From Bottlenecks to Bedside: Returns to NHS Spending on Emergency Care*

Ye Yuan (Job Market Paper)

Damian Clarke

November 21, 2025

See the latest draft here

Abstract

Health care systems routinely face acute crises that disrupt services and threaten patient outcomes. In response, governments often deploy extra-budgetary funding to cushion these shocks—but how effective are such interventions in improving system performance and health outcomes? This paper provides causal evidence on the effectiveness of crisis-responsive health spending. We study a £700 million emergency "Discharge Fund" introduced in England in December 2022 to ease NHS winter pressures by accelerating hospital discharges in emergency care. Using newly assembled administrative data from ambulance services and hospital records, we find that targeted and agile spending can rapidly relieve operational bottlenecks and generate measurable health returns. Following the funding rollout, ambulance handover delays were halved and average response time for the most critical patients fell by 2.5 minutes, relative to levels immediately before the rollout. Patient outcomes improved correspondingly: the rate of return of spontaneous circulation upon hospital arrival rose by 4.2%, and 30-day survival increased by 3.6%. Mortality rates for all deaths also declined modestly in affected areas.

JEL codes: I18, J18; H54.

Keywords: Health inputs; emergency care system; ambulances; mortality; healthcare provision.

*Yuan: Department of Economics, University of Exeter, yy429@exeter.ac.uk; Clarke: Department of Economics, University of Exeter and University of Chile, dclarke@fen.uchile.cl.

1 Introduction

Health systems are routinely exposed to acute shocks, such as pandemics, seasonal surges, or extreme climate change (Williams *et al.*, 2023; Rentschler *et al.*, 2021). Such shocks can quickly overwhelm capacity, disrupt routine care, and consequently harm the patients' health outcomes. When these acute pressures are not managed effectively, they could transform into persistent, chronic system stressors and trigger the next crisis (Witter *et al.*, 2023). Evidence from multiple health systems shows that temporary crises can erode long-run resilience by accelerating workforce burnout and attrition among frontline staff, delaying essential treatments, and generating a backlog of patients whose acute conditions evolve into chronic illness (Yoon *et al.*, 2022; Davino-Ramaya *et al.*, 2023; Walshe *et al.*, 2024; Fetzer and Rauh, 2022).

In this context, governments must develop mechanisms that allow for agile responses to sudden surges in demand. Traditional health budgets which are planned and allocated in advance are too rigid to respond quickly to system shocks. When systems come under acute strain, effective policy requires rapid, targeted spending that relieves specific bottlenecks, preserves capacity and prevents temporary disruptions. Such flexible crisis spending has become an increasingly important component of health system management, yet its effectiveness remains poorly understood.

There are much of the health economics literature studies the returns to planned and budgeted health spending, such as targeted investments in neonatal care (Almond *et al.*, 2010), cardiovascular treatment (Cutler *et al.*, 1998; Cutler, 2007), cancer care (Martin *et al.*, 2008), or hospital quality improvements (Doyle *et al.*, 2015; Gruber *et al.*, 2014), while there is relatively less research which are about the effectiveness of agile, crisis-responsive health spending. Such reactive funding is increasingly deployed to stabilize health systems under urgent pressure, yet its causal impacts on both system performance and population health remain largely unquantified.

This paper addressed this gap by studying the short-run effects of an emergency funding intervention in the United Kingdom's National Health Services (NHS). In December 2022, the UK government announced a £700 million "Discharge Fund" to relieve acute winter pressures by accelerating patient discharge and freeing up the hospital capacity. The funding was implemented rapidly across England between December 2022 and March 2023, offering a natural experiment to test the causal effects of agile targeted health spending.

This setting has several desirable features: (1) this funding is additional and centrally allocated, which alleviates the endogeneity concern if the health spending is budgeted while such spending ties to the local characteristics and policy makers' decision making process; (2)we study the additional emergency funding invested in emergency care system over crisis time, where patients are not able to predict the future change in the system performance and manipulate their timing of accessing to the emergency care, which is helpful to explore the clean causal relationship.

We exploit this abrupt injection of money to identify its effects on emergency care system performance and patient outcomes. Using novel administrative data constructed from a series of Freedom of Information (FOI) requests and NHS datasets, we compile multiple data linking emergency care system performance indicators such as ambulance response times and hospital handover delays, and several health outcomes, which includes return of spontaneous circulation (ROSC), 30-day survival rates for cardiac arrest patients, and population-level mortality. Applying a Regression Discontinuity in Time (RDiT) design around the policy cutoff at January 2023 when we observe the improvement of system performance, we estimate the causal impact of this funding shock on both operational and clinical performance.

Our results show that the 2022–2023 emergency funding led to a sharp and sustained improvement in emergency system performance. Time lost to handover delays fell by 50% immediately after the cutoff. Ambulance response times decreased substantially, with 2.5 minutes improvements observed in the highest priority (C1) incidents. These operational gains translated into significant improvements in patient outcomes: the probability of achieving ROSC upon hospital arrival increased by 4.23% (from an average of 26%), and 30-day survival rose by 3.63% (from 9%). We also find suggestive declines in local mortality rates, consistent with broader spillover benefits from reduced emergency congestion.

Our research contributes to several strands of past literature. First, we provide causal evidence on the relationship between emergency health spending and health outcomes, which has rarely been examined in the existing literature on healthcare spending. We complement this body of work by confirming that such targeted investments are indeed effective, and by extending the analysis to short-term, crisis-responsive spending that is agile and dedicated to alleviating the strain on health systems during public health emergencies.

Second, we contribute to the literature on emergency care system performance and health outcomes. Prior extensive research has emphasized how factors such as response time targets (Pell *et al.*, 2001), public/private provision (Knutsson and Tyrefors, 2022; Chan *et al.*, 2023), delays in response (Lucchese, 2024), austerity (Berman and Hovland, 2024), hospital crowding (Hoe, 2019), incentives in emergency departments (Gruber *et al.*, 2018), and technology adoption (Athey and Stern, 2002) affect performance and patient outcomes. We add to this literature by identifying a new channel—system bottlenecks—as a key mechanism constraining emergency care performance during acute shocks. We show that when additional spending is directed toward relieving these bottlenecks (e.g., higher hospital A&E capacity decreases the ambulance handover delays), the system can respond more effectively, yielding immediate and measurable improvements in health outcomes. Our findings thus highlight both a novel mechanism and a concrete policy lever linking emergency system performance to population health.

The remainder of the paper proceeds as follows. Section 2 introduce the UK emergency care system and the recent pressures. Section 3 describe the data and summary statistics. Section 4 illustrates the identification method, identifying assumptions, and robustness checks. Section 5 reports the main findings, including identification checks and mechanism discussion. Section 6 concludes the paper.

2 Background of Urgent and Emergency Care System in the UK

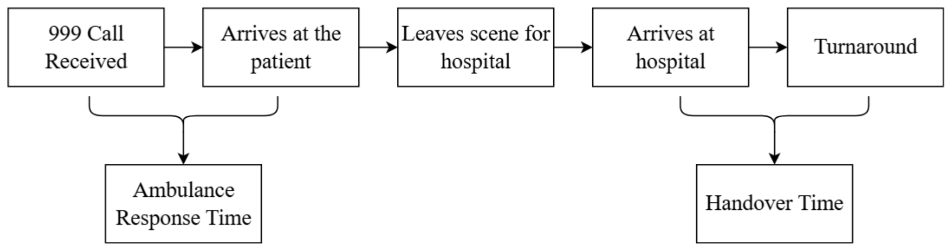
This section introduces the background needed to understand the UK's urgent and emergency care system and the recent pressures it has faced. We first explain how the ambulance system works and how response times are measured. We then describe major shocks, including the COVID-19 pandemic and industrial action in 2022-2023, which created significant strain and also reflected underlying stresses on the system. Next, we outline the government's policy interventions during the 2022-2023 winter crisis. Finally, we show that these interventions were followed by a sharp improvement in ambulance response times from January 2023, which forms the basis of our empirical analysis.

2.1 The UK Urgent and Emergency Care System

Ambulance services in England are organized into regional NHS ambulance trusts, each trust is responsible for emergency medical services across designated geographical areas. There are currently 10 ambulance trusts covering England, with each trust managing its own fleet, staff deployment, and operational protocols while adhering to national performance standards. Since the 2022 Health and Care Act issued, these trusts work closely with Integrated Care Boards (ICBs), which are responsible for planning and funding health services across their local areas. ICBs hold commissioning responsibility for ambulance services and coordinate between ambulance trusts, hospitals, and other healthcare providers to manage patient flow and resource allocation.

Private ambulance services are also playing a significant role in England's healthcare system. If the public ambulance resources are fully utilised, the NHS trusts can dispatch a contracted private ambulance to respond a 999 call. Private ambulance crew are expected to adhere to the same

Figure 1: Ambulance service operation



clinical standards and protocols, the NHS ambulance trusts will document those events in their own data management. They could also provide stand-alone service directly to their customers, while those are not under the management of NHS trusts.

The UK's emergency medical services operate through a coordinated system where residents access emergency care by dialling 999 (though people can also receive emergency care by going directly to hospital Accident and Emergency departments). Figure 1 presents the whole process. Upon receiving an emergency call transferred from the BT operator, call handlers from the local ambulance service's Emergency Operations Centre (EOC) will assess the urgency, and from the time this first call is received, the clock starts on the ambulance response time. The patient or the person who made the call will answer a series of standardised protocol questions about address, contact numbers, symptoms, medical history, and other basic demographic details to provide sufficient information for the call handler to classify the emergency case.

Table 1 shows the ambulance call classifications in different countries of the UK. In England, ambulance calls are divided into four categories according to the patient conditions revealed to the emergency call handlers:

- Category 1 (C1): Life-threatening emergencies (e.g., cardiac arrest), with a target average response time of 7 minutes.
- Category 2 (C2): Serious but not immediately life-threatening emergencies, with a target average response time under 18 minutes.
- Category 3 (C3) and Category 4 (C4): Urgent or less urgent cases requiring assessment or transport, with time standards specifying responses to 90% of calls within 120 minutes (C3) and 180 minutes (C4), respectively.

The Scottish and Welsh ambulance systems use colour-coded classifications with similar structures to England's numbered categories. The Scottish system employs purple (highest priority),

Table 1: Ambulance Call Classifications and Average Standards (Mins)

	England		Scotland		Wales	
Level	Code	Time	Code	Time	Code	Time
Highest	C1	7	Purple	8	Red	8 (65th)
Second	C2	18	Red	8	Amber	–
Third	C3	120 (90th)	Amber	19	Green	–
Fourth	C4	180 (90th)	Yellow	19		

Notes: The information is sourced from the Handbook to the NHS Constitution for England (Department of Health and Social Care, 2023a), the Scottish Ambulance Service New Clinical Response Model Evaluation Report (Stoddart *et al.*, 2019), and the Welsh Ambulance Services NHS Trust Clinical Response Model (NHS Wales Joint Commissioning Committee, 2025).

red, amber, and yellow (lowest priority) designations, while the Welsh system uses red (life-threatening), amber (serious but not immediately life-threatening), and green (non-urgent) classifications.

This initial triage is a prominent clinical process to determine the priority of each call and the appropriate response that is dispatched. For C1 and C2 patients with shorter response time standards, the resources are dispatched immediately to ensure that the sickest patients get help the fastest. Moreover, the triage determines sending the right level of care to the patients, including type of vehicle and level of paramedic skill required. Once the classification is completed on the call, the ambulance crew will be mobilized and sent off. Upon arrival, the response time clock stops, and initial assessment and treatment will be provided. Thus, ambulance response time measures the critical interval from an emergency call to the moment providers reach the patient, which marks the first opportunity to administer formal treatment.

Ambulance response time is a key measure for the performance of emergency medical services. It directly reflects how quickly potential life-saving medical intervention can begin. For a range of time-sensitive conditions, for instance, the "golden hour" principle in trauma or "time is tissue" in cardiac and stroke care, delays in response are strongly associated with higher patient death risks and/or significant morbidity (Pell *et al.*, 2001; Holmén *et al.*, 2020; Alarhayem *et al.*, 2016). Thus, this paper is also motivated by this and aims to examine the effects of longer response times on patients' health outcomes, particularly in light of the increasing operational pressures experienced by the UK emergency care system in recent years, which we will discuss in the following sections.

2.2 Recent Systemic Pressures and Shocks

2.2.1 The COVID-19 pandemic as a systemic shock

The COVID-19 pandemic placed unprecedented strain on the UK's emergency medical services beginning in 2021. The crisis manifested through multiple channels: increased demand for urgent care due to COVID-related complications, reduced ambulance capacity as staff fell ill or were required to isolate, and prolonged hospital handover delays as acute-care facilities became saturated.

Figure 2 illustrates these pressures clearly. The top panel shows a sharp and persistent increase in average response times, with further deterioration after the relaxation of lockdown restrictions in 2021. The middle panel demonstrates rising incident volumes especially for C1 patients, while the bottom panel reveals elevated mortality rates during peak pandemic periods.

In the short term, COVID-19 substantially increased the demand for emergency care, leading to hospital congestion and longer discharge delays, as many patients required extended inpatient treatment before meeting discharge criteria. In the longer term, survivors of COVID-19 have faced a higher prevalence of chronic health complications (e.g., respiratory, cardiovascular, and mental health conditions), increasing ongoing demand for emergency and urgent care services (Tsampasian *et al.*, 2024; Adu-Amankwaah, 2024; Daines *et al.*, 2022).

