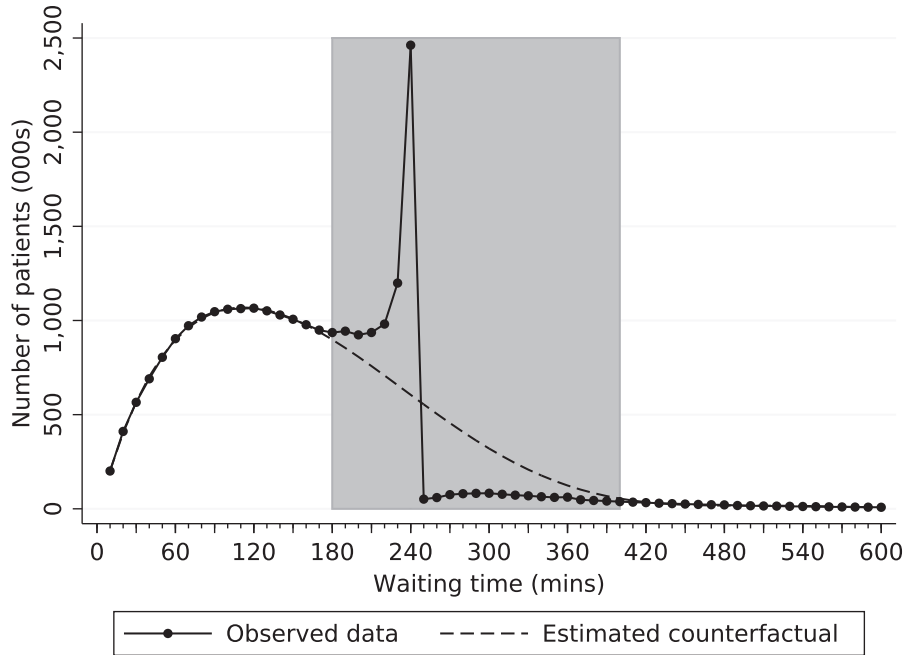
We now extend the analysis to consider outcomes other than the wait time, such as treatment decisions (e.g., inpatient admission) and mortality outcomes. Plotting these outcomes conditional on the wait time shows that they also exhibit “bunching” at the four-hour discontinuity point. Figure 3 gives an example for the likelihood of inpatient admission. The plot shows that admission probability is generally increasing with wait times, and a clear spike is visible in admission probability at 240 minutes. Our analysis decomposes this spike into two channels.

<sup>24</sup>A comparable assumption is required when using bunching techniques to study taxable income responses. In that setting the local effects assumption is often innocuous because the income distribution is the result of optimization decisions of many unrelated individuals, with those situated far from the tax scheme discontinuity having no incentive to adjust behavior. In our setting, the distribution of patient wait times is not determined by patients’ decisions but by the decisions of doctors and nurses, and this raises the concern that there may be an incentive to substitute wait times between patients across different parts of the wait time distribution.

<sup>25</sup>This implicitly assumes the target does not affect short-run patient demand for ED care.

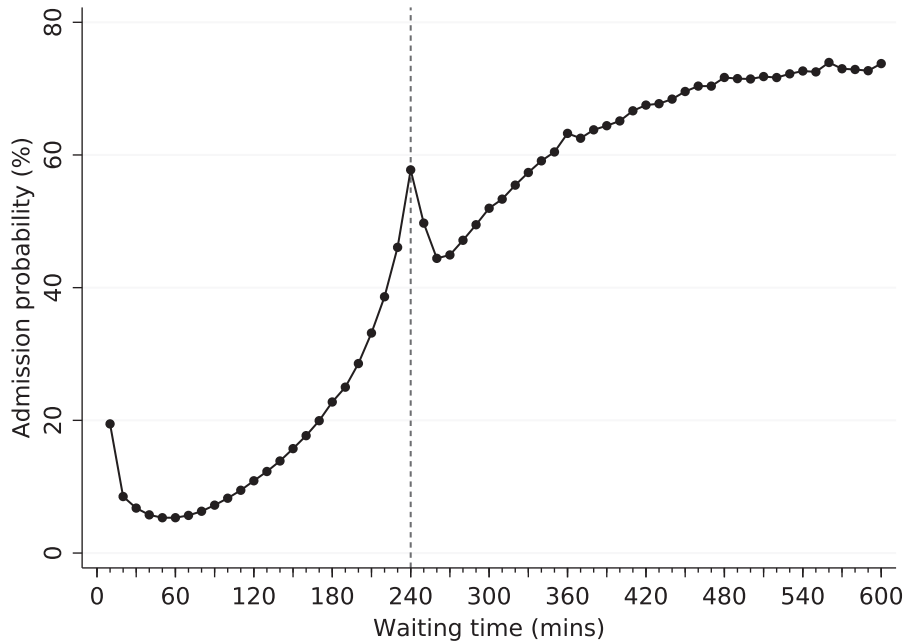


FIGURE 2.—ESTIMATED COUNTERFACTUAL WAIT TIME DISTRIBUTION



(1) Wait time intervals are ten-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) wait times over 600 minutes not shown; (3) 240 minutes are the four-hour threshold specified in the policy; (4) the estimated counterfactual is obtained from a polynomial regression that omits the exclusion window shown in gray.

FIGURE 3.—INPATIENT ADMISSION PROBABILITY CONDITIONAL ON WAIT TIME



(1) Wait time intervals are ten-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) wait times over 600 minutes not shown; (3) 240 minutes are the four-hour threshold specified in the policy.

The first channel is the “composition effect.” As figure 1 suggests, the target causes a substantial number of patients to be moved from later to earlier in the distribution of wait times (a group we refer to as “postthreshold movers”). Since admission probabilities are increasing with wait time, this movement of patients would increase the observed prethreshold admission probability even if the target led to no addi-

tional admissions. This effect arises purely because the target changes the composition of patients observed at each wait time.

Potential also is present for a “treatment change effect” if the target has a direct effect on treatment decisions and health outcomes. The treatment change effect implies identical patients receive different treatment depending on whether or not

the target is in place. In the case of admissions, for example, it would imply that part of the spike in observed outcomes is because the target causes additional admissions, in addition to the composition effect shifting some admissions from after to before the target.

To decompose the two effects we construct a “composition-adjusted counterfactual” (CAC). This is the outcome that would occur in the presence of composition effects but no treatment change effects. Since the observed data contain both effects, the difference between the observed data and the CAC identifies the treatment change effect. Estimates of these effects and tests of whether these are significantly different from zero are the central results of this paper.

