

NBER WORKING PAPER SERIES

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

Jonathan Gruber  
Thomas P. Hoe  
George Stoye

Working Paper 24445  
<http://www.nber.org/papers/w24445>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
March 2018

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, MIT, 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). Author affiliations: Gruber (MIT and NBER); Hoe (University College London and Institute for Fiscal Studies); Stoye (University College London and Institute for Fiscal Studies). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Jonathan Gruber, Thomas P. Hoe, and George Stoye. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers  
Jonathan Gruber, Thomas P. Hoe, and George Stoye  
NBER Working Paper No. 24445  
March 2018  
JEL No. I11,I18

**ABSTRACT**

The emergency department (ED) is a complex node of healthcare delivery that is facing market and regulatory pressure across developed economies to reduce wait times. In this paper we study how ED doctors respond to such incentives, by focussing on a landmark policy in England that imposed strong incentives to treat ED patients within four hours. Using bunching techniques, we estimate that the policy reduced affected patients' wait times by 19 minutes, yet distorted a number of medical decisions. In response to the policy, doctors increased the intensity of ED treatment and admitted more patients for costly inpatient care. We also find a striking 14% reduction in mortality. To determine the mechanism behind these health improvements, we exploit heterogeneity in patient severity and hospital crowding, and find strongly suggestive evidence that it is the reduced wait times, rather than the additional admits, that saves lives. Overall we conclude that, despite distorting medical decisions, constraining ED doctors can induce cost-effective reductions in mortality.

Jonathan Gruber  
Department of Economics, E52-434  
MIT  
77 Massachusetts Avenue  
Cambridge, MA 02139  
and NBER  
gruberj@mit.edu

George Stoye  
Institute for Fiscal Studies  
7 Ridgmount St.  
London, WC1E 7AE  
george\_s@ifs.org.uk

Thomas P. Hoe  
Institute for Fiscal Studies  
7 Ridgmount St.  
London, WC1E 7AE  
thomas.hoe.12@ucl.ac.uk

# 1 Introduction

Perhaps the most complicated node of health delivery in any modern health care system is the emergency department (ED). Patients arrive at the ED 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. And ultimately these providers need to decide whether those going to the ED are to be admitted to the hospital or sent back to their homes – 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 and Aronsky, 2008). Practising 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. 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 the quality of the care that they provide. At the same time, however, it is not clear that ED personnel would maximize patient quality in the absence of such pressures. Emergency rooms are not compensated on the margin based on wait times. Moreover, while health-maximizing ED personnel will internalize the costs of waiting to the extent that they impact patient outcomes, this will be imperfect in the presence of uncertainty and may not account for patient well-being beyond health outcomes. Theoretical ambiguities such as this have motivated a growing number of empirical studies of hospital production in the ED setting (Chan, 2016, 2017; Gowrisankaran et al., 2017; Silver, 2016).

The ‘four-hour wait’ policy in the England provides a natural environment in which to address this critical question. 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 set arbitrary targets for wait times,

---

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

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

initially requiring that 98% of all patients be treated within four hours of arrival. 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, there were strong financial penalties 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.

This policy has been controversial. Some stakeholders have argued that the focus on patient wait times has improved patient care. As one ED nurse quoted in Mortimore and Cooper (2007) said, “it was worse [before the targets were introduced], definitely it just seemed to be more hectic, there were people on trolleys for 12 hours and you’d leave here at 8pm and come back in the morning and there would still be some patients here”. Others have argued that care quality has been sacrificed. One medical student stated, “patients are no longer known by their names or by their conditions, they’re not even known by a number, patients are referred to by their time. By this I mean how long they’ve been in the department, as soon as a patient ticks past 3 hours their name lights up like a Christmas tree. If their stay approaches 3 hours 30, the managers start to appear, they don’t actually care about Mr Jones who is having a heart attack. He’s got to go, wherever it may be, as long as it’s not ED”.

Despite the controversy, there is little consistent evidence from either the UK or other nations that have introduced wait time targets on the impact of those targets on patient costs and health outcomes. This is because the policies are generally introduced nation-wide, 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 wait time policy is that no systematic data on wait times are available before the policy was introduced in 2004.

In this paper we take a different approach. We apply the bunching techniques that have been used widely in other contexts (see Kleven, 2016) to analyze wait times and outcomes. 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. 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. This counterfactual focuses attention on the impact of incentives rather

than technology adoption.

We initially examine the distribution of wait times around the four-hour target, finding a very large spike right at four hours. We then turn to estimating counterfactual distributions of wait times in order to measure the effect of the four-hour policy. We estimate that, relative to the counterfactual, the four-hour target led wait times to be 19 minutes (8%) lower for patients affected by the policy.

We then use these data to study the impact of the policy on patient treatment and outcomes. Without pre-period 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 minimal and testable assumptions we can directly identify policy effects from bunching at the four-hour target.

In particular, to assess the impact of the target on outcomes such as hospital admissions and mortality, we need to separate a ‘composition effect’ (because some patients are moved from after to before four hours of wait time due to the target, and they may not be randomly chosen) and a ‘distortion effect’ (the target itself may have a direct distortion on the treatment of randomly chosen patients). To separately identify the distortion 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. We can test this assumption directly using patient observables, showing that along multiple dimensions there is no difference between these and other patients.

We estimate that there is a significant distortion effect of the English policy. We find that there is more intensive testing of patients in the ED, leading to a modest rise in ED costs. We also find that there is 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% due to 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 outcome improvement that we observe. To do so we exploit heterogeneity across patient groups that are 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 four-hour policy is associated with differential impacts on admission probabilities but little variation in the wait times impacts. We then show that the mortality effect we estimate varies strongly across patient severity, but not across inpatient crowding. Taken together, this evidence suggests that it is the wait time mechanism, and not the admissions mechanism, that is driving our mortality effect.

We contribute to two literatures. First, there is a growing literature that has begun documenting features of hospital production relevant for incentive setting (Chan, 2016, 2017; Gowrisankaran et al., 2017; Silver, 2016). Chan (2016) and Chan (2017), for example, study how ED physicians respond to team environments and work schedules, while Silver (2016) studies peer effects in the ED. Gowrisankaran et al. (2017) also study the ED and estimate different measures of physician skill. Adjacent to these studies, a medical literature has documented robust correlations between mortality rates and measures of ED crowding and wait times (Hoot and Aronsky, 2008). Our contribution is to show how ED production is affected when doctors are put under pressure to make decisions quicker. We find that 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 healthcare 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 and Waseem, 2013), these estimators have now been deployed in other settings such as health insurance (Einav et al., 2015, 2017, 2018), mortgage markets (Best et al., 2017; Best and Kleven, 2018) and education (Diamond and Persson, 2016). We apply these estimators in a healthcare 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 2 provides background information on emergency care in England and on the four-hour target policy. Section 3 describes the data. Section 4 sets out our methodology, beginning with an overview and followed by the details of our analysis of wait times, treatment decisions, and health outcomes. Section 5 describes our results for wait times, treatment decisions and health outcomes. Section 6 explores heterogeneity and mechanisms. Section 7 concludes.

## 2 Background

### 2.1 Emergency care in England

Emergency care in England is publicly funded and is available free at the point of use for all residents. There is no private market for emergency care. The majority of care is provided at emergency departments (EDs) attached to large, publicly owned hospitals. These major emergency departments are physician-led providers of 24-hour services, based in specifically built facilities to treat emergency patients that contain full resuscitation facilities. In 2011/12, 9.2 million patients made 13.6 million visits to 174 emergency departments. In addition, 2.1 million patients made an additional 2.7 million visits to specialist emergency clinics and ‘walk in’ or minor injury centres where simple treatment is provided for less serious diagnoses; as discussed below, we exclude patients from these centres due to 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>3</sup> In 2015/16, there were 11 separate tariffs for ED treatment depending on the severity of the patient and the type of treatments administered.<sup>4</sup> These tariffs ranged from \$77 to \$272 per visit.<sup>5</sup> Revenue from the ED accounted for 5.3% of total hospital income in 2015/16.<sup>6</sup>

Treatment in the ED follows one of two pathways depending upon the method of arrival. Non-ambulance patients register at reception upon arrival, where they must identify themselves

---

<sup>3</sup>Treatments are assigned to a Healthcare Resource Group (HRG), similar to DRGs in the US, with a set of national tariffs for each HRG announced each year by the Department of Health.

<sup>4</sup><https://www.gov.uk/government/publications/confirmation-of-payment-by-results-pbr-arrangements-for-2012-13>

<sup>5</sup>All cost figures in 2017/18 US Dollars. Figures are deflated using the UK GDP deflator, and then converted from sterling to dollars using an exchange rate of 1GBP:1.35USD (US Treasury, 31st Dec 2017, <https://www.fiscal.treasury.gov/fsreports/rpt/treasRptRateExch/currentRates.htm>).

<sup>6</sup>Figures calculated from the 2015/16 UK Department of Health Reference Costs. See: <https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016>

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 either 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 call out. In 2011/12, 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>7</sup> This information feeds into a separate triage process, where patients will be categorized by their severity.

These triage processes sort patients into ‘minor’ and ‘major’ cases. Minor cases require relatively simple treatment, and can often be treated in a short space of time. Major cases are often those who arrive by ambulance, although there are some exceptions to this (for example, 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 will usually require more treatment and investigations 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 emergency department (for example, in the hospital’s ‘urgent care centre’).

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 investigations depend on the symptoms presented by the patients, and range 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.

## **2.2 The 4-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

---

<sup>7</sup>Ambulance staff also provide emergency treatment in the ambulance to patients where required.

or transferred to another hospital within four hours of their arrival. The target level was initially set at 98% when it was first introduced in December 2004, before being relaxed to its current level in November 2010.<sup>8</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>9</sup> Hospital managers who consistently fail to meet this target are likely to be fired, and therefore have a strong incentive to organise emergency care in a way that minimises 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 2011/12, hospitals were fined \$300 for every patient who failed to be treated within 4 hours if the hospital missed the overall 95% target during that week.<sup>10</sup> This compares to an average payment of \$140 per patient in the same year. In 2015, a report commissioned by a number of hospitals indicated that public hospitals paid \$325 million in fines due to missed performance targets (including the 4 hour target), with total penalties equal to around a third of the average deficit of public hospitals in that year.<sup>11</sup>

Hospital staff therefore face pressure from hospital management to meet the target. As a result, the organisation of EDs has changed significantly since the target was introduced.<sup>12</sup> Changes include the use of new IT systems, which track patient wait times in real time. The exact systems vary by hospital, but will indicate when patients reach particular waiting thresholds (e.g. 3 hours) and alert physicians (for example through changing the colour of the computer screen).<sup>13</sup> Most departments also now employ specific members of staff to monitor the progress

---

<sup>8</sup>Interviews with hospital managers, doctors and regulators suggest that it is the ‘four-hour’ component of the target that matter to hospitals rather than the absolute level of the target. Hospitals attempt to meet the target on a daily basis, and aim to achieve the highest proportion possible. This suggests that certain behaviours, such as relaxing or improving performance in later parts of the reporting period, are unlikely. Consistent with this, we do not find any systematic evidence of differences in our results by time period or at different points of the reporting period.

<sup>9</sup>For example, see <http://www.mirror.co.uk/news/uk-news/ae-crisis-exposed-only-three-9801509>.

<sup>10</sup>This penalty was decreased to \$170 in 2015.