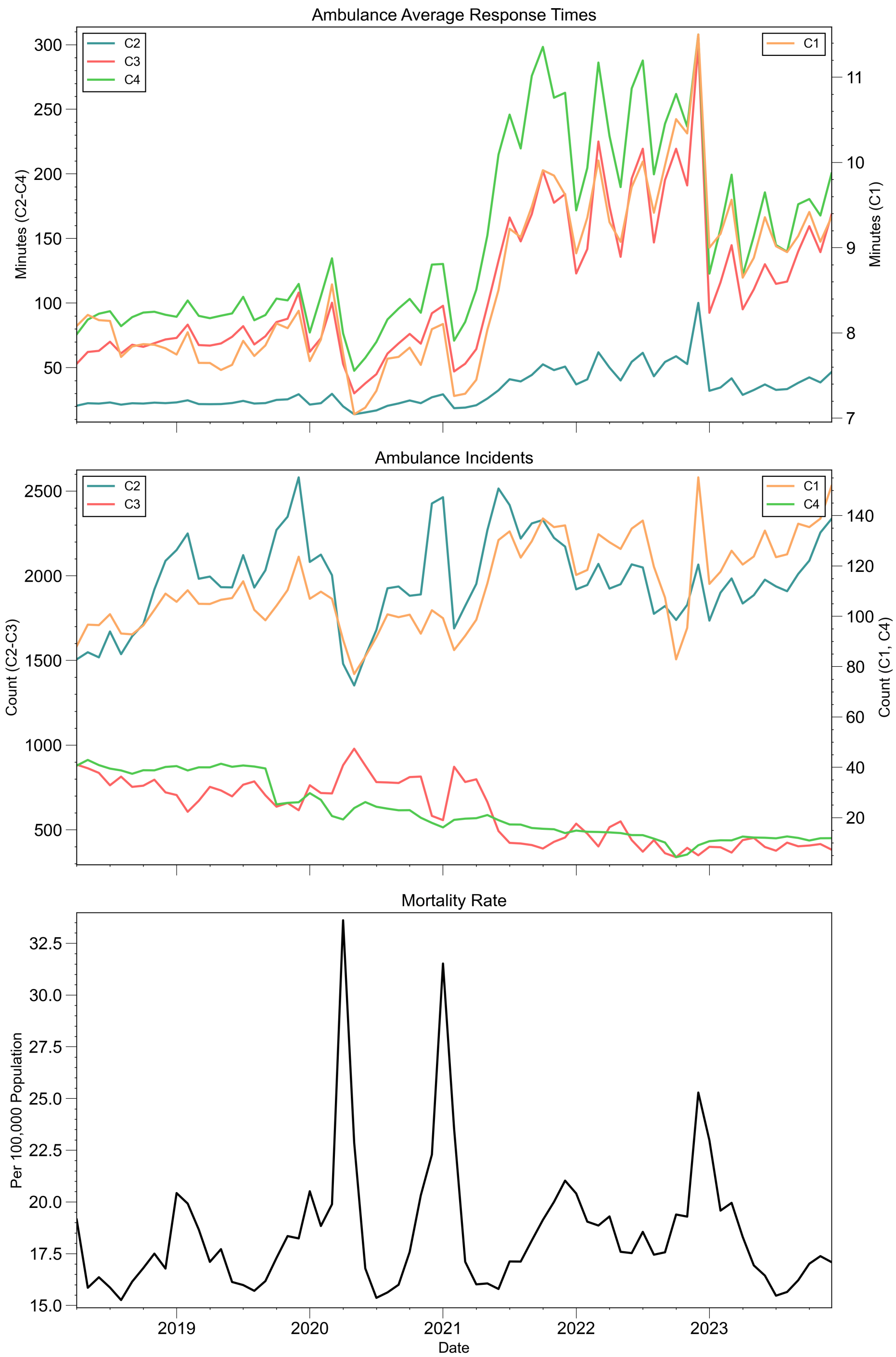
2.2.2 Industrial action in 2022 - 2023

The sustained pressure from the pandemic culminated in the largest ambulance strike in over 30 years during late 2022 and early 2023. Ambulance workers across England and Wales participated in coordinated industrial action involving all three major unions (GMB, UNISON, and Unite). The strikes began on December 21, 2022, and intensified through January 2023, with the largest coordinated action occurring on January 23, 2023.

The industrial action was triggered by the government's pay offer, which unions argued represented a significant real-terms pay cut given high inflation rates, but reflected deeper systemic issues including deteriorating workloads and rising vacancy rates in ambulance services. The disputes were eventually resolved in March 2023 following government negotiations that resulted in improved pay settlements, though this led to new minimum service level legislation restricting future strike action in emergency services.

It is noteworthy that the East of England Ambulance Service NHS Trust (EEAST) was exempt

Figure 2: Monthly Trends of Ambulance Data and Mortality Rate



from the December 2022 strikes. UNISON's initial ballot for ambulance workers at EEAST failed to meet the 50% turnout threshold required for legal strike action in November 2022 (UNISON, 2023). This regional variation in strike participation provides useful contextual variation across trusts and will be revisited in our empirical analysis, where we examine local differences in response times and health outcomes during this period.

2.3 Government Interventions During the 2022-2023 Winter Crisis

In response to the severe winter pressures on the emergency care system and hospitals in 2022-2023, the UK government introduced a series of policy interventions aimed at alleviating the hospital discharge bottleneck. The first of these was the £500 million Adult Social Care Discharge Fund, announced in September 2022, with detailed allocation guidelines released in mid-November. In early December 2022, the first tranche of the fund (£200 million, 40% of the total) was distributed to local authorities and Integrated Care Boards (ICBs). The remaining £300 million was scheduled for release on 31 January 2023, conditional on meeting specific requirements, including submission of spending plans by 16 December 2022 and subsequent fortnightly activity reports documenting delivery of the planned interventions (Department of Health and Social Care, 2023b).

Spending under this scheme was primarily directed towards workforce support and service capacity expansion. On the workforce side, local areas used the funds to purchase additional staff hours and to increase minimum wages in an attempt to attract and retain workers, given the high turnover rates in the social care sector. On the service capacity side, authorities increased capacity by block-purchasing care home beds and buying more home care packages. These measures aimed to accelerate patient discharge from hospitals, thereby freeing up capacity for incoming patients during a period of heightened seasonal demand.

The logic of this intervention rests on the close interdependence between hospital bed availability, ambulance handover time, and response times. High hospital bed occupancy reduces hospitals' ability to admit new patients, leading to longer ambulance handover delays. As ambulances remain occupied waiting to transfer patients, both vehicles and crews are prevented from returning to service, reducing the system's effective capacity. This dynamic directly contributes to longer ambulance response times for patients waiting in the community, thereby raising mortality risks for time-sensitive conditions. Moreover, prolonged handover delays result in inefficient use of ambulance staff time, increasing stress and sickness-related absences, which further exacerbate capacity shortages. The urgent funding was therefore designed to ease these systemic bottlenecks and prevent further deterioration in service performance.

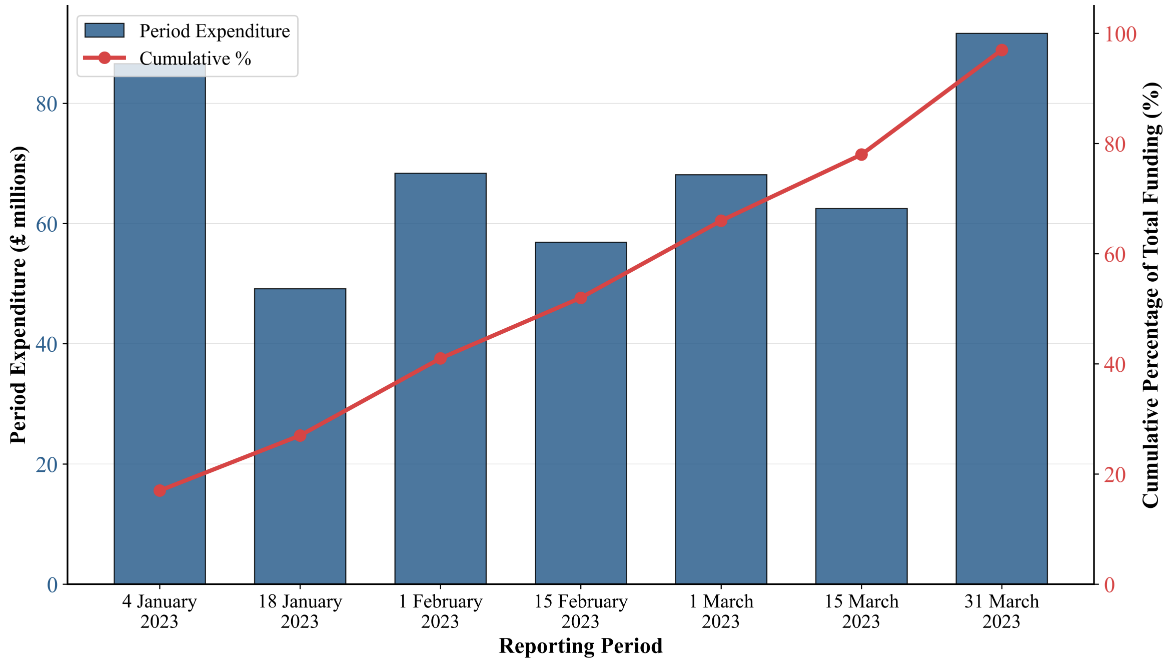
A key challenge, however, was the short notice and compressed implementation timeline. After receiving detailed allocations in November, local authorities were required to design spending plans and implement interventions within a matter of weeks. Immediate implementation was often infeasible due to several constraints. For instance, providers of home care packages may have responded to sudden increases in demand by raising prices, limiting the effective purchasing power of the fund. In addition, the administrative burden of planning, reporting, and compliance may have diverted attention and resources from ongoing winter plans. As a result, even where funding was formally allocated, it was unlikely to translate rapidly into observable improvements in system performance.

Figure 3 illustrates the distribution of spending under the Adult Social Care Discharge Fund, covering the reporting period from 23 December 2022 to 31 March 2023. Spending was highest at the beginning and end of the period, consistent with the compressed timelines and final reporting deadlines. In total, £483 million was spent to support the health and social care system in managing the winter crisis.

The second major intervention was announced on 9 January 2023, when Health Secretary Steve Barclay unveiled an additional £200 million emergency funding package, which specifically targeted at further reducing hospital discharge delays. This scheme operated as a reimbursement-based programme administered through ICBs, which were required to immediately secure bedded step-down capacity in care homes and other community settings. This direct funding mechanism allowed ICBs to respond flexibly to local demand, complementing the earlier £500 million discharge fund. In addition, £50 million in capital funding was allocated to create new ambulance hubs and expand hospital discharge lounges, addressing physical infrastructure constraints that contributed to handover delays.

Without these interventions, the health system risked entering a vicious cycle: overcrowded hospitals would lead to longer ambulance handovers, delayed treatments, and worsening patient conditions, further increasing healthcare service demand. The combined policy response therefore represented a quite sharp and temporally isolated intervention aimed at mitigating systemic strain, improving ambulance response times, and ultimately enhancing patient outcomes at the local level. We will discuss the effects of those policies on improving the emergency care system performance over this period in the following sections.

Figure 3: Temporal Distribution of Adult Social Care Discharge Fund Disbursements



2.4 Sharp Improvement in Ambulance Response Times

The policy interventions described above led to a marked improvement in ambulance response times across all categories from January 2023 onward. Figure 2 presents the temporal patterns of ambulance response times over the study period. Following the first week of 2023, all categories experienced substantial reductions, representing an sharp temporal improvement in emergency service performance. This shift coincides with the implementation of the £500 million Adult Social Care Discharge Fund, and the subsequent £200 million emergency fund reinforced these gains, helping maintain lower response times from the second week of 2023 onward.

Figure 4 provides visual evidence of this systematic improvement across all emergency categories. The figure displays regression discontinuity plots for each category (C1–C4) around the cutoff point at the beginning of 2023. The vertical dashed line at week 0 marks the first week of 2023. A clear discontinuous drop is visible across all categories immediately after this threshold. Prior to the cutoff, response times exhibited gradually increasing trends, particularly during the COVID-19 period, while the post-cutoff period shows sustained lower response times across all categories—closely aligning with the timing of the government's hospital discharge funding.

The consistency of this pattern across call categories is particularly noteworthy. For life-threatening emergencies (Category 1 calls), the weekly average ambulance response time fell from approximately 11 minutes to around 9 minutes. For emergency but non-life-threatening incidents (Category 2), response times declined sharply—from roughly 70 minutes to about 40 minutes. Even for lower-priority cases (Categories 3 and 4), which typically face the longest delays, average response times dropped substantially—from pre-cutoff levels of roughly 250–300 minutes to around 150 minutes after the intervention.

The magnitude and uniformity of these improvements across categories suggest a system-wide operational shift rather than category-specific interventions. This interpretation is consistent with the structural relief provided by increased discharge capacity and workforce resilience funded by the national policy measures.

This sharp temporal discontinuity provides a basis for implementing a Regression Discontinuity in Time (RDiT) design to estimate the causal effect of reduced ambulance response times on local health outcomes. The cutoff—the first week of 2023—represents a policy-induced sharp change in response times rather than a demand-driven shift. Patients experiencing emergencies cannot strategically time the onset of their conditions, especially for acute or life-threatening events. Therefore, there is no manipulation around the cutoff on the demand side, satisfying a key identification assumption for the RDiT framework.