We construct estimates of the CAC as a weighted average of counterfactual outcomes for patients who are observed in the prethreshold part of the wait time distribution (i.e., between  $w^-$  and  $w^*$ ). This includes two separate groups: patients shifted by the target from the postthreshold to the prethreshold period (“postthreshold movers”) and patients who would have been treated prior to the threshold even without the target (“prethreshold nonmovers”). From the wait time analysis, we know how many patients are moved from the postthreshold part of the wait time distribution to the prethreshold part of the distribution as a result of the target. The weights are therefore defined by the observed and counterfactual wait time distributions.

We then construct the required counterfactual outcomes by applying bunching techniques to the *expected outcomes conditional on the wait time*.<sup>26</sup> This relies on two key assumptions.

**Assumption 2 (Local outcome effects).** *Outcomes outside of an “exclusion window,” defined locally around the threshold  $w^*$ , are unaffected by the target:*

$$E[y_{1i} | w_{1i} = w] = E[y_{0i} | w_{1i} = w], \quad \forall w_{1i} \notin [w^-, w^* + \varepsilon], \quad (3)$$

where  $E[y_{0i} | w_{1i} = w]$  is the expected outcome in the absence of the target conditional on observed wait time, and  $E[y_{1i} | w_{1i} = w]$  is the expected outcome under the target conditional on observed wait time (i.e., the observed outcome).

Assumption 2 rules out treatment change effects outside of the prethreshold period. It is the parallel of assumption 1 for the conditional expectation function. In this case the exclusion window ends at  $w^* + \varepsilon$ , where  $\varepsilon$  is a small “overhang period” that extends past the four-hour threshold. This overhang period allows for the empirical fact that the bunching in outcomes extends slightly past the threshold (see figure A2 in appendix A). We interpret the overhang as being a case of treatment change effects for patients that are narrowly dis-

charged or admitted after the threshold. For example, it may be that doctors admit additional patients in an attempt to meet the target, but not all of the excess admits occur before the threshold as some patients may be delayed for unexpected reasons. We determine the size of the overhang period visually, setting  $\varepsilon = 20$  in the baseline analysis, and note that our findings are robust to more conservative (larger) overhang periods.<sup>27</sup>

**Assumption 3 (No selection).** *Nontargeted regime outcomes conditional on the nontargeted wait time are comparable for postthreshold movers and postthreshold nonmovers:*

$$\begin{aligned} & E[y_{0i} | w^- < w_{1i} < w^*, w_{0i} = w] \\ &= E[y_{0i} | w^* < w_{1i} < w^+, w_{0i} = w] \\ &= E[y_{0i} | w_{0i} = w], \quad \forall w_{0i} \in \underline{w}_0^+, \end{aligned} \quad (4)$$

where  $E[y_{0i} | w^- < w_{1i} < w^*, w_{0i} = w]$  is the expected outcome for postthreshold movers under the nontargeted regime,  $E[y_{0i} | w^* < w_{1i} < w^+, w_{0i} = w]$  is the expected outcome for postthreshold nonmovers under the nontargeted regime, and  $\underline{w}_0^+ = w^* < w_{0i} < w^+$ .

Assumption 3 rules out composition effects in the postthreshold period. It states that after conditioning on the nontargeted wait time, no selection occurs when the postthreshold movers are assigned. The assumption is consistent with doctors randomly selecting patients to get shorter wait times in response to the target. In that sense it is equivalent to an unconfoundedness assumption in traditional instrumental variables (IV) terminology.<sup>28</sup> Although this is a strong assumption we believe it is plausible and, most importantly, are also able to evaluate the assumption empirically using placebo tests. We discuss this assumption and the results of these tests in detail shortly.

Together assumptions 2 and 3 imply no composition or treatment change effects outside of the exclusion window  $[w^-, w^* + \varepsilon]$ . We can therefore apply the bunching estimator in the same way as for the wait times but to the conditional expectation function  $E[y_1 | w_1]$ . The estimated counterfactual delivered by the bunching estimator is then  $E[y_0 | w_0]$ . This is shown in figure A2 in appendix A. This directly gives us counterfactual outcomes for the prethreshold nonmovers and, given assumption 3, provides us with the counterfactual outcomes for the postthreshold movers. Taking the weighted average of these outcomes therefore yields an estimate of the composition-adjusted counterfactual.

Under these assumptions, we can test whether treatment change effects exist by taking the differences in the observed

<sup>27</sup>Our estimates of the treatment change effect, which relate to the prethreshold period, do not capture distortions in the overhang period. These omitted effects are small: the number of patients in the overhang period is 1.3% of the number of patients in the prethreshold period.

<sup>28</sup>Similarly, in IV terminology, the postthreshold movers would be compliers, the postthreshold nonmovers would be never-takers, and the prethreshold nonmovers would be always-takers. We implicitly make the assumption of no defiers.

<sup>26</sup>This is in contrast to a typical bunching application that would work with the distribution of a variable that is subject to a discontinuity in incentives. Here we work with outcomes conditional on a variable that is subject to a discontinuity in incentives. Our approach is similar in spirit to Diamond and Persson (2016) and Gerard, Rothe, and Rokkanen (2018).

outcomes and the estimated composition-adjusted counterfactuals ( $\Delta D$ ). Figure C1 in appendix C provides a visual example of how we construct the CAC and the test of treatment change effects for the probability of inpatient admission. Tests for treatment change effects are then simply hypotheses tests that these differences are equal to zero. We compute statistical significance for the test using nonparametric bootstrapped standard errors clustered at the hospital organization level.<sup>29</sup>

The next section explores the validity of these assumptions in detail. To evaluate the validity of the no-selection assumption, we devise a test based on observable patient characteristics that cannot be altered by the hospital. This includes age, sex, whether the patient arrived in an ambulance, three health measures based on hospital use in the prior year (Charlson comorbidity index [CCI], number of emergency admissions, days spent in the hospital), and predicted mortality and admission. These variables, conditional on the wait time, also exhibit bunching at the four-hour point, but in these cases the spike can be explained only by a composition effect because no treatment change effect is seen by definition. If the no-selection assumption is valid, then for these variables the observed data and the CAC should be equal (i.e., the estimated treatment change effect is equal to zero). We therefore estimate the treatment change effect for each of these variables. This acts as a placebo test, where an estimated treatment change effect significantly different from zero would suggest that the no-selection assumption has been violated.<sup>30</sup> We pass these placebo tests for the majority of tested variables, and where they fail (sex, past-CCI) the magnitudes are very small, whereas any bias from selection on (unobservable) severity (if it mirrors observed severity) would likely make our mortality estimates conservative. We discuss these results, and further tests of this assumption, in detail in section VB.