<sup>11</sup><https://www.theguardian.com/society/2016/mar/29/nhs-bosses-slam-600m-hospital-fines-over-patient-targets>

<sup>12</sup>Interviews with senior member of the Emergency Care Improvement Programme (ECIP), a clinically led programme intended to improve the performance of EDs, clearly describe significant changes to the technology used in EDs since the target was introduced. One manager in the programme claimed that “This [the target] is the most monitored part of the entire healthcare system with software specifically designed for it”.

<sup>13</sup>One medical student in an ED describes the IT system in the following way: “Displayed prominently

of all patients against the clock, and to alert physicians that an admission decision is required soon.

## 3 Data

### 3.1 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.<sup>14</sup>

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 definition of the policy. This includes time being examined and treated. We calculate ED wait times as the time elapsed between arrival and the admission decision.

The data also include a hospital identifier, whether the patient is admitted or discharged, details of basic diagnoses, the number and types of ED investigations 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. The data contain the 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>15</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 Healthcare Resource Group (HRG) code, similar to Diagnosis Related Groups (DRGs) in the US. English hospitals are compensated by the government through a system of national tariffs for each HRG.<sup>16</sup> We calculate ‘costs’ for

---

on an electronic whiteboard is a list of all the patients currently in A&E and waiting to be seen, and the second a patient ticks past a 3 hour wait, their name lights up like a Christmas tree in bright red”. See: <https://imamedicalstudentgetmeoutofhere.blogspot.co.uk/2008/03/there-is-338-in-bay-5.html>

<sup>14</sup>Data on EDs is available prior to 2011, covering 2008 and 2010, although data from the earlier period is less complete than in the years we study.

<sup>15</sup>Senior doctors in England are known as ‘consultants’, and are equivalent to attending physicians in the US.

<sup>16</sup>National tariffs are calculated for each HRG on the basis of annual cost reports submitted by hospitals to the UK Department of Health. These tariffs are meant to reflect the average cost of providing the procedure.

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 30 day period to estimate the cost associated with each ED visit and any follow-up treatment.

Mortality outcomes are recorded in administrative records made available by the UK 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 the date of death for all individuals who died in the UK, or UK citizens who died abroad, between April 2010 and March 2014. We create indicators of whether a patient dies within 30, 90 and 365 days of an ED visit. An indicator of in-hospital mortality is also calculated using HES.

### 3.1.1 Sample construction

Our analysis focuses on a sample of emergency patients treated in ‘major’ emergency departments.<sup>17</sup> We exclude patients treated at specialist clinics that treat only particular diagnoses (e.g. dental) and minor injury (‘walk in’) centres. Patients treated by these units typically have simple diagnoses and short wait times, and are therefore unlikely to be affected by the target. This excludes 18% of emergency visits.

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. Dropping patients with some missing information reduces the number of visits in the sample by 14.5%.<sup>18</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.

### 3.1.2 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

---

Payments are then adjusted for unavoidable regional differences in providing care, and unusually long hospital stays.

<sup>17</sup>Major emergency departments are defined as consultant-led providers of 24-hour services, based in specifically built facilities to treat emergency patients that contain full resuscitation facilities.

<sup>18</sup>Results are unaffected by the inclusion of patients with full information relating to treatment times and decisions, but who are missing demographic information.

admission at the same hospital. 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 30-day treatment costs were \$1,676, of which 89% was accounted for by subsequent inpatient treatment. In the short term, mortality among ED patients is relatively rare. 2% of patients died within 30 days of visiting the ED. This increases to 3% over a 90 day period, and 5% during the following year.

Table 1 also shows summary statistics separately for visits that result in 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 30 days, 16% over a year) are substantially higher than in the main sample. ED treatment is more intense for this sample, with a higher mean number of treatments and investigations than in the main sample. Their treatment is also more expensive, with an average total cost over a 30-day period of \$4,762.

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 analysing the impact of the target.

Figure 1 shows the distribution of ED wait times. There is a noticeable discontinuity in the proportion of patients who exit the ED in the period immediately prior to 4 hours. This spike is unlikely to naturally occur, and is instead induced by the target. We cannot illustrate the absence of this spike prior to the wait times target, since we do not have systematic data available from that period. But it is worth noting, as we do in Online Appendix Figure 1, that such a spike is not present in data on ED wait times from a major U.S. hospital.<sup>19</sup>

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 that this is not the case. First of all, 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 2011/12. If recoding explained the spike then those hospitals should do more recoding to avoid the penalty altogether. Second, we show below that

---

<sup>19</sup>Of course, different ED objectives and technologies across countries means that the U.S. data does 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.

there are comparable spikes in a number of real outcomes, such as hospital admissions, costs, and mortality, which are inconsistent with this simply being a coding response.

## 4 Empirical methodology

We now set out our empirical methodology. We begin with an outline of our approach and then describe our analysis of wait times followed by our analysis of treatment decisions and health outcomes.

### 4.1 Overview

A key challenge when analysing the four-hour target is that without pre-policy data or a control sample, quasi-experimental methods cannot be used to construct the counterfactual outcome. 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 that these methods can be used in our setting to estimate the counterfactual outcomes that would occur if the target were removed but other aspects of hospital production were held constant. This allows us to quantify the short-run impact of the policy.

We first apply a bunching estimator to the distribution of wait time outcomes. This involves interpolating how the wait time distribution would look in the absence of the target. As is typical in other bunching settings, we make a ‘local effects’ assumption; namely, that the target only affects the wait time distribution within a certain segment of the distribution. We argue that this assumption holds if hospitals do not substitute resources between patients located in different segments of the wait time distribution, and present empirical evidence that supports this assumption. The estimated counterfactual distribution from the bunching estimator allows us to quantify the impact of the target on wait times.

We then turn to an analysis of treatment decisions and health outcomes. Plotting these outcomes conditional on the wait time shows that they also exhibit ‘bunching’ at the four-hour discontinuity point. Figure 2 gives an example for the likelihood of inpatient admission. The plot shows that admission odds are generally increasing with wait times, and there is a clear spike in admission odds at 240 minutes. Our analysis decomposes this spike into two channels.

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 ‘post-threshold movers’). Since admission probabilities are increasing with wait time, this movement of patients would increase the observed pre-threshold admission probability even if the target led to no additional admissions. This effect arises purely because the target changes the composition of patients observed at each wait time.

There is also potential for a ‘distortion effect’ if the target has a direct effect on treatment decisions and health outcomes. The distortion 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 these two effects we construct a ‘composition-adjusted counterfactual’ (CAC). This is the outcome that would occur in the presence of composition effects but the absence of distortion effects. Since the observed data contains both effects, the difference between the observed data and the CAC identifies the distortion effect. Estimates of the distortion effects and tests of whether these are significantly different from zero are the central results of this paper.

We construct estimates of the CAC by first showing it can be written as a weighted average of counterfactual outcomes for patients situated in different parts of the wait time distribution. We then argue that the required counterfactual outcomes can be constructed by applying bunching techniques to the *expected outcomes conditional on the wait time*.<sup>20</sup> This relies on a ‘no-selection’ assumption about the distribution of post-threshold movers: that those patients moved forward in time are representative of all post-threshold patients.

To evaluate the validity of the no-selection assumption, we devise a test based on observable patient characteristics such as age. These variables, conditional on the wait time, also exhibit bunching at the four-hour point but in these cases the spike can only be explained by a composition effect since there is no distortion effect by definition. If the no-selection assumption is valid then for these variables the observed data and the CAC should be equal. Tests of this hypothesis therefore act as a placebo test, where rejection of the null hypothesis would suggest that the no-selection assumption has been violated. We pass these placebo tests for a range of demographic variables.

---

<sup>20</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.

We proceed by outlining each of these steps and assumptions more formally.

## 4.2 Wait times

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 non-targeted 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 10-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

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

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.

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:*

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

As we explain shortly, 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>21</sup>

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

---

<sup>21</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 their behaviour. 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.

an iterative procedure that equates the excess mass in the period  $[w^-, w^*]$  with the missing mass in the period  $(w^*, w^+]$ .<sup>22</sup> In the baseline analysis we use a polynomial of order 10 and set  $w^- = 180$ . After applying the iterative procedure this produces an upper cut-off of  $w^+ = 400$ .

The observed data and our estimated counterfactual distribution are shown in Figure 3, which indicates that the target moves a number of patients from the post-threshold period to the pre-threshold period ('post-threshold movers'). We later use these distributions to estimate the impact of the target on wait times.

#### 4.2.1 Testing for substitution effects

If hospitals respond to the target by substituting time or resources between patients inside the exclusion window and those outside the window then Assumption 1 will not hold. A particular concern is that the target may induce doctors to substitute time or resources from patients typically discharged early in the wait time distribution (often high severity patients with unambiguous symptoms, e.g. knife attack victims) to patients that might be at risk of breaching the four-hour target (often high severity patients with uncertain diagnoses, e.g. headaches). By making this type of substitution, doctors would extend some shorter wait times in order to treat a greater proportion of patients within four-hours and thus perform better relative to the target. The wait time distributions, with and without the target, would then differ outside of the exclusion window and violate Assumption 1.

We test for two types of substitution effects. The first is 'planned substitution', where the target causes a permanent change in the priority given to certain types of patients. The second is 'temporary substitution', where the target causes short-term deviations from planned priorities when the ED is momentarily overrun with patients.

To test for planned substitution effects we exploit variation in expected volumes of ED arrivals. This variation changes how tightly the target binds since when there are higher volumes of arrivals the target is more challenging to meet in relative terms. With planned substitution effects we anticipate that hospitals would change patient prioritisation across periods that are expected to be more or less busy. While the volume of arrivals may be correlated with other factors, such as the number of doctors scheduled to be on shift, this would not necessarily impact the patient prioritisation that we compare in this test.<sup>23</sup>

---

<sup>22</sup>This implicitly assumes that the target does not affect patient demand for ED care in the short-term.

<sup>23</sup>One potential concern could be that any increase in scheduled doctors may offset any increase in expected arrivals. If we repeat the same test but use shocks to ED arrivals then we find similar results.

Figure 4 plots average wait times for each percentile of predicted mortality (as a measure of severity) for patients that arrive during ‘busy’ and ‘non-busy’ periods. We define busy periods by first predicting the number of patients present in the ED during each hour in our data, using a regression with hospital-specific week-of-year, day-of-week, and hour-of-day fixed effects. We then divide periods into the top-third of predicted volumes (busy) and bottom-third of predicted volumes (non-busy).

The plot shows that higher severity patients typically have longer wait times, and that busier periods have longer wait times for patients of all severity. Most importantly for our purposes, the relative wait times of high and low severity patients are very similar in both types of period. This suggests that as the target binds more or less tightly, hospitals maintain the same prioritisation of patients and there are no planned substitution effects.

To test for temporary substitution effects, we examine whether there is any evidence that hospitals substitute resources away from patients that we would expect to exit in the early part of the distribution in order to ensure that patients approaching the target do not wait over 4 hours. Intuitively, we compare the wait times of newly arrived patients on the basis of how many patients in the ED have waited almost four hours. If there are temporary substitution effects 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.

We examine two groups of newly arriving patients. We first focus on ‘early exit’ patients (those predicted to have wait times below 180 minutes, such that they exit prior to the exclusion window) and regress their wait times on the volume of existing patients wait ahead of them at each 10-minute interval of the queue. We then compare these results to an equivalent analysis of ‘late exit’ patients (those predicted to have wait times above 180 minutes). The late exit group act as a control group in the sense that Assumption 1 allows for temporary substitution effects to occur for this group (inside the exclusion window) but not for the early exit group (outside the exclusion window). We predict early or late exit using a regression of wait times on age, gender, diagnosis fixed effects and an ambulance indicator.