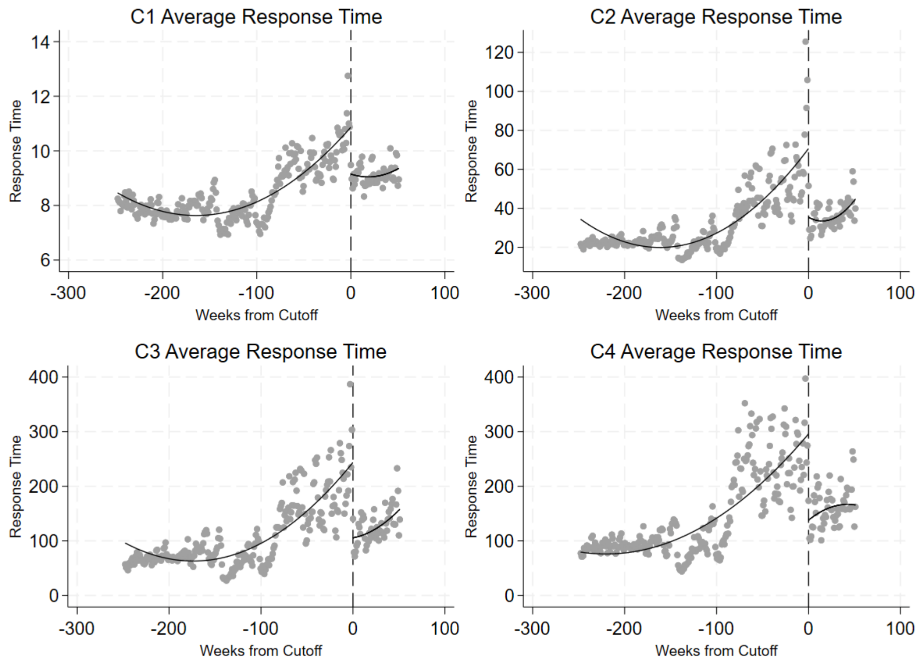
In the following sections, we describe our data and empirical design, and present results quantifying the impact of this policy-induced improvement in several health outcomes.

3 Data

We assemble a set of novel administrative and official datasets to examine how improvements in the performance of the emergency care system affected patient health outcomes in England. The analysis combines multiple sources that together capture both operational measures of system performance (ambulance response times and hospital handover delays) and clinical outcomes, including the Return of Spontaneous Circulation (ROSC) and 30-day survival rates for cardiac arrest patients, as well as local mortality rates.

A key contribution of this study is the construction of a unique ambulance performance dataset obtained through a series of Freedom of Information (FOI) requests to ten UK ambulance trusts. These data provide postcode district-by-week level information on incident volumes and average response times, allowing unprecedented local and temporal coverage across England from 2018 to 2023. We complement these data with national NHS England administrative records on ambulance handover delays, cardiac arrest outcomes, and hospital capacity, and merge them with local authority level mortality statistics from the Office for National Statistics (ONS). Together, these

Figure 4: The Discontinuity of Ambulance Response Times



sources form a multi-dimensional panel that enables us to trace the effects of emergency care system performance improvements from reduced bottlenecks in ambulance operations to downstream health outcomes at both the patient and population level.

3.1 Emergency Care System Performance

3.1.1 Ambulance response time data

The local ambulance service performance is measured by the average ambulance response time with the number of incidents per week for each local authority. The ambulance response time is defined as the time elapsed from receiving a 999 call to the arrival of an ambulance vehicle at the patient's location. The average response time for all ambulance incidents in a local authority within a week is the key variable. These data come from Freedom of Information requests we made to each ambulance trust in the UK, including 10 ambulance trusts covering the major England, Scotland, and Wales[1]. The Northern Ireland area is excluded from the analysis. Data from

[1] Those 10 ambulance trusts are: North East Ambulance Service NHS Foundation Trust; Yorkshire Ambulance Service NHS Trust; North West Ambulance Service NHS Trust; West Midlands Ambulance Service University NHS Foundation Trust; East Midlands Ambulance Service NHS Trust; South East Coast Ambulance Service NHS Foundation Trust; London Ambulance Service NHS Trust; East of England Ambulance

the South Western Ambulance Trust and the South Central Ambulance Trust are not able to be acquired. We use the data from 8 trusts in England in the following analysis, which are with consistent policy context.

The requested data provide postcode district-by-week granularity, which aggregates all incident-level information from a certain postcode district in a week. To match the granularity of mortality data, which is local authority by week level, we merge those data into local authority level, where the number of ambulance incidents are summed together for all postcode districts data in a certain local authority, and the average ambulance response times are averaged with the number of incidents as the weights. Finally, we get a balanced panel dataset covering the time period from April 2018 to the end of 2023.

3.1.2 Hospital handover delay data

As the another part of emergency care system performance, we collect the hospital handover data from the NHS England Urgent and Emergency Care Daily Situation Reports dataset. This dataset provides daily data by NHS Foundation Trusts in England, including number of patients arriving by ambulance, number of ambulance handover delays between 30 and 60 minutes, number of ambulance handover delays greater than 60 minutes, and the total time lost to ambulance handover delays.

Although this dataset provides fine daily granularity, it is reported only for the winter months (from November to April) in each year. In our analysis, we use data from 14 November 2022 to 2 April 2023, which fully covers the period of the policy intervention.

3.2 Health Outcomes

3.2.1 Cardiac arrest outcomes

To examine the direct and clinically relevant health outcomes caused by the emergency care system, we utilize the NHS England Ambulance Quality Indicators dataset. This dataset provides two key health outcomes for cardiac arrest patients.

First, we use the *Return of Spontaneous Circulation* (ROSC) rate in our analysis, which represents the proportion of cardiac patients who had their pulse back before arriving at the hospital.

Service NHS Trust; The Welsh Ambulance Services NHS Trust which covers the whole of Wales; The Scottish Ambulance Service which covers the whole of Scotland.

This indicator measures the success rate of ambulance crew's resuscitation efforts, it is also the indicator of the pre-hospital emergency care system performance. When ambulance response times are shorter, patients receive life-saving interventions sooner, increasing the likelihood of achieving ROSC instead of being directly dead.

Second, we utilize *Survival Rate at 30 days* following cardiac arrest for ambulance patients in the analysis to examine whether more patients could survive from the cardiac arrest. Achieving ROSC is only the initial step in recovery, it does not guarantee that the patient could definitely survive from the cardiac arrest since the underlying reasons caused the cardiac arrest (e.g., blood clot, or severe trauma) still exists and needs to be fixed. Furthermore, patients often face post-cardiac arrest syndrome after the cardiac arrest shock, in which the brain, heart, and other organs have been starved of oxygen for several minutes or even longer time, causing potential inflammation or organ dysfunction syndrome. Thus, we need to check the exact survival rate of those patients at 30-day after the cardiac arrest, which is determined by the full emergency care system performance, not only the pre-hospital care but also how fast the patient could be treated by the hospital care.

These data are reported monthly at the ambulance trust level from January 2021 to February 2025. We focus on cardiac arrest episodes as they constitute medical emergencies where individuals cannot predict or control the timing of their cardiac events. The random nature of cardiac arrest onset provides a clear background for our identification strategy, which indirectly strengthens the identifying assumption that patients can not manipulate their cardiac arrest timing around the time cutoff. When we see ambulance response times decrease during specific periods, we would expect to observe corresponding improvements in the ROSC and survival rates among cardiac arrest patients receiving emergency services.

For each outcome, we utilize two measures of cardiac arrest patient outcome: (1) the conventional group for all cardiac arrest cases, and (2) the Utstein patient group. The Utstein criteria identify a specific subset of cardiac arrest patients with witnessed arrests, initial shockable rhythms, and other favourable prognostic factors, providing a more homogeneous and credible patient population for comparable analysis even across different context or studies. Both measures offer complementary perspectives on the relationship between emergency response performance and patient outcomes.

3.2.2 Mortality data

We obtained the weekly mortality data by local authority level from the Office for National Statistics (2023) over the time period of 2018-2023, which covers all regions for which we have ambulance service performance data. Mortality data are categorized by location (care home, home, hospice, hospital, other communal establishment, elsewhere) and cause of death (all causes or COVID-19). These mortality data are converted into mortality rates adjusted by local population, measured as deaths per 100,000 population. These data allow us to assess population-level mortality trends in relation to system-wide improvements in emergency care performance.

3.3 Additional Check: Hospital Capacity and Seasonal Pressure Data

To address the potential concerns of confounders which might challenge the validity of our identification, we use the hospital flu bed capacity data which is also obtained from the NHS England Urgent and Emergency Care Daily Situation Reports dataset. The number of beds are reported daily by different NHS Foundation trusts, classified by General & Acute beds and Critical Care beds. This daily flu beds data precisely documents the winter flu pandemic trends in the local healthcare provider level.

3.4 Summary Statistics

Table 2 reports descriptive statistics for the key variables used in the analysis, covering ambulance performance, handover delays, health outcomes, mortality rates, and hospital capacity indicators. Together, these panels provide a comprehensive picture of the structure and variation in emergency care system performance and its potential downstream effects on health outcomes.

Panel A shows substantial variation in ambulance response times across emergency categories. Consistent with the national response targets, Category 1 (C1) calls (the most life-threatening emergencies) have the shortest mean response time of 8.54 minutes, while less urgent calls experience progressively longer delays, reaching an average of 140.32 minutes for C4 cases. The wide standard deviations, especially for lower-priority categories, indicate large temporal and regional heterogeneity in system pressure. Incident volumes display similar dispersion. C2 calls dominate the service workload, with a mean of around 1,970 incidents per week, but substantial variation (standard deviation of 5,416) highlights strong clustering in demand across local authorities and time periods. Panel B summarizes hospital handover delays, a key component of emergency care system congestion. On average, 11.6 patients per NHS Foundation Trust per day experience 30–60 minute delays, while about 9.8 face handovers exceeding one hour. The corresponding time lost to handover delays averages nearly 23 hours per day, with large variation (standard deviation 45.6), underscoring the severity of bottlenecks at the hospital interface even before the 2023 intervention.

Panel C presents cardiac arrest outcomes derived from the NHS England Ambulance Quality Indicators. The mean ROSC rate is 26%, increasing to 48% among Utstein patients—those with witnessed arrests and early CPR. Thirty-day survival averages 9% for all cardiac arrests and 27% for Utstein cases. Panel D reports mortality rates by place of death. Hospital deaths are the largest category (mean of 8 deaths per 100,000 population), followed by home (5.0) and care home deaths (4.1). The total death rate averages 18.5 per 100,000 but varies considerably across time and regions (standard deviation 6.7). COVID-19 deaths have a mean of 1.2 but display the greatest volatility, consistent with pandemic-related spikes.

Panel E summarizes hospital flu bed capacity data used to control for seasonal congestion. General & Acute flu-designated beds average 10.5 per NHS Foundation Trust per day, while Critical Care flu beds average less than one. The large standard deviations highlight the uneven distribution of flu-related pressures across hospitals.

Overall, these summary statistics illustrate a highly heterogeneous emergency care system, characterized by pronounced temporal and spatial variation in both operational performance and health outcomes. The broad dispersion in ambulance response times, handover delays, and mortality underscores the scope for identifying causal effects of system improvements on patient and population health outcomes.

3.5 Trends of Emergency Response and Mortality from 2018 to 2023

Figure 2 presents a series of related graphs about emergency healthcare system performance and its potential relationships with public health outcomes in the UK. All data are merged to the monthly level and averaged at the national level [2].

The top graph shows the ambulance average response times with different categories (C1-C4) over 2018-2023, where C1 is measured on the right axis (7-11 minutes range) , while C2-C4 use the left axis (20-300 minutes range) showing much longer response times compared to C1. We can observe that all kinds of response times show a significant increase starting around 2021, with particularly dramatic increases for C3 and C4 categories. C1, representing the most critical emergencies, remained more stable but still showed similar increasing trend. Then starting from 2023, all emergency categories fell back from the peak, while still not returning to pre-COVID level.

[2] Weekly time series are provided in the Appendix Figure A.1

Table 2: Summary Statistics

Variable	Count	Mean	Std. Dev.	Min	Median	Max
Panel A: Ambulance Response Times (Minutes) and Incidents						
C1 Response Time	75,700	8.54	2.85	0.37	7.95	140.53
C2 Response Time	76,942	32.52	21.08	0.33	27.15	439.37
C3 Response Time	76,618	109.63	78.59	5.60	88.83	1408.20
C4 Response Time	61,537	140.32	144.99	0.07	96.73	3083.02
C1 Incidents	79,411	112.11	182.70	0.00	48.00	3,775.00
C2 Incidents	79,411	1,973.10	5,416.77	0.00	279.00	89,445.00
C3 Incidents	79,411	605.80	1,726.37	0.00	108.00	38,625.00
C4 Incidents	79,411	22.63	52.04	0.00	5.00	621.00
Panel B: Handover Delays						
Handover delay 30-60 mins	20,160	11.59	11.42	0	9	86
Handover delay > 60 mins	20,160	9.78	14.50	0	2	102
Time lost (hours)	20,160	22.92	45.58	0	4.16	575.56
Panel C: Health Outcomes of Cardiac Arrest Patients						
ROSC rate	544	0.26	0.06	0	0.27	0.6
ROSC rate (Utstein)	524	0.48	0.13	0	0.49	1
Survival rate	544	0.09	0.03	0	0.09	0.4
Survival rate (Utstein)	527	0.27	0.12	0	0.27	1
Panel D: Mortality Rates (Per 100,000 Population)						
Care Home Deaths Rate	78,877	4.11	2.73	0.00	3.66	38.40
Home Deaths Rate	78,877	4.99	2.47	0.00	4.67	22.39
Hospital Deaths Rate	78,877	8.00	3.53	0.00	7.55	48.00
Hospice Deaths Rate	78,877	0.93	0.92	0.00	0.78	7.92
COVID-19 Deaths Rate	78,877	1.20	3.00	0.00	0.00	58.67
Total Deaths Rate	78,877	18.45	6.68	1.25	17.83	87.33
Panel E: Flu Beds						
General & Acute Flu Beds	19,177	10.49	23.52	0	2	422
Critical Care Flu Beds	19,177	0.78	2.53	0	0	57

The middle graph displays the ambulance incidents count data for different categories. Again using the dual axis, this graph shows the C1 and C2 representing the most serious emergencies maintained relatively parallel trends. This is because both categories involve patients with similar severe conditions, but C1 has specific symptoms (cardiac arrest) requiring immediate treatment. A patient with very serious conditions may be classified into either C1 or C2, but not C3 and C4. C3 and C4 incidents declined notably after 2021, this shift is potentially driven by the impact of COVID-19, which either escalated what would have otherwise been C3 and C4 cases into more critical C1 and C2 conditions or effectively suppressed lower-priority incidents due to the overwhelming demand for ambulance services. With limited emergency resources, the surge in C1 and C2 incidents may have displaced C3 and C4 incidents, reducing their reported numbers.