## V. Validity of Key Assumptions

Our methodology rests on our assumptions about local effects (assumptions 1 and 2) and selection (assumption 3). We discuss these assumptions and supporting evidence below.

### A. Local Effects Assumption

The local effects assumption is that wait times and treatment decisions before  $w^-$  are unaffected by the target, and we set  $w^-$  at 180 minutes in the baseline analysis. As noted earlier this will not hold if hospitals substitute time or resources between patients that exit before  $w^-$  (“early exit patients”) and after  $w^-$  (“late exit patients”). This assumption rules out certain dynamic responses that may impact the wait

time distribution. We suggest a set of factors that mitigate the importance of this issue and then carry out three empirical tests that support the credibility of this assumption in our setting. In this section we provide an overview of these tests and include detailed results in appendix D.

As noted in section IIC, a number of institutional factors mitigate concerns about a violation of local effects. Concerns about the substitution of resources between patients is limited by the fact that patients are treated across many hospitals, days, and time periods and the organization of staff into “minor” and “major” units. Broader concerns about the dynamic behavior of physicians are limited by the fact that physicians can potentially shorten ED treatment by admitting patients and so have little incentive to speed up the treatment of those in the early part of the wait time distribution. In addition, as hospitals are already attempting to maximize an objective function that prioritizes patient outcomes, making changes to the ordering of very severe or very easy cases is unlikely. These unambiguously high- and low-severity patients are likely to account for a significant proportion of exits before the exclusion window and suggest that if changes are made to treatment decisions, then these are more likely to occur near to  $w^-$  rather than at the very start of the distribution.

We conduct three empirical tests to further evaluate this assumption. First, we expect any dynamic responses to be concentrated near  $w^-$ , and so a natural robustness test is to check whether our results are sensitive to the choice of  $w^-$ . We therefore vary  $w^-$  and assess how sensitive our results are. We show the results of this exercise in appendix A (table A1). The results suggest that our estimates are qualitatively robust to changes in  $w^-$ , with the same sign and significance across all specifications for most variables. Some point estimates do vary in size: for example, the estimated impacts on admission double in magnitude when moving from the earliest to latest starting point. However, importantly, reducing  $w^-$  from our baseline parameter does not result in statistically significant changes to the estimates. This is inconsistent with any meaningful changes to physician behavior in the earlier part of the wait time distribution.<sup>31</sup> In addition, estimated mortality effects are not statistically significantly different from one another. This suggests some sensitivity in the magnitude of our estimates with respect to the starting point of the exclusion window but does not change our overall conclusions.

The remaining tests examine whether evidence of substitution of resources across patient types is found to meet the target. Hospitals may approach the prioritization of patients entirely differently when facing the target compared to an unconstrained scenario. This would undermine the previous test by implying that no part of the distribution is unaffected by the target. If this concern is valid, however, it implies that hospitals should change the priority order assigned to patients

<sup>29</sup>We cluster all results at the trust (organization) level because some trusts do not code patients at the site level. NHS trusts group hospitals in geographical proximity that share common management. All results are robust to clustering at the site level.

<sup>30</sup>Figure C2 in appendix C provides a visual example of the demographic test for age.

<sup>31</sup>In contrast, increasing  $w^-$  is associated with larger changes in the estimates. This is unsurprising given that visual inspection of the waiting times distribution shows clear distortions as wait times approach 240 minutes, and these are already apparent at 200 minutes. We would therefore expect to capture some of these dynamic responses in our estimates as  $w^-$  increases.

TABLE 2.—DEMOGRAPHIC TESTS OF THE NO-SELECTION ASSUMPTION

	Treatment change effect ( $\Delta_D$ )		CAC mean
	Level (1)	% (2)	Level (3)
Panel A: Individual characteristics			
Age	0.417 (0.284)	0.009 (0.006)	46.47
Male	-0.005*** (0.001)	-0.011*** (0.003)	0.487
Ambulance	-0.002 (0.004)	-0.005 (0.010)	0.440
Past-CCI	0.013*** (0.005)	0.043*** (0.016)	0.300
Hospital days (2010)	0.066 (0.058)	0.018 (0.016)	3.67
Emergency admissions (2010)	0.020 (0.014)	0.052 (0.037)	0.385
Panel B: Predicted characteristics			
Predicted admission	0.003 (0.002)	0.008 (0.007)	0.308
Predicted mortality	0.000 (0.000)	0.015 (0.015)	0.019

(1) CAC mean is measured over the prethreshold period,  $E[y_0 | \underline{w}_1^-]$ ; (2) predicted variables defined using a regression of the variable on past-CCI score, number of emergency admissions and days spent in hospital in 2010, and a fully interacted set of age, gender, and ambulance-arrival fixed effects; (3) bootstrapped standard errors clustered at the hospital trust level (199 repetitions).

based on how tightly the target is anticipated to bind: with a nonbinding target, hospitals are unconstrained by the target. Our second test is therefore to exploit variation in the expected volumes of ED arrivals—with the target binding more tightly as volumes increase—to see if it impacts patient prioritization, especially at earlier wait times.<sup>32</sup>

Figure D1 in appendix D plots the proportion of patients who exit within 180 minutes for each percentile of predicted mortality for patients that arrive during more or less busy periods (based on their predicted patient volumes).<sup>33</sup> It shows that a smaller proportion of high-severity patients leave the ED within 180 minutes and that busier periods have longer wait times for patients of all severity. Most importantly, the relative probabilities of exits within 180 minutes for high- and low-severity patients are very similar in both types of period, that is, a parallel downwards shift is seen in the relationship. This is precisely what should happen if hospitals do not change the patient prioritization in response to the target binding more or less tightly, as implied by the local effects assumption. These results suggest that as the target binds more or less tightly hospitals do not change the prioritization of patients, consistent with our assumption.

As a final test, we consider whether hospitals temporarily substitute resources between patients if they experience a demand shock (e.g., the ED is momentarily overrun with patients) and this causes short-term deviations from planned priorities to meet the target. One specific example that we consider is whether a hospital that has a build-up of patients close to breaching the target temporarily substitutes resources away from newly arriving patients to clear the backlog. The local effects assumption would be violated if we saw evidence

that actual wait times of patients with expected waiting times of under 180 minutes (i.e., outside of the exclusion window) were particularly affected by the presence of patients about to breach the target.