To implement the test we aggregate the data to the hospital-period level, where periods are defined at 10-minute intervals, and estimate the following equation

$$w_{ht}^g = \sum_k \beta_k q_{h,t-k} + \mu_{hw} + \delta_{hd} + \gamma_{hp} + e_{ht} \quad (3)$$

where  $w_{ht}^g$  is the mean wait time for newly arriving patients of type  $g$  (early or late exit) at hospital  $h$  in period  $t$  (e.g. between 12:01 and 12:10),  $q_{h,t-k}$  is the number of existing patients waiting ahead in the queue at horizon  $t-k$  (e.g. the number of patients that have been waiting 1-10 minutes, 11-20 minutes, and so on), and  $\mu_{hw}$ ,  $\delta_{hd}$  and  $\gamma_{hp}$  are hospital-specific week-of-year, day-of-week, and period-of-day fixed effects.

Figure 5 presents the estimated  $\beta_k$  coefficients from Equation (3). We normalise coefficients so they can be interpreted as the impact of a one standard deviation increase in the queue length at each horizon on newly arriving patients' wait times. Looking first at the early exit group, the plot shows that longer queues increase wait times and the impacts decline with the time horizon. There is no evidence of disproportionate impacts around the four-hour threshold. Looking now at the late exit group, there is again evidence of longer queues increasing wait times but for this group there is clear evidence of a discontinuity at the four-hour threshold. This indicates that, for the late exit group, doctors actively substitute resources away from newly arriving patients towards those patients that are at risk of breaching the target. These results suggest that there are temporary substitution responses for patients predicted to be within the exclusion window (late exits) but not for those predicted to be in the earlier part of the distribution (early exits).

Taken together, Figures 4 and 5 suggest that there are no planned substitution responses, and temporary substitution responses do not occur outside of the exclusion window. This is consistent with Assumption 1.

#### 4.2.2 Interpreting the counterfactual

The counterfactual that the bunching estimator delivers in our context is the short-run outcome that would occur if the four-hour discontinuity in incentives were removed. The counterfactual holds constant other aspects of hospital production, such as patient prioritisation, capital and labour 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.

Our counterfactual differs from the pre-policy or long-run outcomes. To give an example of the difference, we know from anecdotal evidence that the pre-policy outcome had different production inputs (particularly the volume of staff) and different production technology (e.g. IT systems). The full policy impact relative to the pre-policy situation would include the impact

of these changes as well as the discontinuity in incentives introduced by the target.

We refer to our results as the ‘impact of the target’ for brevity but with the above understanding in mind. This interpretation applies to the results for wait times and other outcomes.

### 4.3 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. We first introduce some notation to define the different channels through which the target can affect outcomes and then show how we identify and estimate the ‘distortion effects’ of the target.

#### 4.3.1 Composition and distortion effects

Letting  $y_t$  be an outcome (treatment decision or mortality outcome) and  $w_t$  be the wait time in regime  $t \in \{0, 1\}$ , we define two conditional expectation functions. The first is  $E[y_t | w_t]$ , which is the expected outcome conditional on the wait time. This allows us to express average outcomes (either in the targeted or non-targeted regime) for groups of patients located in different parts of the wait time distribution (either in the targeted or non-targeted regime). For example, the observed data can be written as  $E[y_1 | w_1]$ . It is also possible to think about  $E[y_0 | w_0]$ , outcomes in the absence of the target, and combinations such as  $E[y_0 | w_1]$  which are the outcomes in the non-targeted regime for patients at certain points of the wait time distribution in the targeted regime.

We also define  $E[y_t | w_1, w_0]$ , which is the expected outcome for patients with wait time  $w_1$  in the targeted regime and wait time  $w_0$  in the non-targeted regime. This notation allows us to denote outcomes for groups of individuals that have had a change in wait time due to the target. For example,  $E[y_t | w^- < w_1 \leq w^*, w^* < w_0 < w^+]$  is the expected outcome for post-threshold movers. Since we will repeatedly refer to this and other related groups, we abbreviate these conditioning inequalities in the following way:  $E[y_t | \underline{w}_1^-, \underline{w}_0^+]$ .

Using this notation we can decompose the observed outcomes in the pre-threshold period. Note that, from the wait time analysis, we know that the target causes a number of patients to shift from the post-threshold to the pre-threshold period (‘post-threshold movers’). So with the target, outcomes in the pre-threshold period are a weighted-average of pre-threshold non-movers and post-threshold movers. Abbreviating the pre-threshold period as  $\underline{w}_1^-$ , outcomes can

be written as

$$E[y_1 | \underline{w}_1^-] = \rho E[y_1 | \underline{w}_1^-, \underline{w}_0^-] + (1 - \rho) E[y_1 | \underline{w}_1^-, \underline{w}_0^+], \quad (4)$$

where  $\rho \equiv [F_0(w^*) - F_0(w^-)] / [F_1(w^*) - F_1(w^-)]$  and  $F_t$  is the *cdf* of wait times. The parameter  $\rho$  is defined by the observed and counterfactual wait time distributions, where  $\rho$  is the proportion of pre-threshold non-movers and  $1 - \rho$  is the proportion of post-threshold movers.

The composition and distortion effects are then defined as follows.

**Definition 1** (Composition effect). *The composition effect is the change in expected outcomes conditional on the wait time that occurs in the pre-threshold period because the target shifts some patients into this period from the post-threshold period:*

$$\Delta_C \equiv \rho(E[y_0 | \underline{w}_1^-, \underline{w}_0^-] - E[y_0 | \underline{w}_1^-, \underline{w}_0^+]) + (1 - \rho)(E[y_0 | \underline{w}_1^-, \underline{w}_0^+] - E[y_0 | \underline{w}_1^-, \underline{w}_0^-]) \quad (5)$$

$$= (1 - \rho)(E[y_0 | \underline{w}_1^-, \underline{w}_0^+] - E[y_0 | \underline{w}_1^-, \underline{w}_0^-]). \quad (6)$$

**Definition 2** (Distortion effect). *The distortion effect is the change in expected outcomes conditional on the wait time that occurs in the pre-threshold period because the target has a direct effect on the outcomes in each regime:*

$$\Delta_D \equiv \rho(E[y_1 | \underline{w}_1^-, \underline{w}_0^-] - E[y_0 | \underline{w}_1^-, \underline{w}_0^-]) + (1 - \rho)(E[y_1 | \underline{w}_1^-, \underline{w}_0^+] - E[y_0 | \underline{w}_1^-, \underline{w}_0^+]). \quad (7)$$

With these definitions the observed outcomes in the pre-threshold period can be written as

$$\underbrace{E[y_1 | \underline{w}_1^-]}_{\text{Targeted regime (observed)}} = \underbrace{E[y_0 | \underline{w}_0^-]}_{\text{Non-targeted regime}} + \underbrace{\Delta_C}_{\text{Composition effect}} + \underbrace{\Delta_D}_{\text{Distortion effect}} \quad (8)$$

which can be verified by substituting in Equations (4), (6) and (7) and rewriting the non-targeted regime outcome as  $E[y_0 | \underline{w}_1^-, \underline{w}_0^-]$ .

### 4.3.2 Identification of the distortion effect

To identify the distortion effect we make use of the following definition.

**Definition 3** (Composition-adjusted counterfactual). *The composition-adjusted counterfactual (CAC) is the outcomes from the non-targeted regime in the pre-threshold period that would occur*

in the presence of the composition effect only:

$$E[y_0 | \underline{w}_1^-] \equiv E[y_0 | \underline{w}_0^-] + \Delta_C \quad (9)$$

$$= \rho E[y_0 | \underline{w}_1^-, \underline{w}_0^-] + (1 - \rho) E[y_0 | \underline{w}_1^-, \underline{w}_0^+]. \quad (10)$$

where the second line follows from the definition of  $\Delta_C$ .

With this definition it is straightforward to show that the distortion effect is identified as the difference between the observed data and the CAC:  $\Delta_D = E[y_1 | \underline{w}_1^-] - E[y_0 | \underline{w}_1^-]$ . Moreover, Equation (10) shows the CAC can be constructed as a weighted average of the counterfactual outcomes for two groups, the pre-threshold non-movers and the post-threshold movers, where the weights can be constructed from the observed and counterfactual wait time distributions.

### 4.3.3 Estimating counterfactual outcomes

We now revisit the bunching estimator and show it can be used to obtain the counterfactual outcomes in Equation (10). We require two assumptions for this purpose.

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

$$E[y_1 | w_t] = E[y_0 | w_t] \quad \forall w \notin [w^-, w^* + \varepsilon]. \quad (11)$$

Assumption 2 rules out distortion effects outside of the pre-threshold 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.

The overhang period allows for the empirical fact that the bunching in outcomes extends slightly past the threshold (see Figure 2). We interpret the overhang as being a case of distortion effects for patients that are narrowly discharged or admitted after the threshold. For example, it may be that doctors admit additional patients in attempts to meet the target but not all of the excess admits occur prior to 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>24</sup>

---

<sup>24</sup>Our estimates of the distortion effect, which relate to the pre-threshold 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 pre-threshold period.

**Assumption 3** (No-selection). *Non-targeted regime outcomes conditional on the wait time are comparable for post-threshold movers and post-threshold non-movers:*

$$E[y_0 | \underline{w}_1^-, \underline{w}_0^+] = E[y_0 | \underline{w}_1^+, \underline{w}_0^+] \quad (12)$$

Assumption 3 rules out composition effects in the post-threshold period. It states that after conditioning on the wait time, there is no selection when the post-threshold movers are assigned. This assumption is consistent with doctors randomly selecting which patients get a shorter wait time in response to the target. While this is strong assumption we believe it is plausible. For example, doctors routinely work with incomplete information and this will be exacerbated when they are forced to make earlier admission or discharge decisions (e.g. they may not yet have conducted all tests, or received all test results) and, as a result, may not be able to systematically select which patients to move forward. Importantly, we are also able to evaluate this assumption empirically using placebo tests and discuss this further below.

Together Assumptions 2 and 3 imply that there are no composition or distortion effects outside of the exclusion window  $[w^-, w^* + \varepsilon]$ . We can therefore apply the bunching estimator in the same way as before 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 directly gives us  $E[y_0 | \underline{w}_1^-, \underline{w}_0^-]$  and, given Assumption 3, also provides us with  $E[y_0 | \underline{w}_1^+, \underline{w}_0^+]$ , which are the two terms required to construct Equation (10).

Figure 6 presents an example showing the observed data and our estimated counterfactual for the likelihood of inpatient admission, where the exclusion window is highlighted in grey and we have set  $\varepsilon = 20$ .

#### 4.3.4 Testing for distortion effects

Recalling the definition  $\Delta_D = E[y_1 | \underline{w}_1^-] - E[y_0 | \underline{w}_1^-]$ , and noting that this can now be constructed from the observed data and Equation (10), the test for distortion effects is simply a hypothesis test that  $\Delta_D = 0$ . Estimates of this difference and tests of this null hypothesis form the central results of this paper. We compute statistical significance for the test using non-parametric bootstrapped standard errors clustered at the hospital organisation level.<sup>25</sup>

---

<sup>25</sup>Throughout the analysis we cluster results at the trust (organisation) level. NHS trusts include groups of one or more hospitals in close geographical proximity that share common management. We do not use hospital site codes due to some organisations entering data only at the trust level. All results are robust to clustering at the site level.