The last graph tracks national level mortality rate per 100,000 population, showing several significant spikes. The largest spikes appear in early 2020 and early 2021, with another smaller spike in early 2023. These mortality peaks align temporally with known COVID-19 waves.

3.6 Geographical Distribution of Emergency Response and Mortality

We mapped the ambulance and mortality data to provide a comprehensive spatial analysis of emergency healthcare metrics across England for three pivotal years (2019, 2022, 2023). We particularly present those three years as providing a cross-time comparison of before COVID (2019), COVID time (2022), and post-COVID (2023). These visualizations allow us to examine both the geographical patterns and evolution of ambulance service indicators and health outcomes at two complementary administrative levels.

Figure A.2 illustrates the spatial distribution at the postcode district level, the original granularity of our dataset, offering a detailed perspective on variations across smaller geographic units. The left column represents the logarithmic count of C1 incidents, while the right column depicts the C1 average response time. Both variables exhibit substantial spatial disparities, which persist across different years. Despite the significant disruptions caused by COVID-19, the geographical heterogeneity in ambulance service provision remains evident over time.

Figure A.3 presents these metrics aggregated to local authority level, with the addition of mortality rates that are available at this administrative level. This administrative level aligns with governance and policy implementation boundaries, making it especially relevant for analysing institutional response and resource allocation decisions. From left to right column, it shows logarithmic count of C1 incidents, C1 average response time, and local mortality rate across different years. As with the postcode district maps, significant geographical heterogeneity persists even at this more aggregated administrative level.

This multi-scale spatial analysis complements the earlier time-series findings by revealing not only when but precisely where healthcare system pressures manifested most severely. The consistent geographic patterns with the dramatic spatial shifts in response times and mortality rate suggests that some structural factors (resource allocation decisions, funding models, or labour market constraints) have contributed to geographically uneven impacts on emergency healthcare delivery and population health outcomes.

4 Method

We study the short-run effects of the 2022–2023 emergency funding on (1) emergency care system performance and (2) downstream patient outcomes using a Regression Discontinuity in Time (RDiT) design centred at the beginning of 2023. The running variable is calendar time t, re-indexed so that $t = 0$ corresponds to the first week (or month) of 2023, and the treatment indicator $\mathbb{1}(t \geq 0)$ switches on mechanically at that cutoff. Intuitively, we compare observations just before and just after the cutoff within a local window, flexibly controlling for smooth time trends.

4.1 Discussion of the Policy Cutoff

In a Regression Discontinuity in Time (RDiT) framework, the running variable is calendar time rather than a cross-sectional score. Identification relies on locating a precise policy cutoff where treatment intensity shifts sharply, such that observations just before and after the threshold are comparable except for treatment exposure. Establishing this cutoff correctly is crucial, since in time-based settings multiple institutional or seasonal events may coincide.

In our context, the NHS *Hospital Discharge Fund* was announced and disbursed to local systems in early December 2022. At first glance, one might consider December as the natural treatment onset. However, several institutional and operational factors make December an unsuitable cutoff, while January 2023 provides a cleaner, empirically validated threshold.

Why December 2022 is not an effective cutoff. Although funding allocations were confirmed in early December, converting allocations into operational discharge capacity required time. The funds were primarily used to procure step-down care beds, arrange home-care packages, and expand discharge coordination teams—activities that are typically contracted or rostered on weekly or monthly cycles. Implementation was further delayed by the Christmas and New Year period, when NHS services operate on reduced capacity and local authorities have limited procurement activity. Consequently, December represents a transitional period in which spending plans were being finalized, but new discharge capacity had not yet been mobilized uniformly across systems. Using December as a cutoff would therefore mix treated and untreated observations, attenuating estimated effects and violating the sharp-treatment assumption underlying RDiT.

Why January 2023 represents the true implementation cutoff. By the beginning of January, systems began to deploy the new discharge capacity at scale. Administrative data in Figure 3 indicate that by the end of the first reporting period (4 January 2023), roughly £90 million—about 18% of the total fund—had already been spent. This pattern suggests that implementation accelerated rapidly in the first full week of January. Consistent with this, our first-stage analysis in Section 2.4 and 5.1 shows a clear and sustained improvement in ambulance response times beginning precisely in week 1 of 2023, with no comparable discontinuity in December. This aligns with the interpretation that early-December fund disbursement translated into operational improvements only after a short implementation lag of roughly one month. Because RDiT identifies local average treatment effects at the point of discontinuity, it is appropriate to set the cutoff at the time when the treatment turns on at scale and performance indicators respond sharply—namely, the start of 2023.

Granularity and alignment across outcomes. To maintain consistent temporal alignment across datasets of varying frequency, we define equivalent cutoffs for each data source: *1 January 2023* for daily measures (handover delays and influenza bed occupancy), the *first week of 2023* for weekly outcomes (ambulance response times and local mortality), and *January 2023* for monthly outcomes (ROSC and 30-day survival rates). These correspond to the same effective implementation threshold expressed at different temporal resolutions.

4.2 RDiT Setup and Estimating Equation

We estimate the effects of the 2022–2023 emergency discharge funding using a local-polynomial Regression Discontinuity in Time (RDiT) design centred at the beginning of 2023 (Calonico *et al.*, 2014; Cattaneo *et al.*, 2024; Hausman and Rapson, 2018). The running variable is calendar time t, normalised so that $t = 0$ corresponds to the cutoff (the first week or month of 2023, depending on data frequency). For any outcome Y_{it}—measured for unit i (e.g. ambulance trust or region) at time t—the RDiT specification is:

$$Y_{it} = \alpha_i + \tau \mathbb{1}(t \geq 0) + f(t) + \varepsilon_{it}, \tag{1}$$

where α_i captures unit-specific fixed effects or, equivalently, we use demeaned series[3], $\mathbb{1}(t \geq 0)$ is an indicator for the post-cutoff period, and $f(t)$ is a flexible, piecewise polynomial in t estimated locally on each side of the cutoff using a triangular kernel. The coefficient τ identifies the average discontinuous change in Y_{it} at the cutoff.

Equation (1) is applied consistently across datasets of different temporal resolutions—weekly (ambulance response times), monthly (ROSC and 30-day survival rates), and daily (handover delays or influenza bed occupancy)—with each series normalised to its respective time index. Standard errors are clustered at the ambulance-trust or local-authority level, depending on the aggregation. We use mean squared error–optimal (MSE) bandwidths and report robust bias-corrected confidence intervals.

Following best practice, the main specifications employ a local linear polynomial ($p = 1$), and we report sensitivity analyses varying both the bandwidth h and the polynomial order $p \in \{1, 2, 3, 4\}$ to ensure robustness of the estimated discontinuity τ.

4.3 Identifying Assumptions and Diagnostic Checks

The RDiT estimand identifies the causal effect of the funding shock under standard RD continuity conditions applied to time:

1. **Continuity of counterfactual trends.** In the absence of the funding shock, the conditional expectation of each outcome is continuous in t at $t = 0$. Practically, nearby weeks/months trace a smooth time path with no structural break at the cutoff apart from the treatment.

2. **No precise manipulation of the running variable.** Agents cannot sort around the cutoff in calendar time. In our context, patients cannot time life-threatening emergencies; providers did not pre-announce the exact timing of implementation in a way that allows precise manipulation at the week/month boundary.

3. **No coincident discontinuities in confounders.** Other determinants of outcomes should not jump at $t = 0$. We assess this with (a) seasonal placebo tests using the January 2022 cutoff, (b) continuity of influenza pressure (G&A and CC beds with lab-confirmed influenza), and (c) estimates on a non-strike subsample (East of England).

[3]Because outcomes are measured repeatedly for each ambulance trust or local area, we remove unit-specific mean differences to isolate within-unit variation over time. In practice, this is equivalent to including unit fixed effects (α_i) in equation (1). Demeaning the series prior to estimation yields identical results and simplifies the implementation of local regressions around the cutoff.

We perform several validity checks to assess whether the observed discontinuities can be credibly attributed to the funding intervention rather than to seasonal or contemporaneous confounders. First, we estimate placebo RDiT models using the same specification but centred at January 2022, one year earlier. The absence of a comparable 2022 discontinuity would support the interpretation that the 2023 break reflects the funding shock rather than recurring winter patterns. Second, we examine the continuity of influenza-related hospital bed occupancy around the cutoff, since seasonal influenza is a major potential confounder for emergency care demand. A smooth evolution in flu pressures through the threshold would indicate that our effects are not mechanically driven by changes in background health risk. Finally, we replicate the analysis for the East of England Ambulance Service, the only region unaffected by industrial strike action during this period, to verify that estimated discontinuities are not confounded by contemporaneous labour disruptions.

4.4 Bandwidth and Polynomial Robustness

Local RD estimators trade bias and variance via the bandwidth h and polynomial order p. Narrow h reduces bias (tighter local comparison) at the cost of precision; wider h increases precision but risks bias from curvature in trends. Likewise, higher-order polynomials can overfit, whereas local linear is typically preferred for RD near the threshold.

We therefore: (1) report main estimates with the MSE-optimal bandwidth and $p = 1$; (2) vary h over a symmetric grid around the optimum to assess stability of sign and magnitude; and (3) vary $p \in \{1,2,3,4\}$ at the optimal h. We expect (and find) that estimates remain positive and statistically significant across reasonable h and p, with larger magnitudes at very small h (capturing the sharpest local change) and with higher-order polynomials occasionally producing larger point estimates but similar qualitative conclusions. We retain the local linear specification for parsimony and to avoid overfitting, reporting the richer set for transparency.

4.5 Donut RDiT

Implementation frictions or transient disturbances exactly at the threshold can bias RD estimates if very near-cutoff observations are atypical, i.e., the visual discontinuity is mainly generated by outliers around the cutoff and it will disappear if we remove the points around the cutoff. To address this, we implement a "donut" RDiT that drops observations in a symmetric window around the cutoff and re-estimates the model on the remaining data (Barreca *et al.*, 2011; Noack and Rothe, 2023). Given our relatively narrow optimal windows, we remove ±1 month around $t = 0$ for monthly outcomes (and the analogous ± window for weekly performance when reported). If the discontinuity is genuine, estimates should persist with similar sign and slightly wider confidence intervals due to fewer observations. We expect to see that the discontinuity still exists after we impose the Donut RD design.

5 Results

5.1 Validation of the Discontinuity in Response Times

As documented in Section 2.4, ambulance response times across all emergency categories experienced a sharp and sustained improvement beginning in the first week of 2023. This section formally tests for a discontinuity at that cutoff to validate the key identifying assumption for our RDiT design.

Table 3 reports the estimated discontinuities. We find statistically significant reductions in response times across all categories. C1 calls, representing life-threatening emergencies, show an average drop of approximately 2.5 minutes ($p < 0.001$), while C2 calls decline by around 94 minutes ($p < 0.001$). C3 and C4 categories also exhibit sharp declines significantly. This evidence validates the use of the first week of 2023 as the cutoff for our RDiT analysis and supports interpreting subsequent changes in patient outcomes as responses to an exogenous shift in emergency service performance.

Table 3: Estimated Discontinuities in Ambulance Response Times: Detailed Results

Category	Estimated Jump (τ)	Std. Error	p-value	Bandwidth (weeks)
C1 Response Time (minutes)	-2.49^{***}	(0.23)	< 0.001	35.70
C2 Response Time (minutes)	-94.13^{***}	(5.09)	< 0.001	17.16
C3 Response Time (minutes)	-233.90^{***}	(10.19)	< 0.001	20.14
C4 Response Time (minutes)	-184.62^{***}	(16.24)	< 0.001	27.36

Notes: RDD estimates using local polynomial regression (order 2) with triangular kernel. Bandwidth selected using MSE-optimal procedure. Standard errors clustered at local authority level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Identification Check: Winter Influenza Pressure

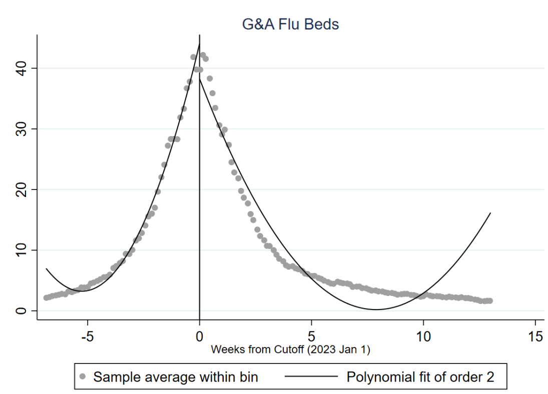
A key identifying assumption in our design is that the discontinuity observed in ambulance response times at the start of 2023 was not driven by concurrent changes in other health system pressures that could independently affect patient outcomes. One major confounding concern is the intensity of the winter influenza season. Influenza outbreaks typically peak in December and

January and are associated with substantial seasonal fluctuations in hospital demand and mortality. If influenza prevalence dropped sharply at the same time as the improvement in ambulance response times, our estimated effects on patient outcomes could in part reflect a shift in background health risks rather than the causal impact of emergency service performance.