Appendix D sets out a formal test for such behavior. Intuitively, we compare wait times of newly arrived patients on the basis of how many existing patients have waited almost four hours. If temporary substitution effects occur between these individuals, we would anticipate large effects of the presence of existing patients near the four-hour threshold on the wait times of new patients. However, as figure D2 in appendix D shows, we do not find any evidence that such behavior takes place. Patients predicted to wait for less than 180 minutes do, on average, wait longer when more patients are already present upon arrival in the ED, but their wait times are most affected by the presence of other recently arrived patients. We find no evidence that the presence of existing patients close to the four-hour target causes disproportionate impacts on their wait times. In contrast, for patients who are expected to wait for more than 180 minutes, we see a disproportionately large impact of the presence of patients who are close to breaching the target at the time that they arrived. This suggests temporary substitution responses for patients predicted to be within the exclusion window but not for those predicted to be in the earlier part of the distribution, as is consistent with our local effects assumption.

Taken together, we interpret these tests as providing strong empirical support for the plausibility of assumptions 1 and 2 in our setting and proceed on that basis.

### B. No-Selection Assumption

We set out our methodology for a test of assumption 3, based on observable demographic and prior health variables, in section IVB. Table 2 presents the results of the relevant

<sup>32</sup>Repeating the same test using shocks to ED arrivals delivers similar results.

<sup>33</sup>See appendix D for details on how we classify busy and nonbusy periods.

tests. Column 1 presents estimates of the treatment change effect, and column 2 presents estimates of the treatment change effect as a proportion of the counterfactual mean. Panel A presents results using individual variables, where we test using age, a male indicator, an indicator for whether the patient arrived via ambulance, and the number of emergency admissions, total number of days spent in hospital, and the Charlson comorbidity index (“past-CCI”) based on the twelve months of hospital admissions before the beginning of our ED data. Each of these variables should be unaffected by decisions made in the ED and thus allow us to test our selection assumption.<sup>34</sup>

For age, ambulance arrival, past number of emergency admissions, and past days spent in hospital, we cannot reject the no-selection hypothesis. In contrast, we reject the no-selection hypothesis for gender and past-CCI. The gender result suggests that postthreshold movers are more likely to be female than the postthreshold nonmovers. However, the extent of this selection effect is small: the difference between the observed and composition-adjusted counterfactual proportion of females in the pretarget period is 0.5 percentage points (1.1% of the baseline).<sup>35</sup> With regard to the past-CCI results, the positive estimate suggests postthreshold movers are on average less healthy than postthreshold nonmovers, with a past-CCI score that is 4% higher. Although this estimate is small in magnitude, this is consistent with physicians responding to the target by prioritizing patients with a worse health record.

Panel B in table 2 presents results for variables that are linear combinations of the individual demographic variables. We use predicted admission and predicted mortality, where the predictions are obtained from linear regressions of the outcome on a flexible specification of the demographic variables (past-CCI score, previous hospital days and emergency admissions, and a fully interacted set of age, gender, and ambulance-arrival fixed effects). The  $R^2$  statistic from these predicted regressions is 0.22 and 0.06. An advantage of using these predicted variables is that they weight individual demographic variables according to their relative importance for clinical outcomes. Weighting factors on this basis is useful because selection on factors that do not impact these outcomes is unlikely to bias our estimates. Looking at the estimates, the demographic tests for these predicted variables cannot reject the hypothesis of no selection. So even though the gender and past-CCI tests reject the hypothesis, the contribution of these

variables to salient medical outcomes, and thus the likelihood of bias, is low.

As a direct test of whether gender and past-CCI introduce meaningful bias to our estimates, we computed estimates conditional on these observables and compared them to our baseline estimates that we present below. The two sets of results were very similar, suggesting that any selection does not introduce substantive bias to the estimates.<sup>36</sup> We also note that any bias from selection on (unobservable) severity, if it mirrors the past-CCI result, would attenuate our estimates toward zero and thus make our mortality estimates conservative.

As a final probe of the no-selection assumption, we simulated how selection of different degrees would manifest itself in the observed data. To do this we built a simulated data set using the counterfactual wait time, age, and past-CCI distributions and then artificially assigned postthreshold movers using different selection rules. We describe this process in appendix E. The simulation highlights three facts about selection in our setting. First, the observed data on age and past-CCI look very similar to the simulated data with random selection. Second, even very modest selection is predicted to have a clear impact by creating a spike in outcomes in the prethreshold period and a very pronounced “dip” in outcomes in the postthreshold period, neither of which is seen in our data. Third, an advantage of our test is that it has potential to detect selection on unobservables even though it relies purely on observables. This follows because a test based on age or past-CCI, for example, would reveal selection on another unobservable variable as long as it is sufficiently correlated with age or past-CCI respectively.

Together these results indicate that the no-selection assumption is plausible in this setting. On its face, this is perhaps a surprising finding. Although patients themselves do not make the selection decision, hospitals do make these choices, and selecting certain patients may be in their interest. But on a day-to-day basis, hospitals are treating patients at different times, and this limits the scope for selection. If the data are segregated into hospital-hour periods, for example, then the number of patients approaching the target at any given point in time is actually small, at around three or four. This compares to an average of 3.5 physicians that are on shift in a typical ED, suggesting physicians rarely have a choice between multiple “potential breach” patients.<sup>37</sup> We therefore view breaches of the target as more likely to occur because of idiosyncratic events and delays (e.g., staff shortages) rather than being a result of selection on patient characteristics. Although we cannot rule out that such events could be correlated with patient characteristics, our demographic tests suggest that this is not the case. In practice, we therefore treat those patients observed with wait times in excess

<sup>34</sup>Several variables we considered for the test have missing data around the threshold (e.g., ED diagnosis). As a result we restrict our attention to examining health variables that were recorded before the ED visit, and whose recording will not be affected by the target itself.

<sup>35</sup>Exploring the male indicator more carefully shows that, unlike the other variables we study, it is poorly correlated with ED wait times. This causes the polynomial regression, which we use to determine the counterfactual outcomes, to fit the data less robustly (i.e., it is sensitive to the choice of polynomial). The demographic test is therefore not reliable for this variable. The same is also true for other variables that are poorly correlated with ED waits.

<sup>36</sup>To compute the conditional estimates, we apply our methodology to subgroups of patients defined by gender and past-CCI and then aggregate these results up to be comparable to the baseline estimates using the sample weights associated with each subgroup.

<sup>37</sup>We sent a data request to all NHS trusts for staffing figures, receiving responses from 40%.