Figure 7 provides a visual example of how we construct the CAC and the test of distortion effects for the probability of inpatient admission. The pre- and post-threshold periods are shown in different shades of grey. In each of these periods the horizontal thin dashed line gives the conditional expectation in Equation (10). The CAC, which is a weighted average of these two conditional expectations, is shown in the horizontal thick dashed line in the pre-threshold period.<sup>26</sup> In comparison, the horizontal thick solid line in the pre-threshold period is the mean observed outcome in the pre-threshold period. Finally, the difference between the thick solid and dashed line is the distortion effect,  $\Delta_D$ , which shows that the observed admission probability in the pre-threshold period is too high to be explained by the composition effect alone. In this case we can reject the null hypothesis that  $\Delta_D = 0$ .

#### 4.3.5 Testing the no-selection assumption

In Assumption 3 we rule out the possibility of non-random selection of post-threshold movers. By adapting our test for distortion effects, it is straightforward to generate placebo tests of this assumption based on observable patient characteristics. The key insight that motivates this is that observed demographics, such as age or gender, are by definition subject to composition effects but not distortion effects. Testing the hypothesis that  $\Delta_D = 0$  for any demographic variable is therefore equivalent to testing the no-selection assumption.

This ‘demographic test’ acts as a placebo test, since we are testing for effects in situations where it is known that none should exist. To the extent that these tests indicate that the no-selection assumption does not hold, our estimated distortion effects will be a combination of distortion and composition effects.

Figure 8 provides a visual example of the demographic test using age, which follows the same format as Figure 7. There is again bunching at the four-hour threshold but in this case it cannot be explained by any distortion effects because patient age is unaffected by hospital treatment decisions. Comparing the observed data and the CAC shows that these now lie very close to one another and indeed a hypothesis test cannot reject the null hypothesis that  $\Delta_D = 0$ . This is consistent with the no-selection assumption: the mean age of post-threshold movers is comparable to the mean age of all post-threshold patients.

---

<sup>26</sup>The weights are obtained from the wait time distributions shown in Figure 3.

## 5 Results

We begin this section by first presenting the wait time results. We then present results from the placebo tests of the no-selection assumption, and finally turn to the results concerning treatment decisions and mortality outcomes. We explore the mechanisms behind the mortality outcomes in Section 6.

### 5.1 Wait times

Figure 3 shows the observed wait time distribution and our estimated counterfactual distribution. 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 post-threshold period (240 to 400 minutes) are moved to the pre-target period (180 to 240 minutes); these are the patients we refer to as post-thresholders 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 is successful in achieving its primary aim of reducing wait times. We estimate that the target reduces mean wait times by 7 minutes. This is equivalent to 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 19 minutes, or 8% of their estimated counterfactual mean.

### 5.2 Demographic tests

Table 2 presents the results of the demographic tests. Column (1) presents estimates of the distortion effect and column (2) presents estimates of the distortion effect as a proportion of the counterfactual mean. Panel A presents results using individual demographic variables, where we test using age, a male indicator, and an indicator for whether the patient arrived via ambulance. We cannot reject the hypothesis of no distortions for age and ambulance-arrival, which supports the plausibility of the no-selection assumption. In contrast, we reject the hypothesis of no distortions for the male indicator. This result indicates that post-threshold movers are more likely to be female than the post-threshold non-movers. However, the extent

of this selection effect is small: the difference between the observed and composition-adjusted counterfactual proportion of females in the pre-target period is 0.5 percentage points (1.1% of the baseline).

Panel B in Table 2 presents results for variables that are linear combinations of the three individual demographic variables. We use predicted admission and predicted mortality, where the predictions are obtained from linear regressions of the outcome on a fully interacted set of male, age-category, and ambulance indicators. The  $R^2$  statistic from these predicted regressions is 0.21 and 0.06. The demographic tests for these predicted variables cannot reject the hypothesis of no distortion. The implication of these tests is that even though the gender test rejects the hypothesis, the contribution of gender to salient medical outcomes, as measured by predicted admission and mortality, is low.

Together these results indicate that, with only gender as a minor exception, the demographic tests support the no selection assumption. In practice this means that patients observed with wait times in excess of 240 minutes (post-threshold non-movers) are comparable to those patients that would have had wait times in excess of 240 minutes in the absence of the target (post-threshold movers), and we can therefore use these post-threshold non-movers as the counterfactual for the post-threshold movers.

### 5.3 Treatments and mortality outcomes

Table 3 presents results of the distortion test for a range of treatment decisions and costs. Each row shows results for a separate outcome. Column (1) presents estimates of the distortion effect and column (2) presents estimates of the distortion effect as a proportion of the counterfactual mean.

Panel A presents estimates for treatment decisions in the ED. We find that, controlling for compositional changes, there is an increase in the odds of admission of 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 or hospitals 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 affects on the number of investigations performed in the ED, such as x-rays, blood-test 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 in order 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 has little effect on the treatments that they provide in the ED.

Panel B examines inpatient treatment decisions. For inpatient treatments, in order 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 that the extra admissions do not receive substantial amounts of care in the hospital. That is, these admissions appear to be largely placeholders in order to avoid the four-hour target.

Nevertheless, the additional admits are costly. Panel C of Table 2 examines the impact of the four-hour target on 30-day patient costs. There is a small rise in ED costs of \$3, or two percent of ED costs. But there is a significant increase in inpatient costs of \$126, which is 5% of inpatient costs. That is, even though most patients appear to be only housed in inpatient departments 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.

Table 4 then extends our analysis to look at patient mortality outcomes. We consider mortality at a variety of time frames, ranging from 30 days after entering the ED to 1 year later. We find significant short term declines in mortality. Mortality over 30 days declines by 0.4 percent, or 14% of baseline. 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 four-hour 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 30 days from admission to the ER rise by 5%, while 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 30 days. Assuming no subsequent costs, but also assuming that the mortality impact only lasts one year, this implies a cost per year of life saved of \$43,000.<sup>27</sup> This is low relative to standard valuations of a life-year in the U.S., where typical benchmarks are around \$100,000 (Cutler,

---

<sup>27</sup>This reflects the cost to the government of the policy 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 due to the policy use fewer resources than the average HRG cost.

2003), and at the upper end of valuations in the U.K., where the national benchmarks are set at \$28,000 to \$42,000 (McCabe et al., 2008).

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

## 6 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 since we essentially have one instrument (the target) and multiple changes in patient treatment.

To address this issue we turn to considering heterogeneous impacts across types of patients. That is, we examine whether there are groups of patients where there are differential effects of the four-hour 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.

In particular, we consider two natural sources of heterogeneity. The first is differences across diagnosis. In particular, we divide patients into 36 diagnosis groups.<sup>28</sup> It seems likely that the largest wait time impacts of the target will show up for those who have the most severe diagnoses, since they are the most likely to hit the wait time target. Indeed, Figure 9 graphs the proportion of patients hitting the wait time target (in the counterfactual wait time distribution) against the severity of the diagnosis. Severity is measured by mean predicted 30-day mortality for patients within each diagnosis. In fact, we see that the odds of hitting 240 minutes are much higher for the most severe diagnoses.

We therefore separately compute the wait time reduction effects, and distortion effects for admissions and 30-day mortality for each diagnosis group. We then assess how the heterogeneity across diagnosis groups translates to each of these outcomes.

---

<sup>28</sup>The data assign patients to 40 diagnosis categories, including a ‘missing’ category. We exclude four diagnoses (nerve injuries, electric shock, near drowning and visceral injury) as small samples do not allow us to separately estimate the impact of the target for these groups.

The results of this exercise are shown in Figure 10. Panel A shows that higher severity diagnoses have larger wait time effects. This is sensible since they are most likely to wait the longest without the four-hour policy. But Panel B shows that the effects of the target on hospital admissions is no higher for more severe diagnoses. That is, the more severe diagnoses are getting treated sooner, but are no more likely than others to have that treatment resolve in an extra hospital admission.

Panel C shows the differential treatment effect on mortality by diagnosis category, where black circles correspond to actual mortality outcomes and red triangles correspond to predicted mortality outcomes. The y-axis shows absolute value of mortality reduction, so that a larger value means a larger mortality reduction. Looking at the black circles, there is a clear upward slope showing that the mortality effect of the four-hour target is strongest for the most severe diagnoses. To ensure that selection is not driving our result, the graph also repeats this exercise for predicted mortality. If our assumption of no-selection (Assumption 3) holds, these effects should not be statistically different from zero. The red triangles shows that this is indeed the case, with all estimates clustered around zero and no systematic relationship between the effects of the target on predicted mortality and the severity of the diagnosis.

Given that there is an effect on wait times, but not admissions, this suggests that it is wait time reductions and not increased admissions that are driving the results. Of course, this set of corresponding facts do not prove this causal mechanism because there may be other factors that cause the effects to differ by diagnosis. So to further test this conclusion we consider a second source of heterogeneity.

We next turn to heterogeneity by the degree of inpatient crowding. In times where the inpatient department is more crowded, EDs may be less able to address their wait time targets 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.

To assess this, we divide the data into 50 quantiles depending on how busy the hospital inpatient department is on the day of admission. For each hospital-day, we calculate the daily number of inpatients treated by the hospital, and use this to assign each hospital-day to one of 50 groups in the hospital-specific distribution of inpatient crowding. Patients are then assigned

to each of these groups depending on their day of arrival.<sup>29</sup> To address differences in casemix during busy and quiet periods, we also split patients into two severity groups. ‘Major’ diagnoses are defined as those with a 30-day mortality rate above the overall 30-day mortality rate (1.6%). Interacting the 50 inpatient crowding groups with severity yields 100 groups. For 95 of these groups we have sufficient sample size to independently compute the effects of the target, and therefore across which to examine heterogeneity in effects.

Figure 11 presents the results of this second heterogeneity test. The figure shows the results for these observations, ranked from least crowded to most crowded. Panel A shows that inpatient crowding has a weak, positive relationship with wait times. Panel B shows a strong, negative relationship between crowded inpatient departments and smaller increases in admission. So this source of heterogeneity gives the opposite results of what we saw for severity: a small effect on wait times and a large effect on admissions. Therefore, if our earlier supposition is correct that it is wait times and not admissions that drives our mortality effects, we should see little differential impact on mortality across these groups.

In fact, that is exactly what we see in Panel C in the black circles: there is no significant relationship between the degree of inpatient crowding and the estimated mortality effect. As in Figure 10c, we repeat this analysis with estimated reductions in predicted mortality (which should be unaffected by the target once we adjust for the composition of patients) to show that these results are not driven by selection. The red triangles show that the predicted mortality effects are again all close to zero, with no significant relationship between predicted mortality reductions and inpatient crowding.

We formalize these graphical results in Table 5. The unit of analysis in this table is either diagnosis groups (columns 1-3) or inpatient crowding by severity (columns 4-6). The dependent variable is the distortion effect on mortality in absolute value for each group. The independent variables are the estimated wait time reduction and the distortion effect for admission probability. Essentially, these regressions report associations between the estimated impact on mortality and the estimated impact on wait times and admissions, using a grouping estimator with groups defined by severity or inpatient crowding. A positive coefficient in these regressions can be interpreted as that margin being associated with a larger policy effect on 30-day mortality.

---

<sup>29</sup>We calculate the inpatient census at the daily level as the data do not contain information on time of arrival at, or discharge from, the inpatient department.