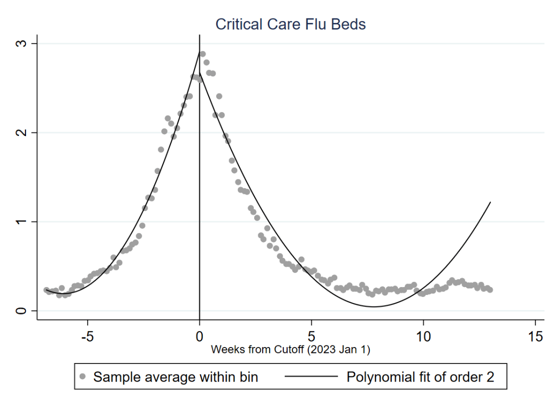
To examine this, we use daily data from NHS England on hospital bed occupancy by patients with laboratory-confirmed influenza. We distinguish between General and Acute (G&A) beds, which capture the broader inpatient caseload, and Critical Care (CC) beds, which represent the most severe flu cases requiring intensive treatment. Figures 5a and 5b present regression discontinuity plots of these two measures around the 1st January 2023 cutoff.

Both figures show a clear seasonal decline in influenza bed occupancy after the late-December peak, consistent with the typical winter pattern. However, this decline occurs smoothly over time, without any significantly visible break or discontinuity at the cutoff. In both G&A and critical care settings, the number of flu-related admissions decreases gradually, indicating that the underlying influenza pressure on the healthcare system evolved continuously through this period. This pattern suggests that the January 2023 discontinuity in ambulance response times and the subsequent changes in health outcomes was not mechanically driven by a sudden alleviation of seasonal flu pressures.

Figure 5: Regression discontinuity plots of beds occupied by patients with laboratory-confirmed influenza



(a) General and Acute (G&A) Beds



(b) Critical Care (CC) Beds

5.3 Estimated Effects on Health Outcomes

This section examines whether the sharp improvement in ambulance response times identified in Section 5.1 translated into measurable changes in patient health outcomes. We first analyse a direct pre-hospital outcome, *Return of Spontaneous Circulation* (ROSC) *rate* on Arrival at Hospital for cardiac arrest patients, which measures the system-wide returns to the shorter ambulance response times. The ROSC rate reflects the effectiveness of pre-hospital intervention, particularly the timeliness of cardiopulmonary resuscitation (CPR) and defibrillation, and therefore serves as a physiological indicator of whether faster response directly improved the chances of initial survival (Neumar *et al.*, 2015; Nolan *et al.*, 2010).

We then turn to *30-day survival rate*, a broader measure that captures both pre-hospital and in-hospital components of emergency care. Faster ambulance response allows patients to receive life-saving treatment earlier at the scene, while reduced handover delays enable quicker transfer to hospital care. Together, these mechanisms reflect how the overall performance of the emergency care system translates into sustained survival improvements beyond the initial incident.

Finally, we examine local *mortality rate*, a population-level measure of health outcomes that extends beyond cardiac arrest cases. Although mortality rate is inherently more diffuse which reflects deaths from both ambulance-related and unrelated causes, it provides a useful benchmark for assessing whether system-wide performance improvements produced broader health benefits across local areas.

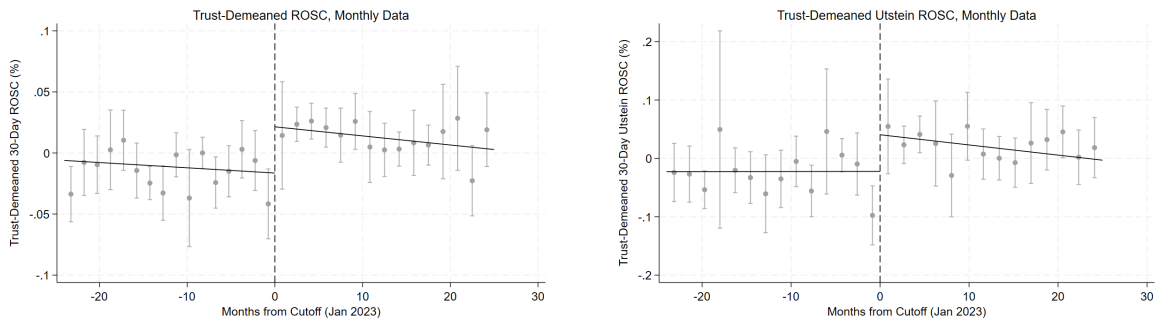
For each outcome, we present visual evidence of the discontinuity around the January 2023 cutoff, followed by the estimated treatment effects from our RDiT models. We also report a series of robustness checks, including placebo cutoffs, donut specifications, and bandwidth sensitivity analyses, to ensure that our findings are not driven by specification choices or short-term confounders.

5.3.1 Immediate pre-hospital outcome: return of spontaneous circulation rate on arrival at hospital

We begin our analysis of health outcomes with the Return of Spontaneous Circulation (ROSC) rate, a pre-hospital indicator of successful resuscitation following cardiac arrest. It stands for the moment the ambulance crew's resuscitation treatment (e.g., CPR or defibrillation) is successful, and the patient's heart, which had stopped, starts beating on its own again. This ROSC rate measures the proportion of patients who regained a palpable pulse upon arrival at hospital among those for whom resuscitation was attempted. This outcome directly reflects the timeliness and effectiveness of pre-hospital emergency response, especially for cardiac arrests, where the probability of survival declines sharply with each minute of delay in initiating CPR or defibrillation.

Figure 6a and Figure 6b present regression discontinuity plots of the ROSC rate and the Utstein ROSC rate.[4] In both panels, there is a clear upward shift at the cutoff corresponding to the beginning of 2023. This discontinuity coincides precisely with the improvement in ambulance response times documented in Section 5.1, this aligns with that in theory, faster ambulance response could translate into higher rates of successful resuscitation at the scene.

Figure 6: Regression Discontinuity in Time: ROSC Rate



(a) ROSC Rate

(b) ROSC Rate for Utstein Patients

The RDiT estimates reported in Table 4 confirm the visual evidence. Across trusts, the ROSC rate increased by approximately 4.23% (p<0.01) following the January 2023 cutoff, while the Utstein ROSC rate experienced a much larger and statistically significant improvement by 14.93% (p<0.01). Placebo tests using the same specification and a January 2022 cutoff yield statistically insignificance, and the difference-in-RD estimates further validate that the observed discontinuity is unique to 2023.[5] Although this difference-in-RD effect is significant for Utstein group not the general group, this gives a benchmark showing how large is the effect when we rule out the seasonal change using the placebo effect at the 2022 cutoff. The positive effect remains robust when we restrict the sample to the East of England Ambulance Service, which was the region not affected by the industrial strikes over this period.

The above RDD treatment effects are the differences of fitted values at the cutoff time estimated by pre-cutoff and post-cutoff data respectively, while they may be sensitive to the choice of bandwidth and polynomial order of the fitting curve. We test this sensitivity and present it in the Table 5. Panel A varies the local bandwidth around the optimal value while holding the polynomial order fixed at one. The estimated effects remain consistently positive and statistically significant across a wide range of bandwidths. For the ROSC rate, coefficients decline gradually in magnitude

[4]The Utstein comparator group includes bystander-witnessed, non-traumatic cardiac arrests with a shockable rhythm, providing a clinically comparable sample across time (Shin *et al.*, 2021).
[5]See Figure A.4 for a visual plot of the RDD around the placebo cutoff.

Table 4: ROSC Rate RDiT: Main Results

Specification	Coefficient	Std Error	P-value	Bandwidth
Panel A: ROSC Rate				
Main RDD (2023)	0.0423***	(0.0148)	0.004	5.13
Placebo Test (2022)	0.0284*	(0.0164)	0.083	4.15
Diff-in-RD	0.0139	(0.0221)	0.530	–
East of England Only	0.0561	(0.0361)	0.120	4.89
Panel B: Utstein ROSC Rate				
Main RDD (2023)	0.1493***	(0.0405)	0.000	4.98
Placebo Test (2022)	0.0115	(0.0401)	0.775	5.22
Diff-in-RD	0.1379**	(0.0570)	0.017	–
East of England Only	0.1876**	(0.0845)	0.026	5.44

Notes: Dependent variables are trust-demeaned ROSC rates. Main RDD uses January 2023 cutoff. Placebo Test uses January 2022 cutoff. Diff-in-RD compares 2023 vs 2022 discontinuities. East England analysis uses non-strike ambulance trust only. All specifications use linear polynomial, triangular kernel, MSE-optimal bandwidth. Standard errors clustered at ambulance trust level. Monthly data, 2021-2024. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

from roughly 0.09 to 0.035 as the bandwidth widens, but the estimates remain highly significant throughout. For the Utstein sample, the point estimates follow a similar pattern, which is larger at smaller bandwidths. This behaviour is typical of RDD settings, as narrower bandwidths capture the sharp local change near the cutoff, while wider ranges incorporate more noise from surrounding periods. Panel B assesses sensitivity to polynomial specification using the optimal bandwidth for each outcome, where we change the polynomial order from 1 to 4. The estimated effect remains stable and significant across all polynomial orders, with larger magnitudes at higher-order fits. For the sake of avoiding the potential overfitting, we choose the linear function to show the main RDD effects reported above, which gives a more conservative estimate.

Finally, to eliminate the concern that our RDD estimates are just driven by other short term unobservable variations, we perform a Donut RDD analysis. The idea of Donut RDD is removing the data points around the cutoff (in the "donut"), then we use the remaining data to estimate the effect. Figure 7a and Figure 7b present the corresponding Donut RDD analyses, in which observations within ± 1 month of the cutoff are excluded[6]. The donut plots reveal that the positive shift in both groups persists even after removing the months closest to the cutoff, confirming our discontinuity of ROSC rates and the estimated impacts are robust.

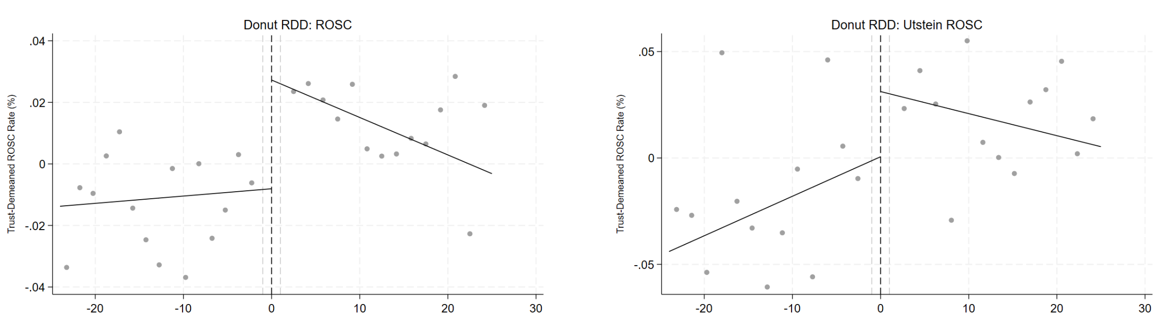
[6] As our optimal bandwidth is very narrow, which is only a 5-month time window, we have to just perform the Donut RDD by removing ± 1 month data points around the cutoff to get a estimation.

Table 5: ROSC Rate Sensitivity Analysis

ROSC Rate			Utstein ROSC		
Parameter	Coefficient	P-value	Parameter	Coefficient	P-value
Panel A: Bandwidth Sensitivity (Linear Polynomial)					
h=2.56	0.0907***	0.002	h=2.49	0.2640***	0.000
h=3.08	0.0771***	0.001	h=2.99	0.2661***	0.000
h=3.59	0.0528***	0.002	h=3.48	0.1842***	0.000
h=4.10	0.0469***	0.002	h=3.98	0.1688***	0.000
h=4.61	0.0448***	0.003	h=4.48	0.1542***	0.000
h=5.13	**0.0423*****	**0.004**	**h=4.98**	**0.1493*****	**0.000**
h=5.64	0.0382***	0.007	h=5.47	0.1383***	0.001
h=6.15	0.0365***	0.009	h=5.97	0.1332***	0.001
h=6.66	0.0360**	0.012	h=6.47	0.1292***	0.001
h=7.18	0.0353**	0.015	h=6.97	0.1270***	0.001
h=7.69	0.0345**	0.018	h=7.46	0.1158***	0.003
Panel B: Polynomial Order Sensitivity (Optimal Bandwidth)					
p=1	**0.0423*****	**0.004**	**p=1**	**0.1493*****	**0.000**
p=2	0.0401**	0.015	p=2	0.1858***	0.000
p=3	0.0857***	0.000	p=3	0.2273***	0.000
p=4	0.1738**	0.012	p=4	0.2591***	0.000

Notes: Dependent variables are trust-demeaned ROSC rates. Panel A tests sensitivity to bandwidth choice using linear polynomial (p=1). Bandwidth values differ between outcomes as each uses multiples of its own optimal bandwidth. Panel B tests sensitivity to polynomial order using optimal bandwidth for each outcome. All specifications use triangular kernel and MSE-optimal bandwidth selection. Standard errors clustered at ambulance trust level. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Figure 7: ± 1 month Donut: ROSC Rate



(a) ROSC Rate

(b) ROSC Rate for Utstein Patients

Taken together, the ROSC results provide strong evidence of the causal mechanism linking system performance and survival. Faster ambulance response times increased the likelihood that patients received timely CPR or defibrillation, allowing spontaneous circulation to be restored before hospital arrival. This immediate gain forms the first link in the broader chain through which improved emergency care system performance translated into higher survival and lower mortality, which we examine next.