TABLE 3.—ESTIMATED TREATMENT CHANGE EFFECTS OF THE TARGET ON TREATMENT DECISIONS, COSTS, AND MORTALITY

	Treatment change effect ( $\Delta_D$ )		CAC mean
	Level (1)	% (2)	Level (3)
Panel A: ED treatment decisions			
Pr(admission)	0.046*** (0.008)	0.122*** (0.022)	0.379
Pr(discharge)	-0.033*** (0.007)	-0.070*** (0.014)	0.472
Pr(referral)	-0.013*** (0.003)	-0.089*** (0.020)	0.150
ED investigation count	0.108** (0.048)	0.046** (0.021)	2.369
ED treatment count	-0.033 (0.028)	-0.016 (0.014)	2.070
Panel B: Inpatient treatment decisions			
Length of stay (days)	0.035 (0.048)	0.015 (0.021)	2.302
Inpatient procedure count	0.000 (0.006)	0.001 (0.020)	0.290
Panel C: Hospital costs			
30-day ED cost	3.040*** (0.911)	0.016*** (0.005)	192.950
30-day inpatient cost	125.793*** (33.992)	0.052*** (0.015)	2,414.087
30-day total cost	128.833*** (34.389)	0.049*** (0.014)	2,607.037
Panel D: Mortality			
30-day mortality	-0.0041*** (0.0006)	-0.138*** (0.019)	0.029
90-day mortality	-0.0040*** (0.0010)	-0.079*** (0.019)	0.048
1-year mortality	-0.0029* (0.0016)	-0.031* (0.017)	0.090

(1) CAC mean is measured over the prethreshold period,  $E[y_0 | w_1^-]$ ; (2) all inpatient variables (e.g. length of stay, costs) take on the value zero for patients that are not admitted; (3) bootstrapped standard errors clustered at the hospital trust level (199 repetitions).

of 240 minutes (postthreshold nonmovers) as comparable to those patients that would have had wait times over 240 minutes in the absence of the target (postthreshold movers), and we can therefore use these postthreshold nonmovers as the counterfactual for the postthreshold movers.

## VI. Results

We first present the wait time results. We then present results related to treatment decisions and mortality outcomes. We explore the mechanisms behind the mortality outcomes in section VII.

### A. Wait Times

Figure 2 shows the observed and estimated counterfactual wait time distributions. The shaded panel is the exclusion window where we estimate the effects of the policy, covering the period between 180 and 400 minutes. The solid line is the observed distribution of patients that exit at each interval, and the dashed line is the estimated counterfactual distribution. The effect of the target on exit times is clear: a large proportion of patients from the postthreshold period (240 to 400 minutes) are moved to the pretarget period (180 to 240 minutes). These are the patients we refer to as postthreshold movers. By comparing the observed wait time distribution

with our counterfactual we can compute the impact of the target on average wait times.

The results indicate that the target successfully reduced wait times. We estimate that the target reduces mean wait times by seven minutes, or 4% of the estimated counterfactual mean. For patients affected by the target (i.e., in the exclusion window), we estimate that the target reduces wait times by 21 minutes, or 8% of their estimated counterfactual mean. Moreover, if we restrict our attention to those patients moved to the prethreshold period from the postthreshold period (the postthreshold movers), then the average wait time reduction is 59 minutes.<sup>38</sup>

### B. Treatments and Mortality Outcomes

Table 3 presents results of the treatment change test for a range of treatment decisions, costs, and mortality outcomes, both in absolute values (column 1) and as a proportion of the counterfactual mean (column 2). Each row shows results for a separate outcome.

<sup>38</sup>Estimating the distribution of wait time reductions would require further assumptions on the ordering of patients. We do not impose these assumptions but note maximum wait time reduction could be as large as 200 minutes (i.e., a patient moved from 390 to 190 minutes).

Panel A presents estimates for ED treatment decisions. We find that, controlling for compositional changes, the probability of admission increases by 4.6%. This is 12.2% of the baseline composition-adjusted counterfactual value, which is sizeable. The results for discharges and referrals out of the ED to specialist clinics offset these admission effects, with roughly three-quarters of the effect coming from decreased discharges and one-quarter from decreased referrals, although as a percentage of the baseline these responses are of comparable magnitude.

We also show target effects on the number of investigations performed in the ED, such as x-rays, blood tests, and CT scans. We find that investigations rose by 0.1 per patient, or 4.6% of the baseline. We do not, however, find any effect on the number of treatments performed in the ED. This suggests that doctors perform more tests to speed up the admission decision for individuals (i.e., they perform an extra test instead of monitoring the patient for a longer period of time) but this has little effect on the treatments that they provide in the ED.

Panel B examines inpatient treatment decisions. To avoid selection, we include all ED patients, even those who did not end up being admitted. As a result, the coefficient represents the incremental amount of treatment due to the four-hour target. We find no evidence of any statistically significant increases in length of stay or the number of procedures. This suggests the extra admissions do not receive substantial amounts of care in the hospital; that is, these admissions appear to be largely placeholders to avoid the four-hour target.

Nevertheless, the additional admits are costly. Panel C examines the impact of the four-hour target on thirty-day patient costs. A small rise is seen in ED costs of \$3 (£2), or 2% of ED costs, but we have a significant increase in inpatient costs of \$126 (£93), which is 5% of inpatient costs. That is, even though most patients appear to be housed in inpatient departments only as a way of avoiding the four-hour target, these admissions generate transfers from the government to hospitals. Total costs rise by roughly 5% relative to the baseline.

Panel D then extends our analysis to look at patient mortality outcomes over a variety of time frames. We find significant short-term declines in mortality. Mortality over thirty days declines by 0.41%, or 14% of baseline. The CAC for thirty-day mortality is shown in appendix figure A3; here, after adjusting for the composition effect, we find that the observed data are lower than the CAC, and this is what produces the negative estimate. This effect fades slightly over time and falls as a share of the baseline, so that at one year it is only 3.1% of baseline. This pattern suggests that the health benefits of the policy are seen in the short term.

This is a sizeable mortality decline given the modest increase in costs documented in table 2. We find that total costs over thirty days from admission to the ER rise by 5%, and mortality falls by 3.1% over a year. Calculating the cost per year of life saved by the policy requires assumptions on how long-lasting is the impact on mortality and on any subsequent costs past thirty days. Assuming no subsequent costs, but also assuming that the mortality impact lasts only one

year, this implies a cost per year of life saved of \$43,000 (£31,850).<sup>39</sup> This is low relative to standard valuations of a life-year in the United States, where typical benchmarks are around \$100,000 (£74,000) (Cutler, 2003), and at the upper end of valuations in the United States, where the national benchmarks are set at \$28,000 to \$42,000 (£20,000 to £30,000) (McCabe, Claxton, & Culyer, 2008).

In summary, our analysis of the four-hour target shows that it led to shorter wait times, more admissions, only marginal additional costs (due to little use of extra inpatient care), and significant reductions in mortality. That is, it appears that constraining hospitals did save lives.