Column 1 shows that across the 36 diagnosis groups, those groups with larger wait time effects have larger mortality effects. The estimated coefficient suggests that each additional minute of wait time reduction increases the mortality reduction by 0.001 percentage points. Earlier, we estimated that wait times fell by 19 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 4 of 3-4 percentage points. Column 2, however, shows that there is no impact of the increase in admissions on mortality. And column 3 shows it is still the case that groups with larger wait time effects, but not larger admit effects, have larger mortality effects when we consider both variables together.

Columns 4-6 repeat this exercise using the estimates by inpatient crowding and patient severity. Once again, we have a highly significant relationship between the wait time reduction and mortality reduction, which a coefficient that is 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.

These results are not surprising given the graphical evidence shown above. The bottom line is that heterogeneity associated with wait time variation appears associated with mortality variation, while 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.

## 7 Conclusion

The Emergency Department is a central node of health care delivery in developed countries around the world. 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, there remains dissatisfaction 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 to estimate that this represents a significant reduction of almost 20 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 do so by introducing an econometric framework that allows us to separate the compositional impacts of individuals shifting from after to before the four-hour target from the distortionary effect of the four-hour target on medical decisions. We find this target led to a significant rise in hospital admissions. These admissions do not appear to involve much new treatment, suggesting that they may just be ‘placeholders’ to meet the target. But there is nonetheless a significant rise in inpatient spending of about 5% of baseline.

At the same time, we find striking evidence that the target is associated with lower patient mortality. There is a 0.4 percentage point reduction in patient mortality that emerges within the first 30 days, amounting to a large 14% reduction in mortality in that interval. This reduction fades slightly over time, so that after one year it amounts to a 3.1% mortality reduction. While modest, this effect is large relative to the extra spending, suggesting a cost of extending life by one year of \$43,000. Finally, we exploit heterogeneity across patient types to show that this effect arises through reduced wait times, not through increased inpatient admissions.

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 of 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 there may be welfare costs from the extra admissions (Hoe, 2017).

Of course, this result only applies 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. More work is clearly needed to understand the proper set of rules and incentives for delivering cost-effective ED care.

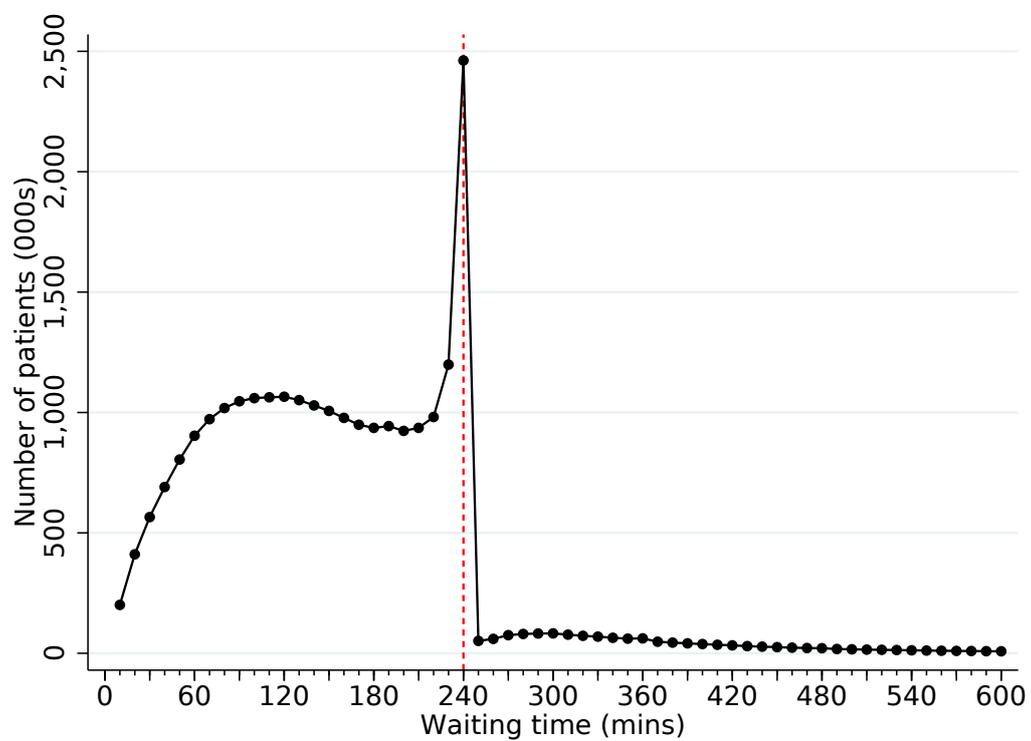
## References

- Best, Michael, Cloyne, James, Ilzetzi, Ethan and Kleven, Henrik J.** (2017). ‘Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches’. Working Paper.
- Best, Michael and Kleven, Henrik J.** (2018). ‘Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK’, *Review of Economic Studies* (85), 157–193.
- Chan, David.** (2016). ‘Teamwork and Moral Hazard: Evidence from the Emergency Department’, *Journal of Political Economy* 124(3).
- Chan, David.** (2017). ‘The Efficiency of Slacking Off: Evidence from the Emergency Department’, *Econometrica* . Forthcoming.
- Chetty, Raj, Friedman, John N., Olsen, Tore and Pistaferri, Luigi.** (2013). ‘Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records’, *Quarterly Journal of Economics* 126(2), 749–804.
- Cutler, David.** (2003), *Your Money Or Your Life: Strong Medicine for America’s Health Care System*, Oxford University Press.
- Diamond, Rebecca and Persson, Petra.** (2016), The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests. Working Paper.
- Einav, Liran, Finkelstein, Amy and Polyakova, Maria.** (2018). ‘Private Provision of Social Insurance: Drug-specific price elasticities and cost sharing in Medicare Part D’, *American Economic Journal: Economic Policy* . Forthcoming.
- Einav, Liran, Finkelstein, Amy and Schrimpf, Paul.** (2015). ‘The Response of Drug Expenditure to Non-Linear Contract Design: Evidence from Medicare Part D’, *Quarterly Journal of Economics* 130(2), 841–899.
- Einav, Liran, Finkelstein, Amy and Schrimpf, Paul.** (2017). ‘Bunching at the kink: implications for spending responses to health insurance contracts’, *Journal of Public Economics* 146, 27–40.
- Gowrisankaran, Gautam, Joiner, Keith A. and Léger, Pierre-Thomas.** (2017). ‘Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments’, *NBER Working Paper No. 24155* .
- Hoe, Thomas P.** (2017), Are Public Hospitals Overcrowded? Evidence from Trauma and Orthopaedics in England. Working Paper.
- Hoot, Nathan R. and Aronsky, Dominik.** (2008). ‘Systematic Review of Emergency Department Crowding: Causes, Effects and Solutions’, *Annals of Emergency Medicine* 52(2), 126–137.
- Kleven, Henrik J.** (2016). ‘Bunching’, *Annual Review of Economics* (8), 435–464.
- Kleven, Henrik J. and Waseem, Mazhar.** (2013). ‘Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan’, *Quarterly Journal of Economics* 128(2), 669–723.
- McCabe, Christopher, Claxton, Karl and Culyer, Anthony J.** (2008). ‘The NICE Cost-Effectiveness Threshold: What it is and What that means’, *Pharmacoeconomics* 26(9), 733–744.

- Mortimore, Andy and Cooper, Simon.** (2007). ‘The ‘4-hour target’: emergency nurses’ views’, *Emergency Medicine Journal* 24(6), 402–404.
- National Audit Office.** (2004). ‘Improving Emergency Care in England’. HC 1075 Session 2003-2004.
- Saez, Emmanuel.** (2010). ‘Do Taxpayers Bunch at Kink Points?’, *American Economic Journal: Economic Policy* 2(3), 180–212.
- Silver, David.** (2016), Haste or Waste? Peer Pressure and the Distribution of Marginal Returns to Health Care. Working Paper.

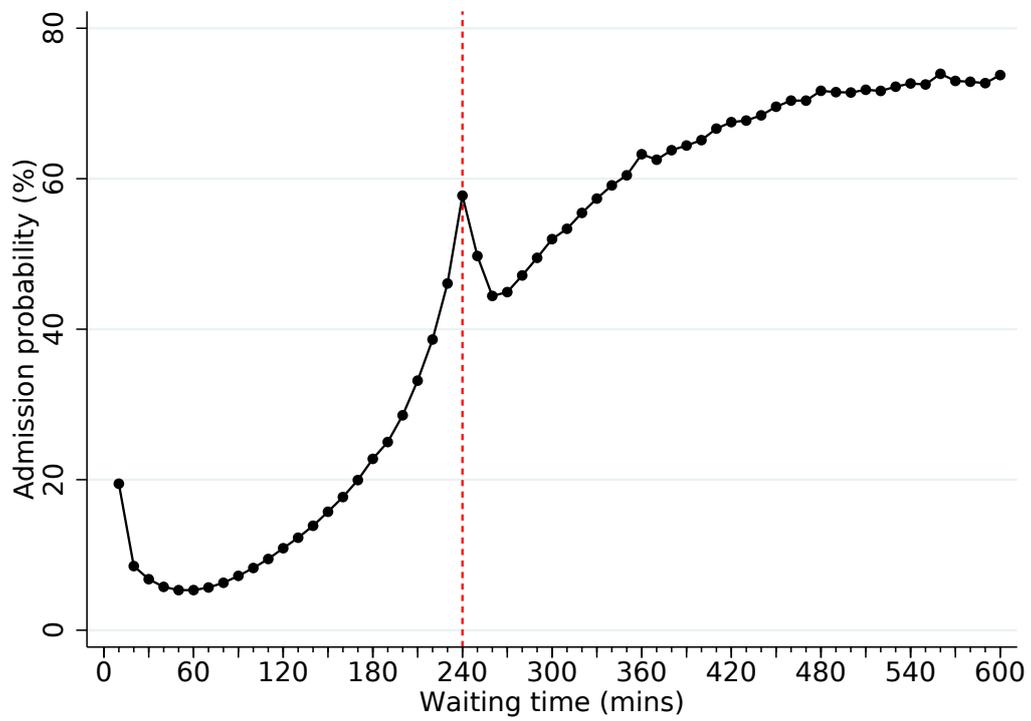
## Figures and Tables

Figure 1: Distribution of wait times



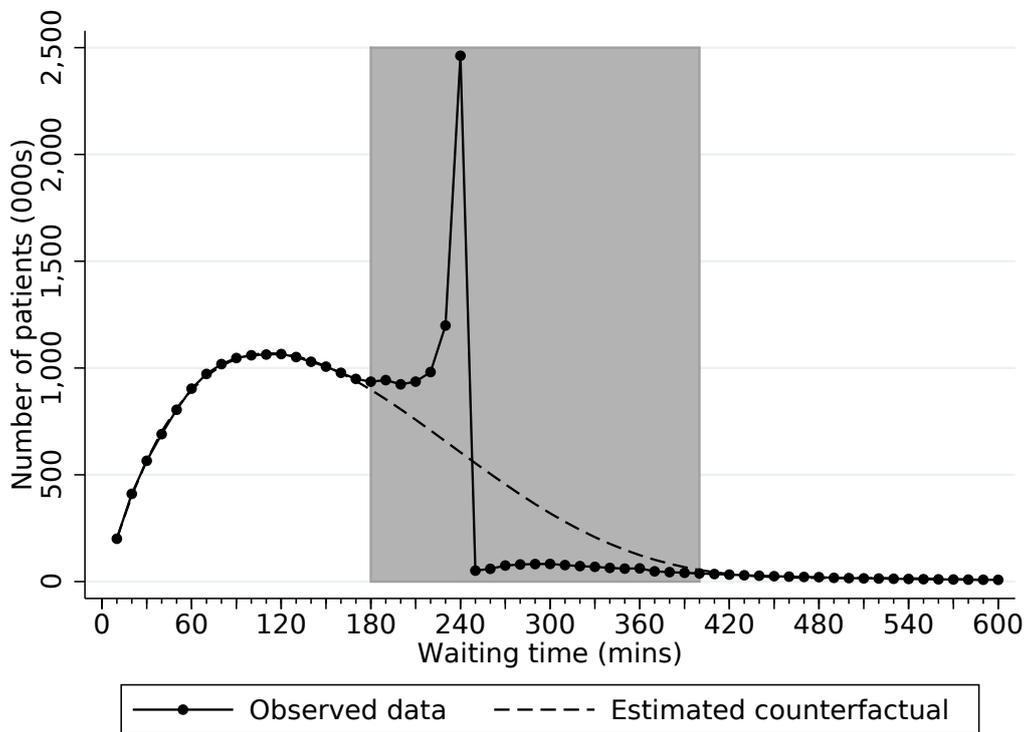
Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy.