5.3.2 Intermediate clinical outcome: survival rate at 30 days

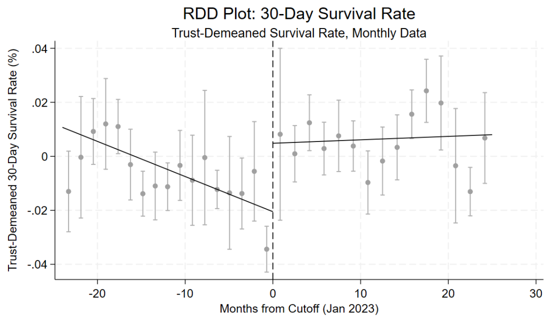
We proceed by examining the 30-day survival rate following cardiac arrest as the primary outcome of interest. This measure provides a direct and time-sensitive indicator of full emergency system performance, as patient survival after out-of-hospital cardiac arrest depends critically on both early resuscitation interventions and timely hospital care.

Figure 8a and Figure 8b present the regression discontinuity plots for the overall 30-day survival rate and the Utstein comparator group, respectively. In both cases, we observe a clear and discrete upward shift in survival immediately after the January 2023 cutoff, consistent with the timing of the sharp reduction in ambulance response times documented earlier (see Section 5.1). The increase is more visible in the general survival rate than in the Utstein comparator group. This pattern aligns with the clinical nature of each outcome: Utstein patients whose cardiac arrests are witnessed and typically receive early bystander intervention face lower death risks and are therefore less sensitive to additional minutes of ambulance delay. In contrast, the broader patient population includes many unwitnessed arrests in which survival depends critically on the speed of professional response. The observed discontinuity thus provides evidence that the improvement in full emergency care system performance at the start of 2023 translated into better survival outcomes, particularly among those most reliant on timely emergency care.

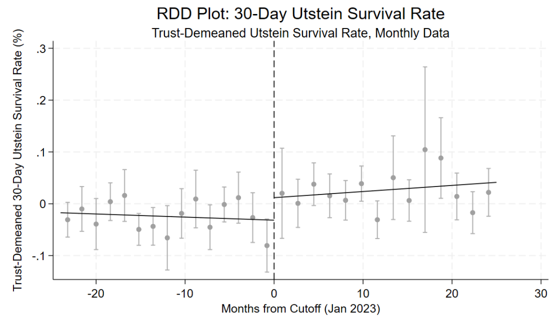
Table 6 summarizes the formal RDiT estimates. For the overall 30-day survival rate, the main RDD coefficient indicates an increase of 0.036 percentage points following the 2023 cutoff ($p < 0.01$), implying a statistically significant improvement in patient outcomes. This magnitude, even though modest, is still meaningful given the low baseline survival rate in out-of-hospital cardiac arrest cases. For the Utstein comparator group, the estimated effect is with larger standard error and only marginally significant ($p < 0.10$). This weaker response aligns with the less prominent discontinuity observed in the Figure 8b.

The placebo test using the January 2022 cutoff shows no comparable discontinuity, suggesting that the estimated 2023 jump is not driven by seasonal variation, reporting artefacts, or unrelated

Figure 8: Regression Discontinuity in Time: Survival Rate at 30 Days



(a) Survival Rate at 30 Days



(b) Survival Rate at 30 Days for Utstein Patients

winter shocks.[7] The difference-in-RD (Diff-in-RD) estimate compares the discontinuities across the two years to net out any common seasonal patterns. Even though those estimates in both groups are not significant, they provide a suggestive positive effect of the improved response times on 30-day survival rates.

We also present the estimates only by using East of England data, where no strike action occurred during December 2022 or January 2023. The estimates are not statistically significant since this specification largely restrict the sample size, but the positive sign still suggests a improved trend of survival rate after the cutoff time in this area.

Sensitivity analyses presented in Table 7 demonstrate that these results are stable across a range of bandwidths and polynomial orders, and Figure 9a–Figure 9b show that the discontinuities persist after excluding observations within one month of the cutoff. These findings indicate that the improvement in ambulance system performance which driven by shorter response times produced immediate, measurable gains in out of hospital cardiac arrest outcomes, which leads to more patients survived.

5.3.3 Broader population-level outcome: mortality rate

While the above analysis provides a direct and clinically precise measure of the benefits from improved emergency care system response, it is restricted to a specific patient group experiencing cardiac arrest. In this section, we further examine broader population-level mortality outcomes. The total mortality rate captures all deaths in a local area, regardless of cause, offering a wider but less targeted view of health impacts. This measure allows us to test whether the improvement

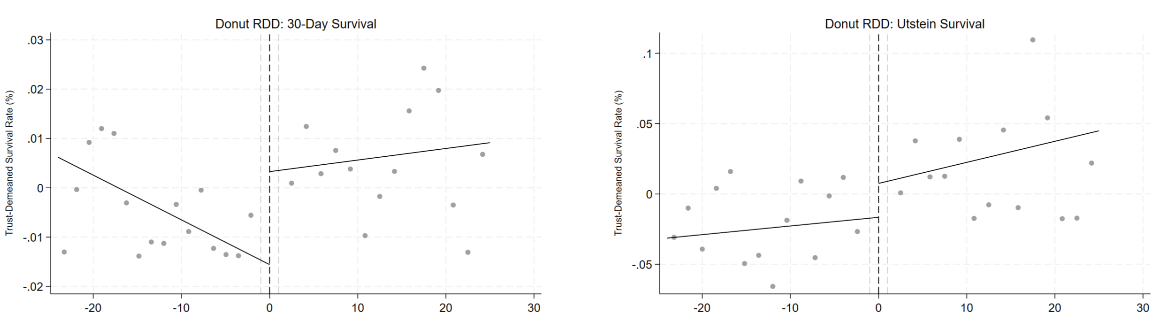
[7] See Figure A.5 for a visual plot of the RDD around the placebo cutoff.

Table 6: Survival Rate RDiT: Main Results

Specification	Coefficient	Std Error	P-value	Bandwidth
Panel A: 30-Day Survival Rate				
Main RDD (2023)	0.0363***	(0.0108)	0.001	5.05
Placebo Test (2022)	0.0115	(0.0135)	0.397	4.69
Diff-in-RD	0.0249	(0.0173)	0.154	–
East of England Only	0.0251	(0.0224)	0.262	6.33
Panel B: 30-Day Utstein Survival Rate				
Main RDD (2023)	0.0769*	(0.0435)	0.077	6.39
Placebo Test (2022)	0.0352	(0.0281)	0.209	6.07
Diff-in-RD	0.0416	(0.0518)	0.423	–
East of England Only	0.0351	(0.0750)	0.640	6.22

Notes: Dependent variables are trust-demeaned survival rates. Main RDD uses January 2023 cutoff. Placebo Test uses January 2022 cutoff. Diff-in-RD compares 2023 vs 2022 discontinuities. East England analysis uses non-strike ambulance trust only. All specifications use linear polynomial, triangular kernel, MSE-optimal bandwidth. Standard errors clustered at ambulance trust level. Monthly data, 2021-2024. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Figure 9: ± 1 month Donut: Survival Rate at 30 Days



(a) Survival Rate at 30 Days

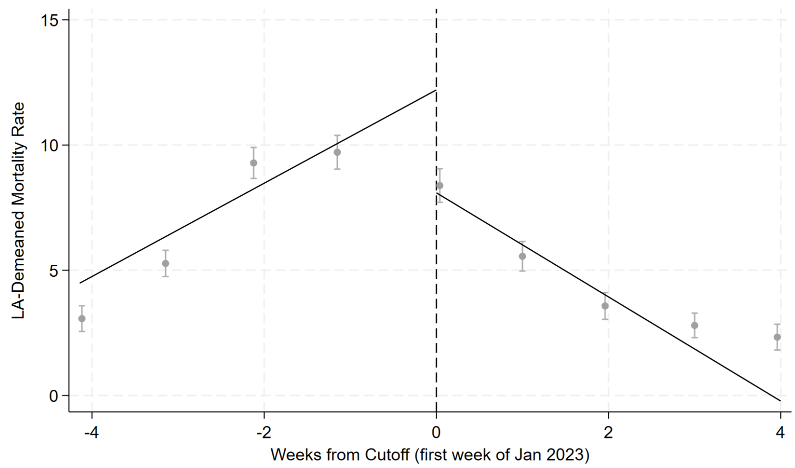
(b) Survival Rate at 30 Days for Utstein Patients

Table 7: Survival Rate Sensitivity Analysis

30-Day Survival			Utstein Survival		
Parameter	Coefficient	P-value	Parameter	Coefficient	P-value
Panel A: Bandwidth Sensitivity (Linear Polynomial)					
h=2.53	0.0599***	0.000	h=3.19	0.1305***	0.004
h=3.03	0.0579***	0.000	h=3.83	0.0996**	0.034
h=3.54	0.0446***	0.000	h=4.47	0.0909*	0.053
h=4.04	0.0414***	0.000	h=5.11	0.0861*	0.065
h=4.55	0.0379***	0.000	h=5.75	0.0793*	0.080
h=5.05	**0.0363*** **	**0.001**	**h=6.39**	**0.0769***	**0.078**
h=5.56	0.0336***	0.002	h=7.03	0.0756*	0.075
h=6.06	0.0324***	0.003	h=7.67	0.0670	0.110
h=6.57	0.0316***	0.002	h=8.31	0.0626	0.130
h=7.08	0.0309***	0.001	h=8.94	0.0602	0.141
h=7.58	0.0293***	0.002	h=9.58	0.0589	0.140
Panel B: Polynomial Order Sensitivity (Optimal Bandwidth)					
p=1	**0.0363*** **	**0.001**	**p=1**	**0.0769***	**0.077**
p=2	0.0382***	0.002	p=2	0.0878*	0.060
p=3	0.0474***	0.001	p=3	0.1237***	0.008
p=4	0.0772***	0.000	p=4	0.1614**	0.034

Notes: Dependent variables are trust-demeaned survival rates. Panel A tests sensitivity to bandwidth choice using linear polynomial (p=1). Bandwidth values differ between outcomes as each uses multiples of its own optimal bandwidth. Panel B tests sensitivity to polynomial order using optimal bandwidth for each outcome. All specifications use triangular kernel and MSE-optimal bandwidth selection. Standard errors clustered at ambulance trust level. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Figure 10: Regression Discontinuity in Time around the Cutoff: Mortality Rate



in ambulance performance also benefited patients with non-cardiac emergencies or other acute conditions requiring timely hospital admission. However, because population-level mortality aggregates deaths unrelated to emergency care access, such as the death happened before or without the need of any emergency service, the true effect of faster emergency response may be diluted or obscured by this additional noise. Nevertheless, documenting mortality around the same cutoff provides a complementary perspective on whether the improvement in ambulance performance translated into measurable effects at the population level.

Figure 10 plots the regression discontinuity in time for the local mortality rate. Although the fitted lines suggest a decline after the first week of 2023, the observed data points do not exhibit a clear drop at the cutoff. The rise of fitted curve before the cutoff is largely driven by overfitting in the pre-period polynomial.

As the total mortality rate does not present clear discontinuity but provides a suggestive declining trend after the cutoff, we intend to further examine the suggestive effects across heterogenous patient groups by locations. Table 8 decomposes the effects by location of death. The most pronounced reductions occur for hospital, home, and care-home deaths, whereas hospice and other locations show no meaningful change. These results suggest that any improvement is concentrated in settings more directly linked to ambulance-delivered patients.

Taken together, the visual RDiT result suggests a declining trend of mortality rate in local areas, although there is not enough evidence showing the discontinuity of total mortality rate. We found this declining trend is only significant in several specific patient groups, such as home, hospital, and care home, but not in hospice, and other locations.