## VII. Mechanisms

### A. Using Patient Heterogeneity to Identify Mechanisms

Our results so far show a number of effects of the wait time target on patient treatment: on wait times, admission probabilities, and treatment costs more generally. We also show a significant effect on patient mortality. Ideally we would like to uncover the mechanism through which the four-hour target impacts patient mortality. This is difficult because we essentially have one instrument (the target) and multiple changes in patient treatment.

To address this issue we turn to considering heterogeneous impacts across patient types, examining whether groups of patients exist where we find differential effects of the target. If those groups have effects that are focused along one channel (e.g., wait times) but not another (e.g., admits), then we can use this to separate the effect of the two channels on outcomes.

We consider two natural sources of heterogeneity. The first is differences across diagnosis. Dividing patients into 36 ED diagnosis groups, figure F1 in appendix F shows that patients with the most severe diagnoses are most likely to hit the target. This suggests that we would expect the largest wait time impacts of the target to show up for those with the most severe diagnoses. We therefore separately compute the wait time reduction effects and treatment change effects for admissions and thirty-day mortality for each diagnosis group. We then assess how heterogeneity across diagnosis groups translates to each of these outcomes.

Figure F2 shows the results of this graphically. Panel A shows that higher severity diagnoses have larger wait time effects. This is sensible because they are most likely to wait the longest without the policy. Panel B shows the effect of the target on hospital admissions is no higher for more severe diagnoses. Panel C shows the absolute value of mortality reduction for each diagnosis group and clearly shows that the mortality effect of the target is strongest for the most severe diagnosis. To ensure selection is not driving our result, the

<sup>39</sup>This reflects the cost to the government due to the increase in HRG transfers to hospitals. The actual cost in terms of resource use will be even lower if the marginal admissions use fewer resources than the average HRG cost.

graph repeats this exercise for predicted mortality and finds no systematic relationship between the effects of the target on predicted mortality and diagnosis severity.

We formalize the results of this exercise by regressing the treatment change effect on mortality in absolute value for each diagnosis on the estimated wait time reduction and the treatment change effect for admission probability. A positive coefficient in these regressions can be interpreted as that margin being associated with a larger policy effect on thirty-day mortality.

Columns 1–3 in table F1 in online appendix F show these results. Column 1 shows that diagnosis groups with larger wait time effects have larger mortality effects. The estimated coefficient suggests a minute of wait time reduction reduces mortality by 0.001 percentage points. Earlier we estimated that wait times fell by nineteen minutes on average. This suggests a mortality reduction of 2.2 percentage points. This is of a similar magnitude to our reduced form estimate in table 3 of 3–4 percentage points. Column 2, however, shows no impact of the increase in admissions on mortality. Column 3 shows the same correlations when we consider both variables together. These results suggest that it is wait time reductions and not increased admissions that are driving the mortality reductions.

Of course, this set of corresponding facts does not prove this causal mechanism because other factors may be causing the effects to differ by diagnosis. So to further test this conclusion we consider a second source of heterogeneity: the degree of inpatient crowding. In times where the inpatient department is more crowded, EDs may be less able to meet the target by admitting patients because the inpatient wards have less spare capacity for these patients to be sent. But it is unclear that inpatient crowding would much affect the marginal wait time impacts of the target. Inpatient crowding therefore provides an opposite test of the diagnosis heterogeneity: an opportunity to observe heterogeneity that drives admission probabilities but not wait times.

We therefore divide the data into fifty quantiles depending on how busy the hospital inpatient department is on the day of admission, with patients assigned to each group based on their day of admission. To address differences in case mix during busy and quiet periods, we also split patients into two severity groups and interact the inpatient crowding groups with severity.<sup>40</sup>

Figure F3 in appendix F shows that inpatient crowding has a weak, positive relationship with wait times, while the most crowded inpatient departments experience the smallest increases in admission probability. However, panel C shows no significant relationship between the degree of inpatient crowding and the estimated mortality effect. When we repeat this analysis with estimated reductions in predicted mortality (which should be unaffected by the target once we adjust for patient composition) we find that these results are not driven

<sup>40</sup>See appendix F for details of how we calculate crowding-severity groups.

by selection: a positive relationship is seen between predicted mortality reductions and inpatient crowding, but this is small in magnitude.<sup>41</sup> Together, this suggests that admissions had little role in driving the mortality effects.

Columns 4–6 of table F1 formalize the results of this analysis. Again, a highly significant relationship is found between the wait time reduction and mortality reduction, with a coefficient similar to column 1. In this case, in column 5, we do see a significant effect of the admissions effect on mortality, albeit with a wrong signed coefficient suggesting that a larger admissions effect leads to a smaller mortality effect. But when both are included in column 6 only the wait time effect persists.

Taken together, the evidence suggests that heterogeneity associated with wait time variation appears to be associated with mortality variation, whereas heterogeneity associated with admissions variation does not. This does not prove that the wait time reductions are driving our mortality reductions, but it is highly suggestive.

### B. Wait Times, Diagnoses, and Causes of Death

This evidence raises the question of how reductions in wait times could lead to lower mortality rates. The most likely mechanism is that reductions in wait times lead to lower time-to-treatment for patients with severe diagnoses. An extensive medical literature makes it clear that rapid treatment is associated with better mortality outcomes for patients across a range of conditions. For example, Seymour et al. (2017) find a strong positive association between time-to-treatment and survival for ED patients with sepsis and septic shock.<sup>42</sup> However, it may be difficult for ED physicians to identify these patients as they arrive: a body of medical evidence suggests that misdiagnosis in the ED is not rare, and disagreement is often seen between ED physician and subsequent specialist diagnosis.<sup>43</sup> This suggests why the target may have been successful in improving outcomes relative to an unconstrained scenario as it leads doctors to speed up treatment for all patients, which is costly but ensures that hard-to-diagnose and time-sensitive patients ultimately get the correct treatment sooner.

We explore the likelihood that this mechanism is driving our results in two ways. First, we examine which parts of

<sup>41</sup>This means that our results may actually understate the mortality reductions in the most crowded periods. Given that these periods are also those with the smallest increases in admissions, this would strengthen the conclusion that mortality reductions are associated with reductions in wait times and not additional admissions.

<sup>42</sup>Many examples can be given from other diagnoses. For example, Saver et al. (2013) find significant improvements in patient outcomes for stroke patients when cutting time to treatment.