Figure 2: Inpatient admission probability conditional on wait time



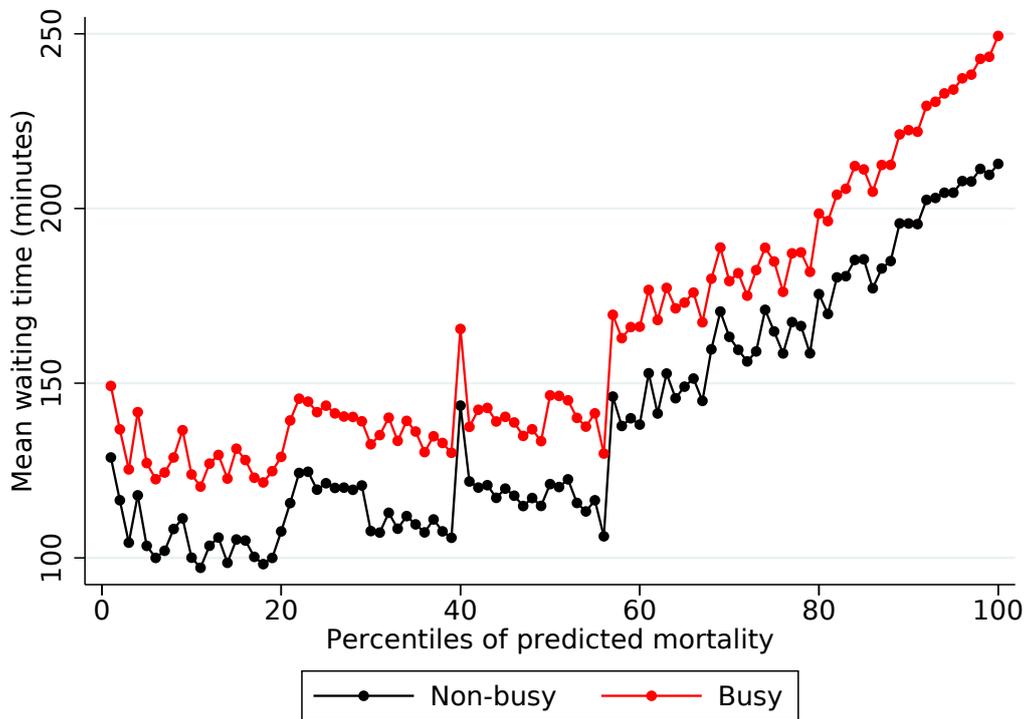
Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy.

Figure 3: Estimated counterfactual wait time distribution



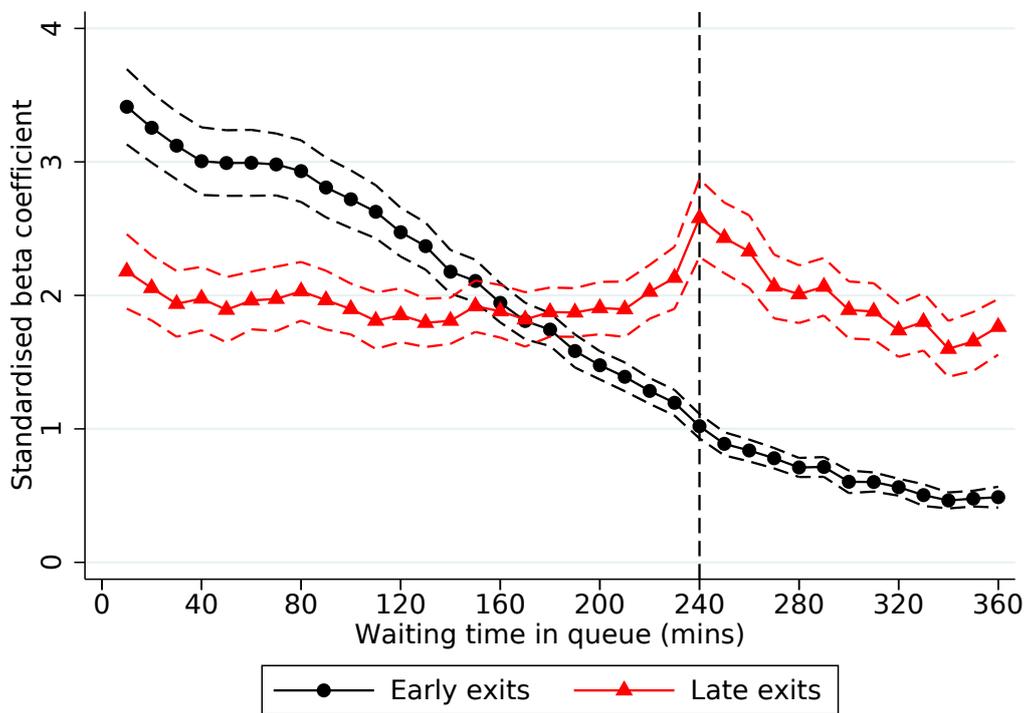
Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy; (4) The estimated counterfactual is obtained from a polynomial regression that omits the exclusion window shown in grey.

Figure 4: Mean wait times by patient severity and expected volumes of ED arrivals



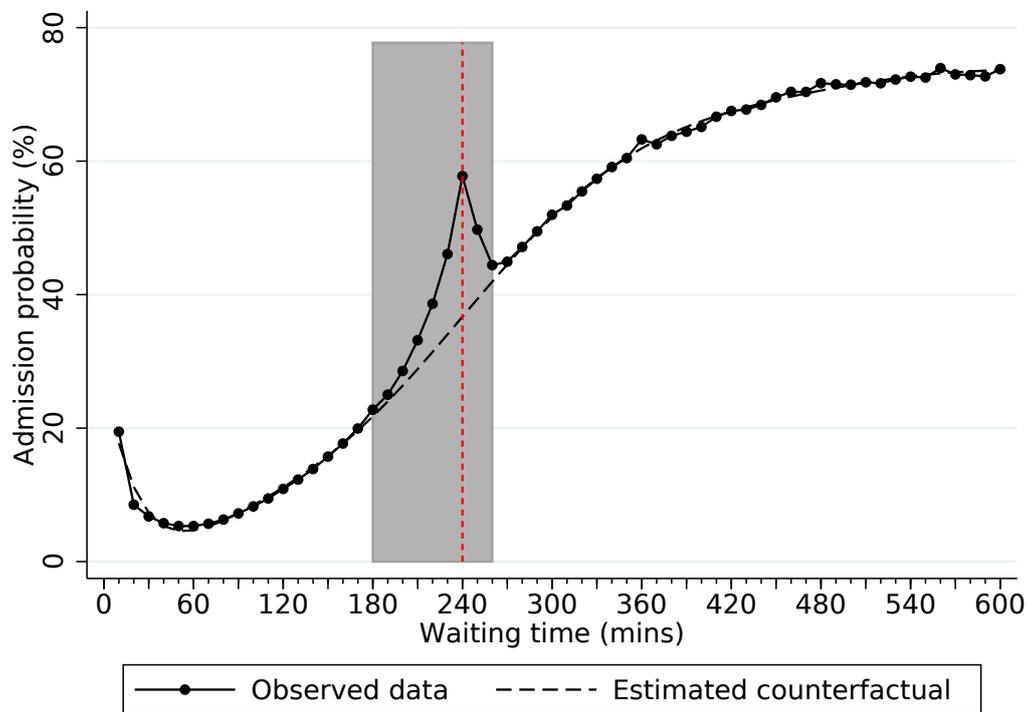
Notes: (1) Wait times defined as the time from arrival in the ED to leaving the ED; (2) Predicted mortality defined using a regression of 30-day in-hospital mortality on a fully interacted set of age, gender, ambulance arrival fixed effects and diagnosis fixed effects; (3) Busy and non-busy periods defined by predicting the volume of ED arrivals during each hour in our data, using a regression with hospital-specific week-of-year, day-of-week, and hour-of-day fixed effects, and then dividing periods into the top-third of predicted volumes (busy) and bottom-third of predicted volumes (non-busy).

Figure 5: Impact of queues on wait times for arriving patients by early- and late-exit groups



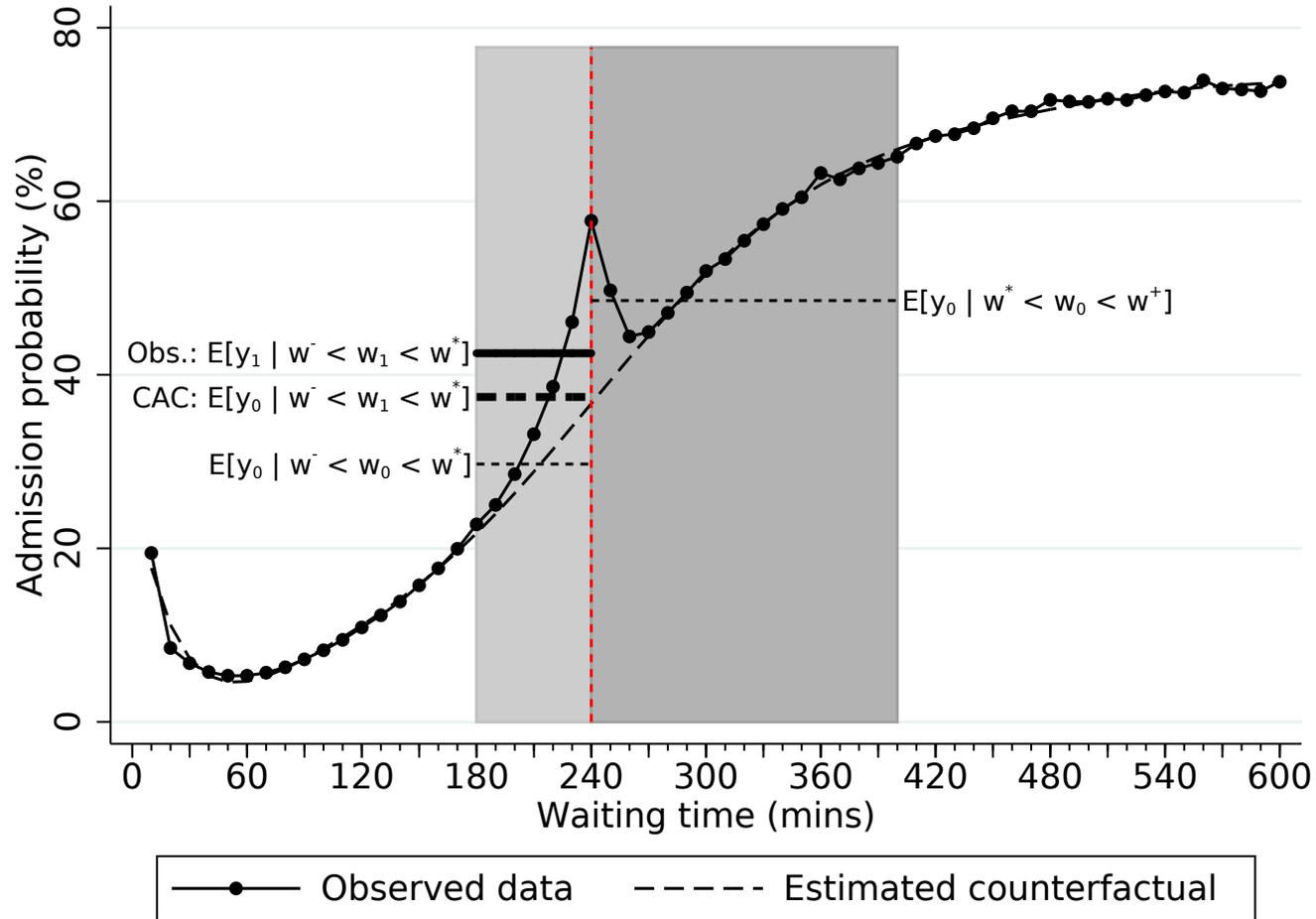
Notes: (1) Wait times defined as the time from arrival in the ED to leaving the ED; (2) We normalise coefficients so they can be interpreted as the impact of a one standard deviation increase in the queue length at each horizon on newly arriving patients' wait times; (3) Early or late exit is defined using a regression of wait times on age, gender, diagnosis fixed effects and an ambulance indicator, from which we predict wait times and group individuals into early (below 180 minutes) and late (above 180 minutes) exit groups.