Table 8: RDiT Effects by Mortality Category

Mortality Category	Coefficient	Std Error	P-value
Total	-4.6114***	(0.5619)	0.000
Home	-1.7036***	(0.1966)	0.000
Hospital	-2.1330***	(0.3284)	0.000
Care Home	-1.2231***	(0.2253)	0.000
Hospice	0.0319	(0.0331)	0.335
Other	-0.0087	(0.0101)	0.387
Else	-0.0293	(0.0269)	0.277

Notes: Each row shows RDD treatment effect for different mortality categories. All outcomes are LA-demeaned mortality rates per 100,000 population. All specifications use triangular kernel with quadratic polynomial and MSE-optimal bandwidth. Standard errors clustered at local authority level. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

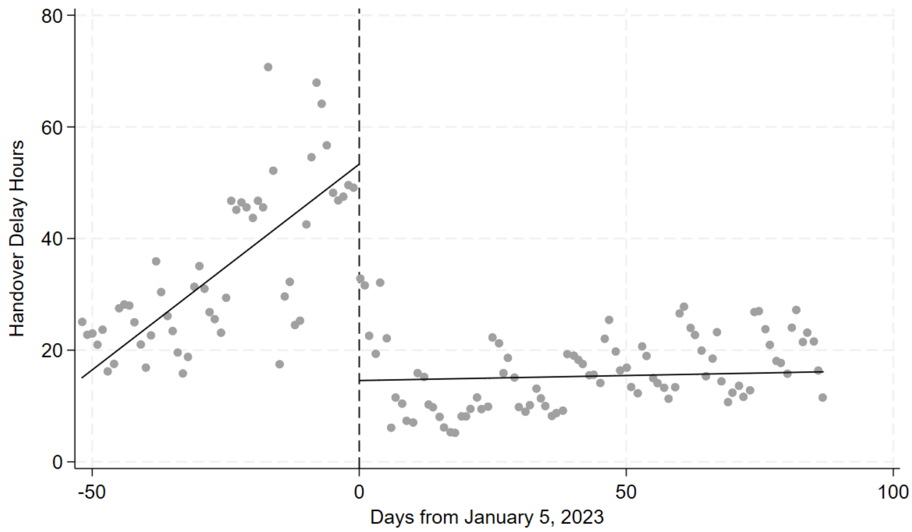
5.4 Mechanisms: Handover Delay to Hospital

A key operational mechanism underlying the improvement in ambulance response times and subsequent survival outcomes is the reduction in hospital handover delays. The handover time refers to the period between an ambulance's arrival at a hospital and the formal transfer of the patient's care to hospital staff. When emergency departments operate near capacity, ambulances have to wait and they are still clinically responsible for the patients. A handover delay occurs when this process exceeds 15 minutes. Long handovers constrain the supply of ambulances in service, therefore resulting in longer ambulance response times.

Figure A.6 plots the daily proportion of ambulance arrivals across NHS Foundation Trusts that experienced handover times exceeding 30 minutes between November 2022 and April 2023. The vertical line marks 1 January 2023, after this cutoff there is the time when all areas experience the sharp drop in response times. The share of delayed handovers peaked in late December 2022 at nearly 45% but fell sharply thereafter, stabilising at roughly 20%–30% through the first quarter of 2023. This visible break aligns closely with the timing of the ambulance response time improvement documented earlier, suggesting that shorter handovers at hospitals freed up ambulance capacity in the field.

Figure 11 provides a regression discontinuity plot of total handover delay hours at the daily trust level. The fitted lines show a sharp structural break at the 5th January 2023 cutoff (the middle of the first week of 2023): total delay hours drop abruptly, consistent with a substantial release of ambulance capacity. The discontinuity is visually strong and statistically clear.

Figure 11: Discontinuity in Total Daily Handover Delay Hours



To examine the timing of this improvement in finer detail, Figure A.7 presents a daily event-study of handover delay hours relative to 1 January 2023. The red shaded area highlights the first full week of 2023 (2–8 January), during which we observe a sharp and statistically significant reduction in total handover delay hours. This period coincides precisely with the week in which the improvement in ambulance response times began in our weekly data. After this point, the estimated coefficients remain persistently negative throughout the sample window. This timing aligns closely with the policy context: during December 2022 to March 2023, the UK government allocated approximately £700 million to emergency system relief, with a central focus on improving hospital patient discharge.

Collectively, these results indicate that the improvement in ambulance response times around January 2023 was closely linked to a sharp and sustained reduction in hospital handover delays. By reducing the time ambulances spent waiting to transfer patients, hospitals effectively increased the number of vehicles available for dispatch, thereby shortening response times to new calls. This mechanism provides a direct operational explanation for the observed discontinuities in emergency performance and the associated improvements in patient survival outcomes.

6 Discussion and Concluding Remarks

This paper provides new causal evidence on the short-run effectiveness of agile, targeted health spending designed to relieve specific system bottlenecks. Using a sharp policy-induced improvement in emergency care system performance at the start of 2023, we evaluate how quickly and through which mechanisms such spending translates into health gains.

The intervention we study is the £700 million discharge fund allocated by the UK government between December 2022 and March 2023, which aimed to reduce hospital congestion by accelerating patient discharges and improving ambulance handover efficiency. This funding produced a sudden and measurable improvement in the emergency care system, as shown by discontinuous reductions in both ambulance handover delays and response times across England.

By combining multiple datasets linking emergency operations to health outcomes, we find clear evidence that the operational gains from this funding translated into meaningful clinical improvement. The probability that a cardiac arrest patient achieved return of spontaneous circulation on arrival at hospital rose significantly, which partially reflects that the improvement of pre-hospital emergency care system performance. Subsequently, 30-day survival rate of those cardiac arrest patients also improved, indicating that faster pre-hospital treatment and reduced handover delays produced sustained health benefits. Broader local area mortality outcomes moved in the same direction, albeit with greater noise, consistent with a diluted effect when non-emergency deaths are included as well.

These results highlight the value of flexible and targeted investment in health system bottlenecks. Rather than increasing general health budgets or inputs, policies that focus on restoring system flow can yield rapid and measurable returns in patient outcomes. In contrast to much of the economic literature that studies returns to standard, pre-planned health spending, our findings emphasize the high marginal impact of responsive emergency spending designed to stabilize the system during acute stress and wide crisis.

Future work will deepen this analysis along two fronts. First, we plan to develop a behavioural model of emergency resource allocation to examine how prioritization of higher-acuity calls (Category 1) may crowd out responses to lower-priority cases (Category 2), shedding light on equity–efficiency trade-offs within constrained systems. Second, we will extend our welfare analysis to assess the cost-effectiveness and policy returns of targeted emergency investments relative to broader NHS funding allocations.

References

ADU-AMANKWAAH, J. (2024). Behind the shadows: Bringing the cardiovascular secrets of long covid into light. *European Journal of Preventive Cardiology*, **32** (6), 499–501.

ALARHAYEM, A., MYERS, J., DENT, D., LIAO, L., MUIR, M., MUELLER, D., NICHOLSON, S., CESTERO, R., JOHNSON, M., STEWART, R., O'KEEFE, G. and EASTRIDGE, B. (2016). Time is the enemy: Mortality in trauma patients with hemorrhage from torso injury occurs long before the "golden hour". *The American Journal of Surgery*, **212** (6), 1101–1105.

ALMOND, D., DOYLE, J., JOSEPH J., KOWALSKI, A. E. and WILLIAMS, H. (2010). Estimating marginal returns to medical care: Evidence from at-risk newborns. *The Quarterly Journal of Economics*, **125** (2), 591–634.

ATHEY, S. and STERN, S. (2002). The impact of information technology on emergency health care outcomes. *The RAND Journal of Economics*, **33** (3), 399–432.

BARRECA, A. I., GULDI, M., LINDO, J. M. and WADDELL, G. R. (2011). Saving babies? Revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics*, **126** (4), 2117–2123.

BERMAN, Y. and HOVLAND, T. (2024). *The impact of austerity on mortality and life expectancy*. Working Paper 139, International Inequalities Institute, London School of Economics and Political Science.

CALONICO, S., CATTANEO, M. D. and TITIUNIK, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82** (6), 2295–2326.

CATTANEO, M. D., IDROBO, N. and TITIUNIK, R. (2024). *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Elements in Quantitative and Computational Methods for the Social Sciences, Cambridge University Press.

CHAN, D. C., CARD, D. and TAYLOR, L. (2023). Is there a VA advantage? evidence from dually eligible veterans. *American Economic Review*, **113** (11), 3003–43.

CUTLER, D., MCCLELLAN, M. and NEWHOUSE, J. (1998). *The Costs and Benefits of Intensive Treatment for Cardiovascular Disease*. Working Paper 6514, NBER.

CUTLER, D. M. (2007). The lifetime costs and benefits of medical technology. *Journal of Health Economics*, **26** (6), 1081–1100.

DAINES, L., ZHENG, B., PFEFFER, P., HURST, J. and SHEIKH, A. (2022). A clinical review of long-COVID with a focus on the respiratory system. *Current Opinion in Pulmonary Medicine*, **28** (3), 174–179.

DAVINO-RAMAYA, C. M., LIPPY, S. H., NATURALE, A., FEIST, J. C., CIPRIANO, P. and CLARK, D. (2023). Burnout and the effect on the global health care workforce crisis: An expert panel discussion. *The Permanente Journal*, **27** (2), 111–122.

DEPARTMENT OF HEALTH AND SOCIAL CARE (2023a). Handbook to the NHS constitution for England.

DEPARTMENT OF HEALTH AND SOCIAL CARE (2023b). Rapid evaluation of the 2022 to 2023 discharge funds.

DOYLE, J. J., GRAVES, J. A., GRUBER, J. and KLEINER, S. A. (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy*, **123** (1), 170–214.

FETZER, T. and RAUH, C. (2022). *Pandemic Pressures and Public Health Care: Evidence from England*. Working paper, arXiv.

GRUBER, J., HENDREN, N. and TOWNSEND, R. M. (2014). The Great Equalizer: Health Care Access and Infant Mortality in Thailand. *American Economic Journal: Applied Economics*, **6** (1), 91–107.

—, HOE, T. P. and STOYE, G. (2018). *Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers*. Working Paper 24445, NBER.

HAUSMAN, C. and RAPSON, D. S. (2018). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, **10** (Volume 10, 2018), 533–552.

HOE, T. P. (2019). *Does Hospital Crowding Matter? Evidence from Trauma and Orthopedics in England*. Working paper.

HOLMÉN, J., HERLITZ, J., RICKSTEN, S., STRÖMSÖE, A., HAGBERG, E., AXELSSON, C. and RAWSHANI, A. (2020). Shortening ambulance response time increases survival in out-of-hospital cardiac arrest. *Journal of the American Heart Association*, **9** (21), e017048.

KNUTSSON, D. and TYREFORS, B. (2022). The quality and efficiency of public and private firms: Evidence from ambulance services. *The Quarterly Journal of Economics*, **137** (4), 2213–2262.

LUCCHESE, E. (2024). How important are delays in treatment for health outcomes? the case of ambulance response time and cardiovascular events. *Health Economics*, **33** (4), 652–673.

MARTIN, S., RICE, N. and SMITH, P. C. (2008). Does health care spending improve health outcomes? evidence from english programme budgeting data. *Journal of Health Economics*, **27** (4), 826–842.

NEUMAR, R. W., SHUSTER, M., CALLAWAY, C. W., GENT, L. M., ATKINS, D. L., BHANJI, F., BROOKS, S. C., DE CAEN, A. R., DONNINO, M. W., FERRER, J. M. E., KLEINMAN, M. E., KRONICK, S. L., LAVONAS, E. J., LINK, M. S., MANCINI, M. E., MORRISON, L. J., O'CONNOR, R. E., SAMSON, R. A., SCHEXNAYDER, S. M., SINGLETARY, E. M., SINZ, E. H., TRAVERS, A. H., WYCKOFF, M. H. and HAZINSKI, M. F. (2015). 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care part 1: Executive Summary. *Circulation*, **132** (18_suppl_2), S315–S367.

NHS WALES JOINT COMMISSIONING COMMITTEE (2025). Clinical response model.

NOACK, C. and ROTHE, C. (2023). *Donut Regression Discontinuity Designs*. Working paper.

NOLAN, J. P., SOAR, J., ZIDEMAN, D. A., BIARENT, D., BOSSAERT, L. L., DEAKIN, C., KOSTER, R. W., WYLLIE, J. and BÖTTIGER, B. (2010). European resuscitation council guidelines for resuscitation 2010 section 1. Executive summary. *Resuscitation*, **81** (10), 1219–1276.

OFFICE FOR NATIONAL STATISTICS (2023). Deaths registered weekly in England and Wales.

PELL, J. P., SIREL, J. M., MARSDEN, A. K., FORD, I. and COBBE, S. M. (2001). Effect of reducing ambulance response times on deaths from out of hospital cardiac arrest: cohort study. *BMJ*, **322** (7299), 1385–1388.

RENTSCHLER, J., KLAIBER, C., TARIVERDI, M., DESJONQUERES, C. and MERCADANTE, J. (2021). Frontline: Preparing healthcare systems for shocks from disasters to pandemics.

SHIN, J., CHOCRON, R., REA, T., KUDENCHUK, P., MCNALLY, B. and EISENBERG, M. (2021). Merits of expanding the utstein case definition for out of hospital cardiac arrest. *Resuscitation*, **158**, 88–93.

STODDART, K., COWIE, J., ROBERTSON, T., BUGGE, C., DONALDSON, J. and ANDREIS, F. (2019). *The Scottish Ambulance Service New Clinical Response Model Evaluation Report*. Tech. rep., University of Stirling.

TSAMPASIAN, V., BÄCK, M., BERNARDI, M., CAVARRETTA, E., DĘBSKI, M., GATI, S., HANSEN, D., KRÄNKEL, N., KOSKINAS, K. C., NIEBAUER, J., SPADAFORA, L., FRIAS VARGAS, M., BIONDI-ZOCCAI, G. and VASSILIOU, V. S. (2024). Cardiovascular disease as part of long covid: a systematic review. *European Journal of Preventive Cardiology*, **32** (6), 485–498.

UNISON (2023). Nurses, paramedics, blood collection workers and other NHS staff in new strike over pay.

WALSHE, K., SMITH, J., LAMONT, T. and LEARY, A. (2024). Workforce research priorities for resilience to future health shocks—and the workforce crisis. *BMJ*, **387**.

WILLIAMS, B. A., JONES, C. H., WELCH, V. and TRUE, J. M. (2023). Outlook of pandemic preparedness in a post-COVID-19 world. *npj Vaccines*, **8**, 1–12.