<sup>43</sup>Shojania et al. (2003) conducted a systematic review into studies of autopsy-detected diagnostic errors over a forty-year period in the United States and found a median error rate of 23.5%. Delays and misdiagnoses are particularly common for neurological and cerebrovascular patients, and many of the existing studies are in this area. For example, Newman-Toker et al. (2014) estimate that between 15,000 and 165,000 cerebrovascular events are misdiagnosed annually in U.S. EDs.



the treatment pathway are compressed. This provides some evidence on whether patients start to receive treatment earlier. Second, we examine variation in mortality reductions across diagnoses and primary causes of death to see whether the greatest mortality reductions are for patients where outcomes are known to be time-sensitive.

To examine how wait times are reduced, we break down the overall impacts on waits into the three separate components that the data allow: time to initial assessment, time between assessment and the beginning of treatment, and duration of ED treatment. The initial assessment is usually conducted by a triage nurse and includes a relatively basic examination. Treatment begins when the patient is first examined by a doctor and ends when the ED makes the decision to admit or discharge the patient. Admitted patients will then receive further treatment from a specialist within the hospital. As noted in section IIA and shown in appendix tables B1 and B2, most ED treatment in England is aimed at stabilizing and diagnosing patients, with more extensive treatments provided by specialists in inpatient wards. Reducing treatment time in the ED therefore means that patients start to receive this specialist treatment sooner.

We repeat the analysis in section IVC using time to initial assessment, time between assessment and the start of treatment, and duration of ED treatment as separate outcomes. The results show that the reductions are achieved mostly by reducing the initial wait for treatment (48% of the overall reduction) and by shortening the duration of ED treatment (45%). These results suggest that the target reduces the wait for both ED and specialist treatment to begin. Patients start to receive treatment from ED physicians sooner and spend less time receiving treatment in the ED. Importantly, shorter periods spent in the ED also mean that admitted patients start to receive specialist inpatient treatment sooner. This specialist treatment often begins with further diagnostic testing (such as CT or MRI scanning, as documented in table B3), and so reducing ED time means patients are likely to receive a detailed diagnosis sooner.

Next we turn to examining which patients benefit most (in terms of mortality reductions) from the target. If quicker treatment is responsible for improving patient outcomes, we should see the greatest improvements for patients with diagnoses that can be affected by time-sensitive treatments. We therefore examine in which diagnosis groups we see the biggest impact of the target on patient outcomes. Table A3 in appendix A shows the estimated impact on mortality and wait times within each of the 40 ED diagnoses categories, ordered by the size of the mortality reduction. The largest impacts of the target on mortality rates are found in patients with septicemia and cerebrovascular (stroke) and other vascular injuries. These are all areas, as noted above, in which medical evidence suggests benefits to patients from reduced time to treatment.

Although the impacts are largest among these diagnoses, the total number of patients saved in each diagnosis group will also depend on the number of patients who attend ED

with each diagnosis. For example, septicemia is a relatively rare condition, whereas respiratory problems are common. We therefore use the estimates to compute the number of patients for each broad ED diagnosis who survived for at least thirty days following their ED visit as a result of the target.

The last column of table A3 reports these estimates as the share of total lives saved among patients with full diagnosis information. The aggregate estimates indicate that in 2012/2013, just over 20,000 patients were saved by the target, or around one patient per hospital every three days. Among patients with nonmissing diagnoses, a third of the lives saved are from ED patients attending with a respiratory problem. Gastrointestinal, cardiac, and cerebrovascular diagnoses also explain substantial shares of the lives saved. Although these categories are still relatively broad, they provide reassurance that the majority of the mortality reductions come from serious conditions where timely treatment can plausibly make a difference to patient outcomes.

An alternative way of analyzing which patients are saved by the target is to examine whether we observe reductions in the specific causes of death (ICD-10 codes of primary cause of death). We use indicators for specific causes of death as outcome variables and test whether the target reduced the prevalence of each cause. We begin by classifying deaths into 23 categories according to the first letter of the ICD-10 code, and repeat our analysis with the 23 dummy variables as outcome variables. Table A4 in appendix A shows the results. We find 70% of the reduction in mortality can be explained by reductions in deaths related to circulatory (30.1%), respiratory (25.7%), and digestive (15.0%) conditions. These are all categories that include specific conditions that are likely to be time-sensitive. In contrast, no significant reduction in mortality is attributed to neoplasms (cancers), which include a number of high-mortality conditions unlikely to be time-sensitive.

Table A5 in appendix A repeats this analysis in more detail, examining the ten most common causes of death for ED patients (60% of all patient deaths). Again, the greatest mortality reductions are found for time-sensitive conditions. The largest reductions are in deaths from cerebrovascular diseases. Deaths from chronic lower respiratory diseases, influenza and pneumonia, and ischemic and pulmonary heart diseases are also substantially reduced. In contrast, again no significant changes are seen in mortality associated with cancers of any type. These are conditions that we would expect to be less sensitive to time to treatment in an acute setting and so act as a convincing placebo test when examining the time-to-treatment mechanism.

Although these results do not provide definitive proof that wait time reductions are causing mortality reductions, they do provide reassuring evidence that many of the mortality reductions occur in diagnoses where timely treatment is known to be important and not in areas where it is less so. Wait times, and specifically time to treatment, therefore do appear to play an important role in explaining patient outcomes.

### VIII. Conclusion

The emergency department is a central node of health care delivery in all developed countries. It is the entry point into the hospital for a large share of patients, and decisions made rapidly by ED staff have fundamental impacts on the entire course of care. Despite the complicated nature of these decisions, dissatisfaction remains in most health care systems with the level of crowding in EDs and the speed with which cases are resolved. This has led in recent years to both open competition on ED wait times and to regulatory interventions to reduce those times.

We study one type of regulatory intervention, the four-hour wait target policy enacted in England. We find that this target had an enormous effect on wait times, as illustrated vividly by the spike in the wait times distribution at the four-hour mark. We use well-established bunching methodologies applied to a new setting to estimate that this represents a significant reduction of around twenty minutes, or 8%, in the average wait time of impacted ED patients.

We then turn to assessing how this change in wait times impacted patient care and outcomes. We introduce an econometric framework that allows us to separate the compositional impacts of individuals shifting from after to before the four-hour target from the treatment change effect of the four-hour target on medical decisions. We find this target led to a significant rise in hospital admissions and, despite involving little new treatment, a 5% increase in inpatient spending.

At the same time, we find striking evidence that the target is associated with lower mortality. A 0.4 percentage point reduction in mortality emerges within the first thirty days, amounting to a large 14% reduction in mortality in that interval. This reduction fades slightly over time: after one year it amounts to a 3.1% mortality reduction. Although modest, this effect is large relative to the extra spending, suggesting a cost of extending life by one year of \$43,000 (£31,000). Finally, we exploit heterogeneity across patient types to show this effect arises through reduced wait times, not through increased inpatient admissions, with the majority of mortality reductions occurring in diagnoses where rapid treatment is known to benefit patients.