Figure 6: Estimated counterfactual admission probability conditional on wait times



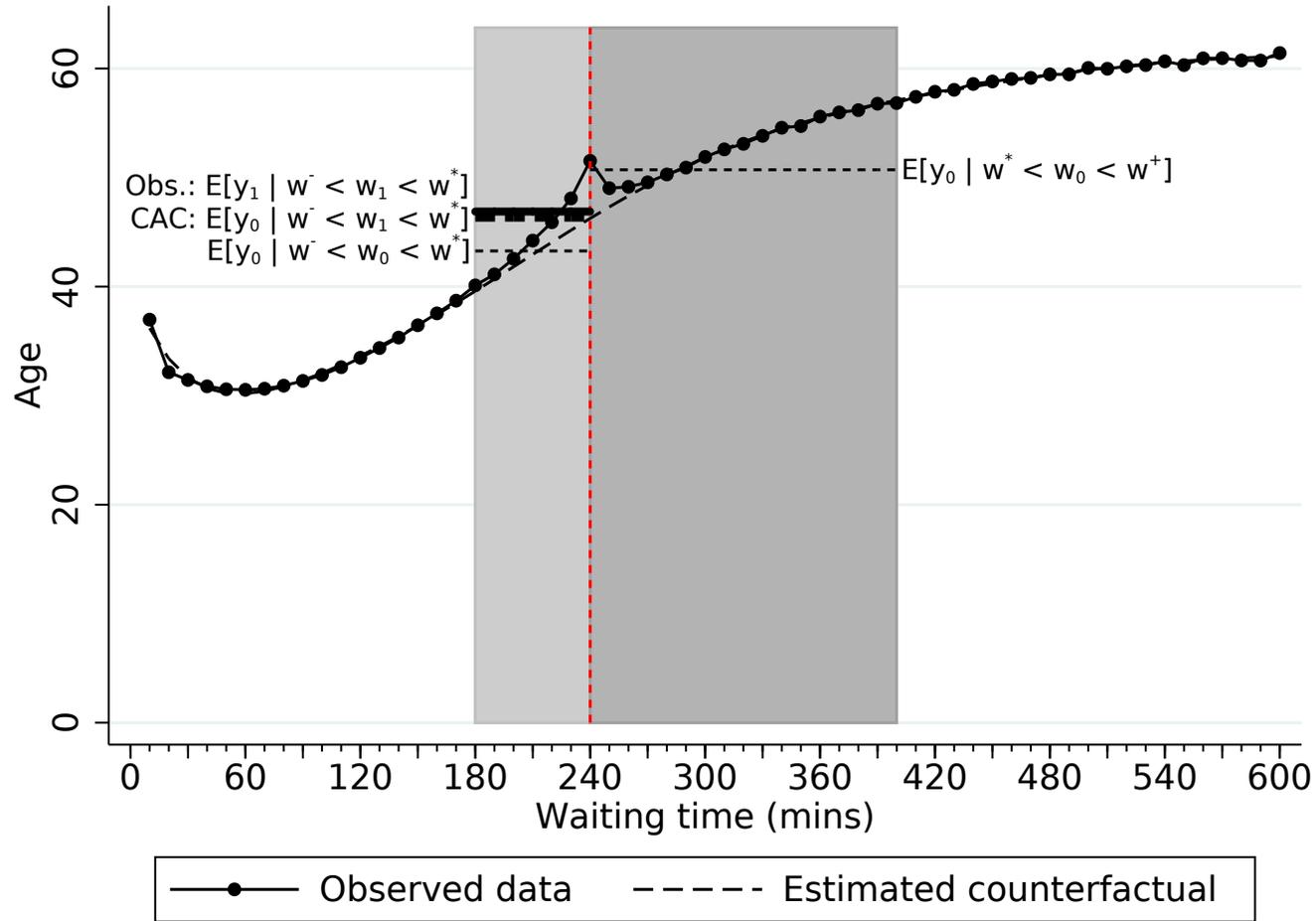
Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy; (4) The estimated counterfactual is obtained from a polynomial regression that omits the exclusion window shown in grey.

Figure 7: Constructing the composition-adjusted counterfactual for admission probability



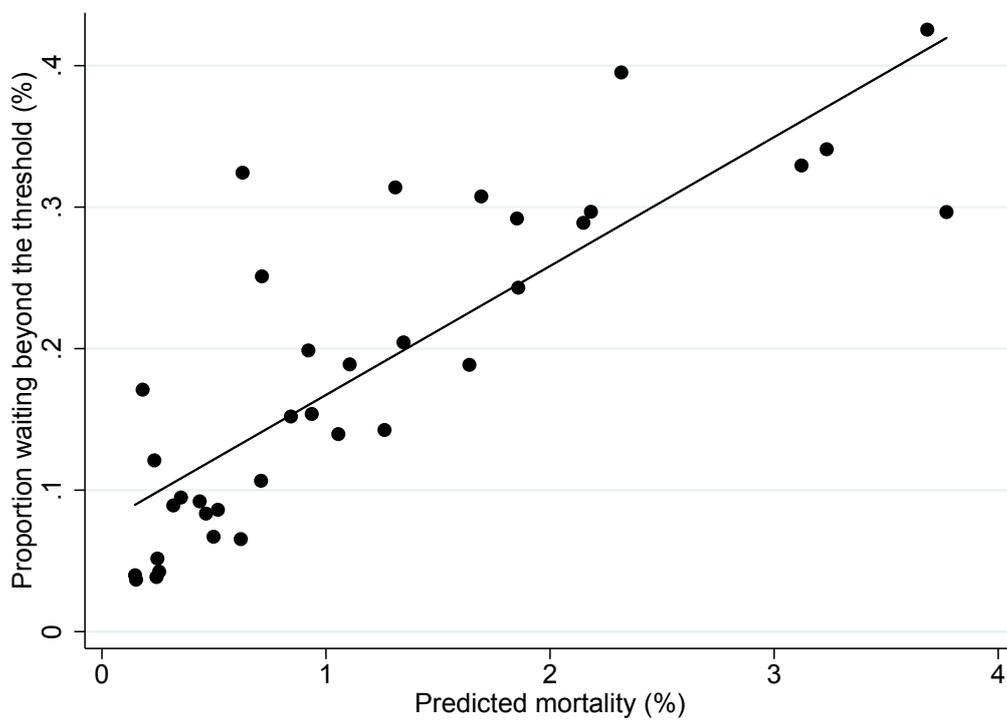
Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy; (4) The horizontal thin dashed lines in the light grey (dark grey) region give the counterfactual outcome in the pre-threshold (post-threshold) period,  $E[y_0 | w_0]$ ; (5) The horizontal thick dashed line in the pre-threshold period is the composition-adjusted counterfactual,  $E[y_0 | \underline{w}_1^-]$ ; (6) The horizontal thick solid line in the pre-threshold period is the observed observed admission probability,  $E[y_1 | \underline{w}_1^-]$ ; (7) The distortion effect is the gap between the thick solid and dashed line,  $\Delta_D = E[y_1 | \underline{w}_1^-] - E[y_0 | \underline{w}_1^-]$ .

Figure 8: Demographic test of the no-selection assumption using age



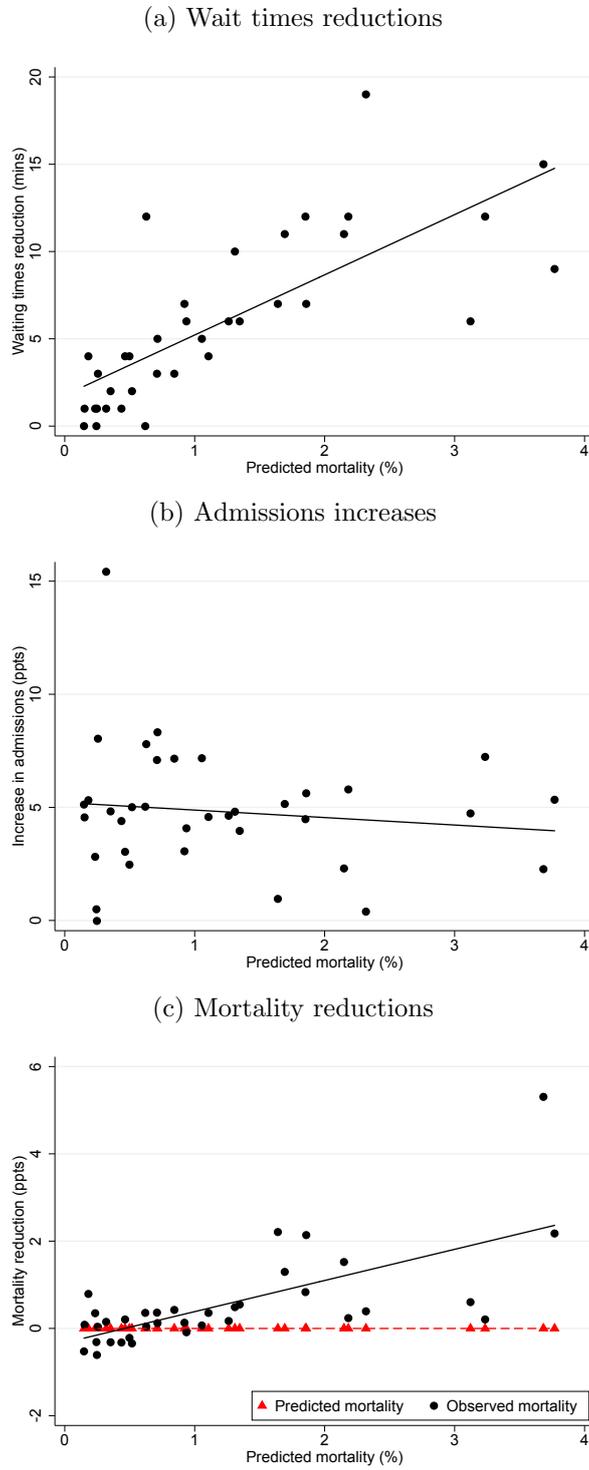
Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy; (4) The horizontal thin dashed lines in the light grey (dark grey) region give the counterfactual outcome in the pre-threshold (post-threshold) period,  $E[y_0 | w_0^-]$ ; (5) The horizontal thick dashed line in the pre-threshold period is the composition-adjusted counterfactual,  $E[y_0 | w_1^-]$ ; (6) The horizontal thick solid line in the pre-threshold period is the observed observed admission probability,  $E[y_1 | w_1^-]$ ; (7) The distortion effect is the gap between the thick solid and dashed line,  $\Delta_D = E[y_1 | w_1^-] - E[y_0 | w_1^-]$ .