WITTER, S., THOMAS, S., TOPP, S. M., BARASA, E., CHOPRA, M., COBOS, D., BLANCHET, K., TEDDY, G., ATUN, R. and AGER, A. (2023). Health system resilience: a critical review and reconceptualisation. *The Lancet Global Health*, **11**, e1454–e1458.

YOON, S., GOH, H., CHAN, A., MALHOTRA, R., VISARIA, A., MATCHAR, D., LUM, E., SENG, B., RAMAKRISHNAN, C., QUAH, S., KOH, M. S., TIEW, P. Y., BEE, Y. M., ABDULLAH, H., NADARAJAN, G. D., GRAVES, N., JAFAR, T. and ONG, M. E. (2022). Spillover effects of COVID-19 on essential chronic care and ways to foster health system resilience to support vulnerable non-COVID patients: A multistakeholder study. *Journal of the American Medical Directors Association*, **23** (1), 7–14.

Online Appendix – Not for Print

The Long and Short of It: Ambulance Response Times and Mortality

Damian Clarke & Ye Yuan

A Appendix Figures and Tables

Figure A.1: Weekly Trends of Ambulance Data and Mortality Rate

Figure A.2: Ambulance Incidents and Response Times by Postcode Districts

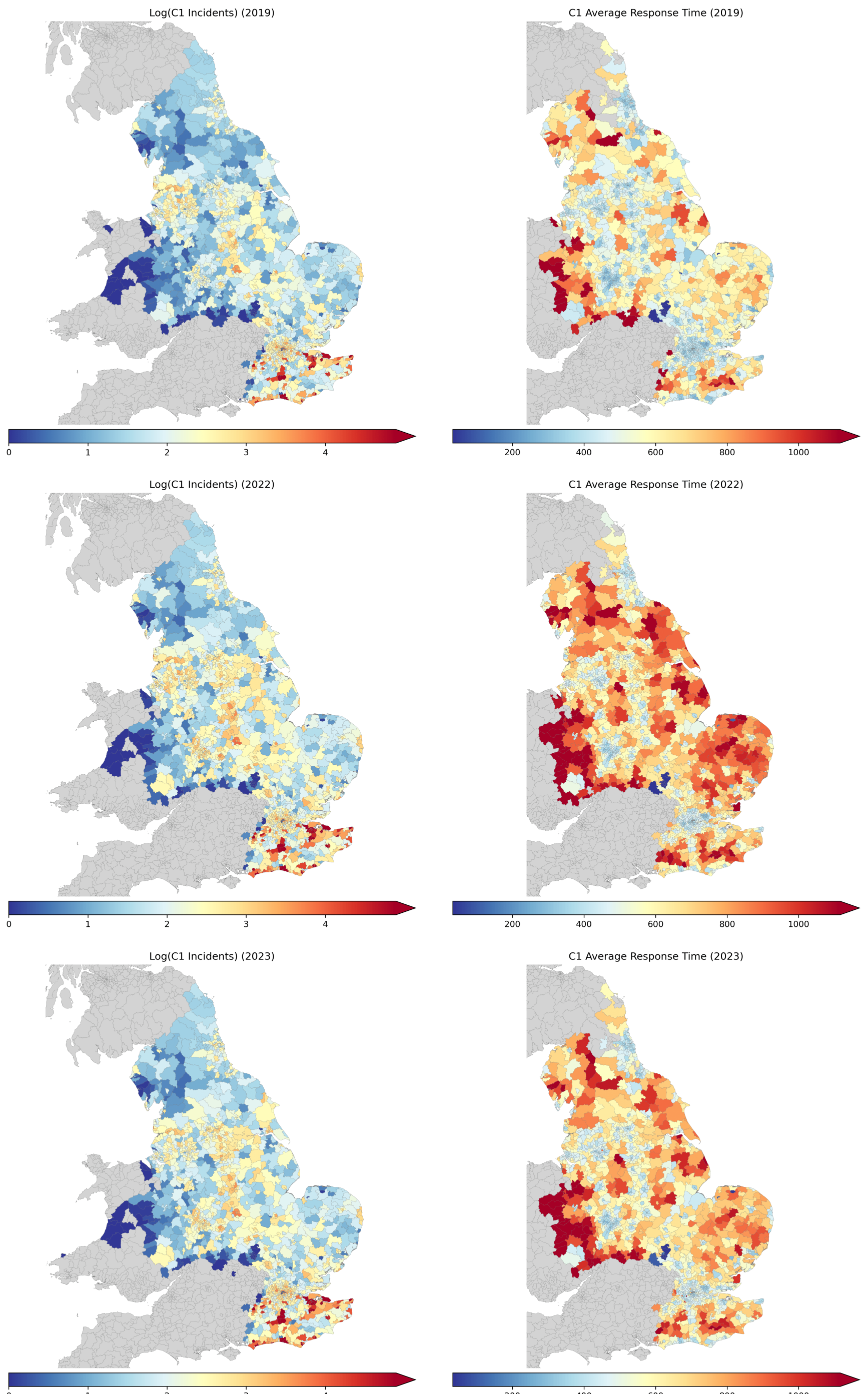


Figure A.3: Ambulance Data and Mortality Rate by Local Authorities

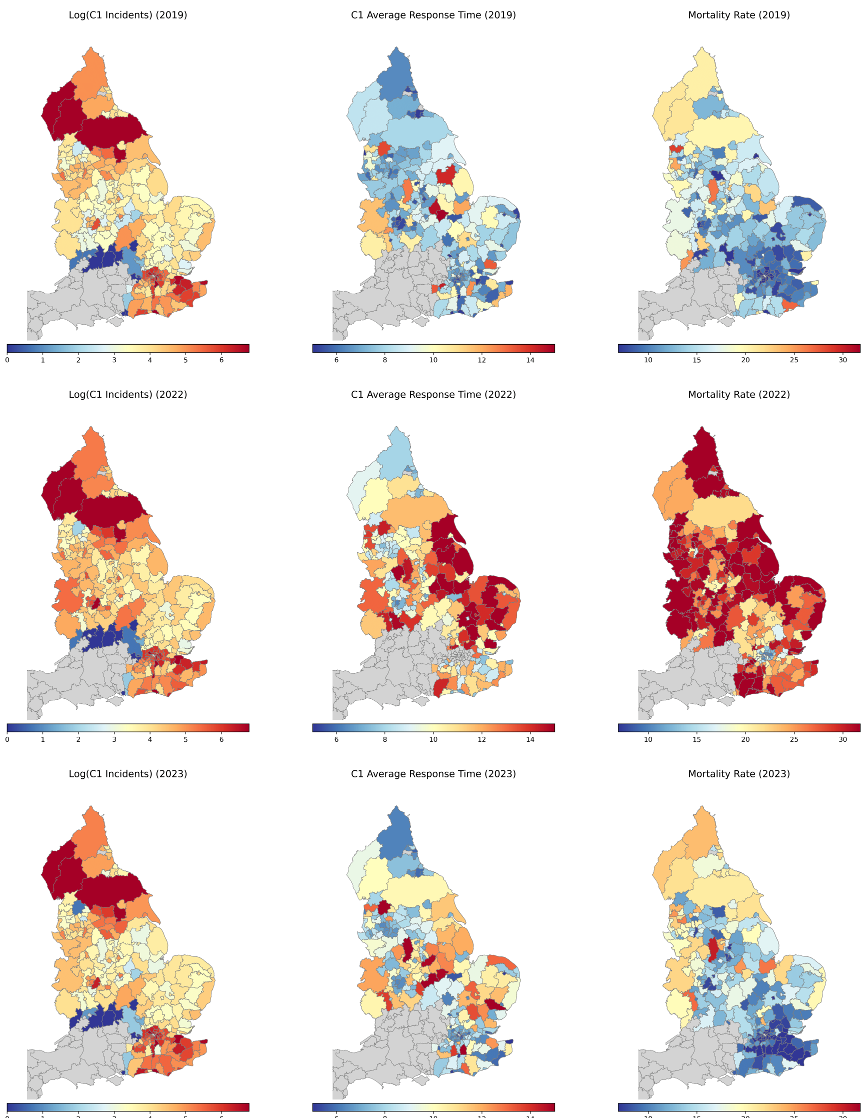
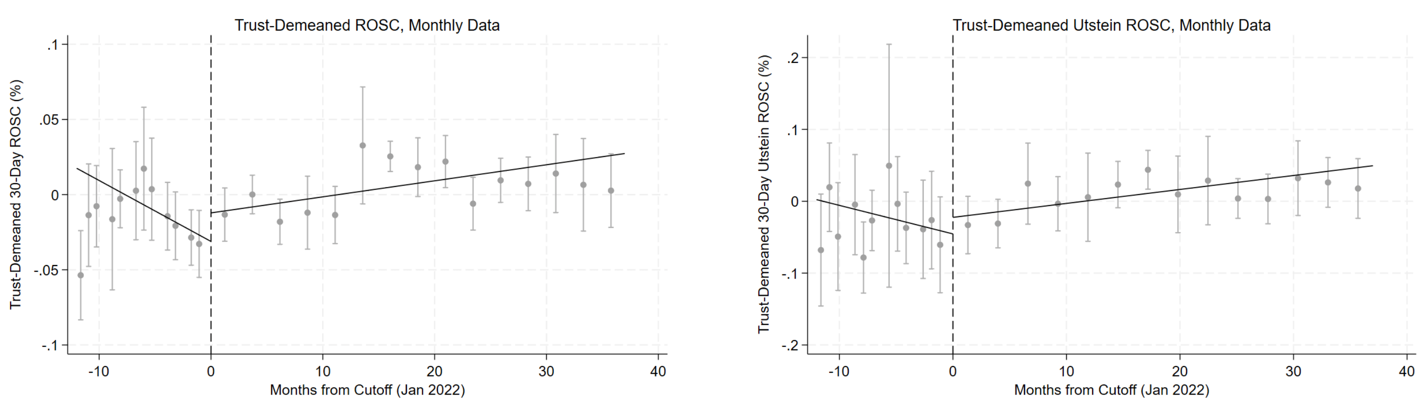


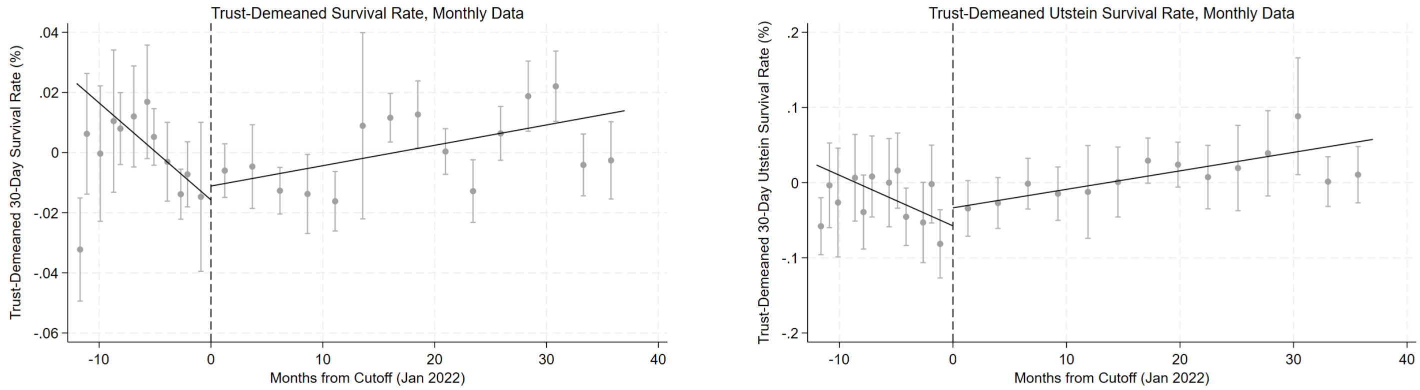
Figure A.4: Regression Discontinuity in Time: ROSC Rate at the placebo time cutoff January 2022



(a) ROSC Rate

(b) ROSC Rate for Utstein Patients

Figure A.5: Regression Discontinuity in Time: Survival Rate at 30 Days at the placebo time cutoff January 2022



(a) Survival Rate at 30 Days

(b) Survival Rate at 30 Days for Utstein Patients

Figure A.6: Share of ambulance arrivals experiencing handover delays exceeding 30 minutes

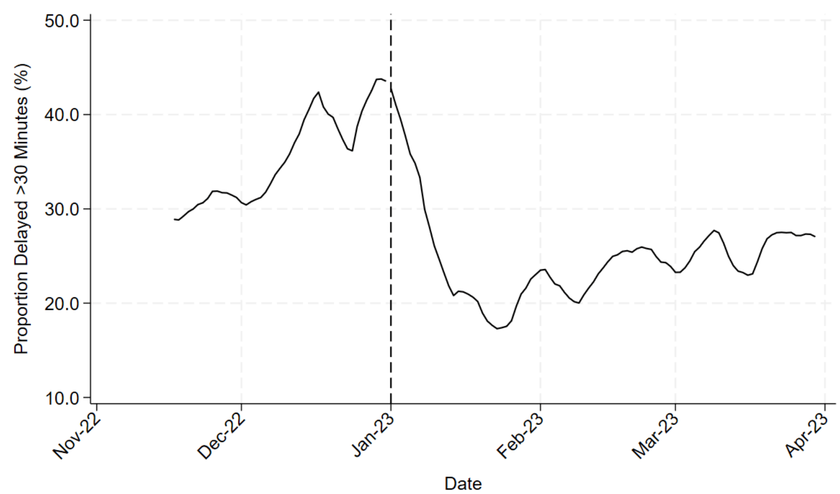


Figure A.7: Daily event study of total handover delay hours