The implications of our finding is that, unconstrained, EDs in England are not making optimal decisions on patient wait times. By reducing wait times, the four-hour target induced cost-effective mortality reductions. This is likely a lower bound on the welfare gains due to the target, as it does not value the other benefits to consumers from waiting shorter times, although welfare costs may be due to the extra admissions (Hoe, 2022).

Of course, this result applies only to the specific target studied here and does not necessarily imply that other limits would have equal effects. It is also unclear how this result applies to other nations with different means of rewarding or incentivizing EDs. A question raised by our results is why physicians and EDs do not optimize wait times in the absence of the policy. One credible explanation is that physicians are

simply imperfect agents for their patients, a long-standing concern in medical markets (Arrow, 1963). This seems especially plausible in our setting where physicians are dealing with patients prior to their full diagnosis being revealed. Alternatively, physicians may lack information on the relative benefits of timely treatment for certain patients. Unfortunately, we are unable to separate these potential explanations here.

Importantly, however, in both cases our results suggest that ED physicians working in an unconstrained setting appear to systematically keep patients in the ED too long, such that an information-free policy (such as the four-hour target) delivers better outcomes for patients. This suggests that better targeted interventions could potentially deliver further improvements for patient outcomes. More work is clearly needed to understand informational constraints and the proper set of rules and incentives necessary for delivering cost-effective ED care.

### REFERENCES

- Arrow, Kenneth J., "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review* 53:5 (1963), 941–973.
- Best, Michael, and Henrik J. Kleven, "Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK," *Review of Economic Studies* 85:1 (2018), 157–193. 10.1093/restud/rdx032
- Chan, David, "Teamwork and Moral Hazard: Evidence from the Emergency Department," *Journal of Political Economy* 124:3 (2016), 734–770. 10.1086/685910
- "The Efficiency of Slacking Off: Evidence from the Emergency Department," *Econometrica* 86:3 (2018), 997–1030. 10.3982/ECTA13565
- Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri, "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics* 126:2 (2013), 749–804. 10.1093/qje/qjr013
- Cutler, David, *Your Money or Your Life: Strong Medicine for America's Health Care System* (Oxford: Oxford University Press, 2003).
- Diamond, Rebecca, and Petra Persson, "The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests," National Bureau of Economic Research working paper (June 2016).
- Einav, Liran, Amy Finkelstein, and Maria Polyakova, "Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D," *American Economic Journal: Economic Policy* 10:3 (2018), 122–153. 10.1257/pol.20160355
- Einav, Liran, Amy Finkelstein, and Paul Schrimpf, "The Response of Drug Expenditure to Non-Linear Contract Design: Evidence from Medicare Part D," *Quarterly Journal of Economics* 130:2 (2015), 841–899. 10.1093/qje/qjv005
- "Bunching at the Kink: Implications for Spending Responses to Health Insurance Contracts," *Journal of Public Economics* 146 (2017), 27–40. 10.1016/j.jpubeco.2016.11.011
- Gerard, François, Christoph Rothe, and Miikka Rokkanen, "Bounds on Treatment Effects in Regression Discontinuity Designs with a Manipulated Running Variable, with an Application to Unemployment Insurance in Brazil," NBER working paper 22892 (May 2018).
- Hoe, Thomas P., "Are Public Hospitals Overcrowded? Evidence from Trauma and Orthopaedics in England," *American Economic Journal: Economic Policy* 14:2 (2022).
- Hoot, Nathan, and Dominik Aronsky, "Systematic Review of Emergency Department Crowding: Causes, Effects and Solutions," *Annals of Emergency Medicine* 52:2 (2008), 126–137. 10.1016/j.annemergmed.2008.03.014
- Kleven, Henrik J., "Bunching," *Annual Review of Economics* 8 (2016), 435–464. 10.1146/annurev-economics-080315-015234
- Kleven, Henrik J., and Mazhar Waseem, "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *Quarterly Journal of Economics* 128:2 (2013), 669–723. 10.1093/qje/qjt004

- Locker, Thomas, and Suzanne M. Mason, "Are These Emergency Department Performance Data Real?," *Emergency Medicine Journal* 23:7 (2006), 558–559. 10.1136/emj.2005.032748
- McCabe, Christopher, Karl Claxton, and Anthony Culyer, "The NICE Cost-Effectiveness Threshold: What It Is and What That Means," *Pharmacoeconomics* 26:9 (2008), 733–744. 10.2165/00019053-200826090-00004
- National Audit Office, "Improving Emergency Care in England," HC 1075 Session 2003-2004 (2004).
- Newman-Toker, David E., Ernest Moy, Ernest Valente, Rosanna Coffey, and Anika L. Hines, "Missed Diagnosis of Stroke in the Emergency Department: A Cross-sectional Analysis of a Large Population-Based Sample," *Diagnosis* 1:2 (2014), 155–166. 10.1515/dx-2013-0038
- Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy* 2:3 (2010), 180–212. 10.1257/pol.2.3.180
- Saver, Jeffrey L., Gregg C. Fonarow, Eric E. Smith, Mathew J. Reeves, Maria V. Grau-Sepulveda, Wenqin Pan, DaiWai M. Olson, Adrian F. Hernandez, Eric D. Peterson, and Lee H. Schwamm, "Time to Treatment with Intravenous Tissue Plasminogen Activator and Outcome from Acute Ischemic Stroke," *Journal of the American Medical Association* 309:23 (2013), 2480–2488. 10.1001/jama.2013.6959
- Seymour, Christopher W., Foster Gesten, Hallie C. Prescott, Marcus E. Friedrich, Theodore J. Iwashyna, Gary S. Phillips, Stanley Lemeshow, Tiffany Osborn, Kathleen M. Terry, and Mitchell M. Levy, "Time to Treatment and Mortality during Mandated Emergency Care for Sepsis," *New England Journal of Medicine* 376:23 (2017), 2235–2244. 10.1056/NEJMoa1703058
- Shojania, Kaveh G., Elizabeth C. Burton, Kathryn M. McDonald, and Lee Goldman, "Changes in Rates of Autopsy-Detected Diagnostic Errors over Time: A Systematic Review," *Journal of the American Medical Association* 289:21 (2003), 2849–2856. 10.1001/jama.289.21.2849
- Silver, David, "Haste or Waste? Peer Pressure and the Distribution of Marginal Returns to Health Care," *Review of Economic Studies* 88:3 (2021), 1385–1417.