Figure 9: Proportion wait beyond the threshold vs. predicted mortality by diagnosis groups



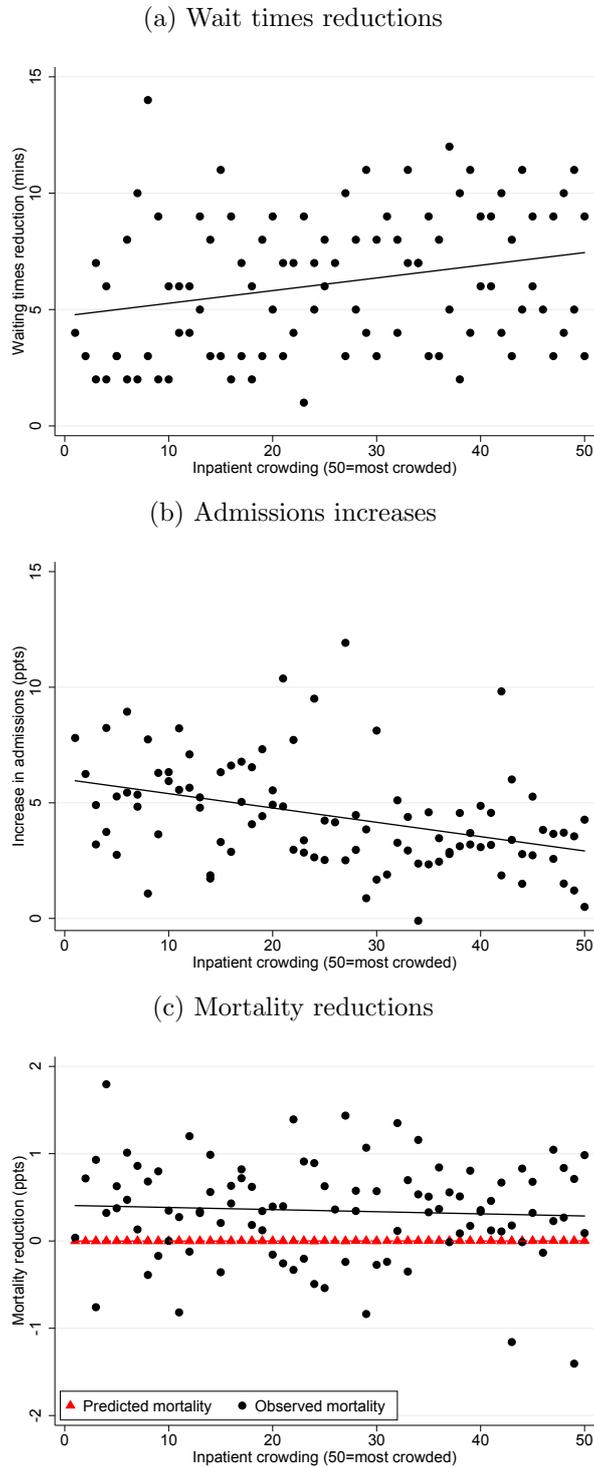
Notes: (1) Each data point corresponds to a diagnosis group average; (2) Proportion waiting beyond the threshold defined using the counterfactual distribution of wait times; (3) Predicted mortality defined using a regression of 30-day in-hospital mortality on a fully interacted set of age, gender, ambulance arrival fixed effects and diagnosis fixed effects.

Figure 10: Estimated effects of the target vs. predicted mortality by diagnosis groups



Notes: (1) Each data point corresponds to a diagnosis group average; (2) Predicted mortality defined using a regression of 30-day in-hospital mortality on a fully interacted set of age, gender, ambulance arrival fixed effects and diagnosis fixed effects.

Figure 11: Estimated effects of the target vs. inpatient crowding by crowding-severity groups



Notes: (1) Each data point corresponds to an inpatient crowding-severity group average; (2) Inpatient crowding groups defined according to the number of inpatients treated per hospital-day, which we then use to split into 50 quantiles; (3) Severity is defined as diagnoses with a mean 30-day mortality rate above the mean overall 30-day mortality rate; (4) Predicted mortality defined using a regression of 30-day in-hospital mortality on a fully interacted set of age, gender, ambulance arrival fixed effects and diagnosis fixed effects.

Table 1: Summary statistics

	All patients		Admitted inpatients	
	Mean	Std. dev.	Mean	Std. dev.
<i>Patient characteristics</i>				
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
<i>Treatment decisions</i>				
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 (mins)	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
<i>Costs</i>				
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
<i>Mortality outcomes</i>				
30-day in-hospital mortality	0.01	0.11	0.04	0.20
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

Notes: (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.

Table 2: Demographic tests of the no-selection assumption

	Distortion effect ( $\Delta_D$ )		CAC mean
	Level (1)	% (2)	Level (3)
<i>Panel A: Individual characteristics</i>			
Age	0.417 (0.284)	0.009 (0.006)	46.468
Male	-0.005*** (0.001)	-0.011*** (0.003)	0.487
Ambulance	-0.002 (0.004)	-0.005 (0.010)	0.440
<i>Panel B: Predicted characteristics</i>			
Predicted admission	0.002 (0.002)	0.006 (0.007)	0.323
Predicted mortality	0.000 (0.000)	0.015 (0.015)	0.019

Notes: (1) CAC mean is measured over the pre-threshold period,  $E[y_0 | \underline{w}_1^-]$ ; (2) Predicted admissions and mortality use regressions with fully interacted variables from Panel A; (3) Bootstrapped standard errors clustered at the hospital trust level (199 repetitions).

Table 3: Estimated distortion effects of the target on treatment decisions and costs

	Distortion effect ( $\Delta_D$ )		CAC mean
	Level (1)	% (2)	Level (3)
<i>Panel A: ED treatment decisions</i>			
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
<i>Panel B: Inpatient treatment decisions</i>			
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
<i>Panel C: Hospital costs</i>			
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

Notes: (1) CAC mean is measured over the pre-threshold period,  $E[y_0 | \underline{w}_1^-]$ ; (2) Predicted admissions and mortality use regressions with fully interacted variables from Panel A; (3) All inpatient variables (e.g. length of stay, costs) take on the value zero for patients that are not admitted; (4) Bootstrapped standard errors clustered at the hospital trust level (199 repetitions).

Table 4: Estimated distortion effects of the target on mortality

	Distortion effect ( $\Delta_D$ )		CAC mean
	Level (1)	% (2)	Level (3)
30-day mortality	-0.004*** (0.001)	-0.138*** (0.019)	0.029
90-day mortality	-0.004*** (0.001)	-0.079*** (0.019)	0.048
1-year mortality	-0.003* (0.002)	-0.031* (0.017)	0.090

Notes: (1) CAC mean is measured over the pre-threshold period,  $E[y_0 | \underline{w}_1^-]$ ; (2) Bootstrapped standard errors clustered at the hospital trust level (199 repetitions).

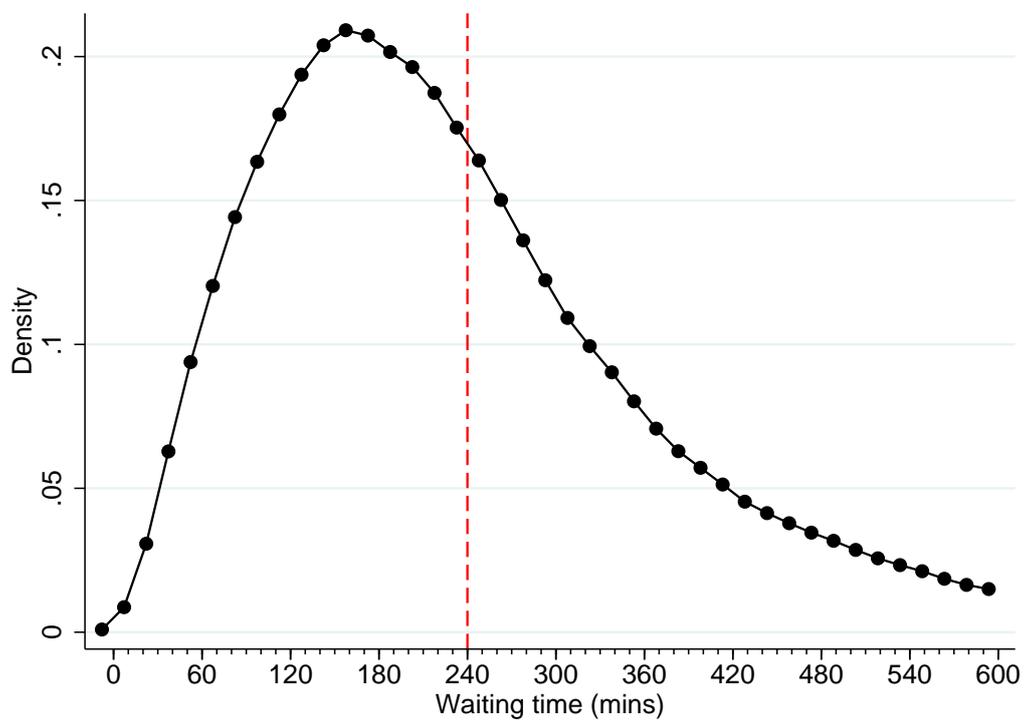
Table 5: OLS regressions of the estimated 30-day mortality reductions on other effects of the target

	Diagnosis groups			Crowding-severity groups		
	(1)	(2)	(3)	(4)	(5)	(6)
Wait time	0.118*** (0.034)		0.115*** (0.034)	0.083*** (0.018)		0.066*** (0.022)
Admission probability		-0.059 (0.065)	-0.029 (0.058)		-0.088*** (0.024)	-0.037 (0.028)
N	36	36	36	95	95	95

Notes: (1) Dependent variable is the absolute value of the target impact on 30-day mortality measured as % of the CAC mean over the pre-threshold period; (2) Independent variables are the absolute value of the target impact on the respective variable, measured as a % of the CAC mean over the pre-threshold period.

## Online Appendix: Additional Figures and Tables

Figure 1: Distribution of wait times at a large hospital in California



Notes: (1) The English data displays a sharp discontinuity in the wait time distribution at four hours (see Figure 1). Here we present the wait time distribution from a large hospital in California to illustrate that the discontinuity in the English data is unlikely to naturally occur, and is instead induced by the target; (2) We thank David Chan for providing the data for this chart.