Revisiting the Causal Impact of Response Time on Health Outcomes: a Non-Parametric Estimation Using Air Temperature as an IV*

Thomas Pelloquin[†], Baptiste Haon[‡], Pauline Chauvin[§], Lise Rochaix[¶]

March 1, 2025

Abstract

Although delays in treatment adversely affect patient health outcomes, establishing the precise nature of this relationship raises important methodological challenges. Healthcare production is not only time-dependent but also varies greatly between individuals; this heterogeneity arises from unobserved or endogenous variables, particularly the severity of the condition, the rate of health depreciation or patient health literacy. This paper leverages original large-scale individual data, containing precise time measures for 10,250 episodes of acute ischaemic strokes (AIS), starting at symptom onset, as well as precise data on patient health outcomes -transformed into a utility index-, to address this complex relationship over the complete response time segment. By choosing an original instrument —air temperature— in a non-parametric estimation, we identify the average treatment effect (ATE) of an increase in response time, as well as the average treatment effect on the treated (ATT). Furthermore, the heterogeneity in patients' returns to reducing response time is examined. Comparing the estimates of the relationship identified using an IV method with existing results in the clinical literature highlights the value of a robust identification strategy in correcting for effects' underestimation. Our approach suggests a near-linear relationship between delays and outcomes, challenging the conventional view of bounded effects in the management of AIS. Our ATT results support existing clinical recommendations, but policy-relevant treatment effects require a more precise account of patient benefits' distribution and health loss valuations.

JEL Codes: I10, C31, D61

Index

AIS: Acute Ischemic Stroke
CSC: Comprehensive Stroke Centres
EMS: Emergency Medical Services
IV: Instrumental Variable
LVO: Large Vessels Occlusion
POC devices: Point of Care devices
RT: Response Time
MT: Mechanical Thrombectomy

*The authors express their sincere gratitude to Mikael Mazighi, principal investigator of the RHU BOOSTER project, for his continuous support and interest, and to Bertrand Lapergue who provided access and guidance to the ETIS database, as well as members of the RHU BOOSTER for their contributions. This work is supported by a government grant managed by the French National Research Agency (ANR) as part of the future investment program integrated into France 2030, under grant agreement No. ANR-18-RHUS-0001.

†Corresponding author: thomas.pelloquin@psemail.eu - Paris School of Economics (Hospinnomics) - French school of public health

‡Paris School of Economics (Hospinnomics) - GATE, Lyon 2

§Université Paris Cité, LIRAES, F-75006 Paris, France

¶Paris School of Economics (Hospinnomics) - Paris 1 Panthéon-Sorbonne

1 Introduction

Healthcare, as an economic good, possesses distinctive characteristics, particularly for time-sensitive conditions. The theoretical framework established by Becker [2007] and Grossman [1972] conceptualizes health as a capital stock subject to depreciation. Unlike most commodities, delays in healthcare delivery can result in irreversible health deterioration, with waiting costs escalating non-linearly over time. While timely access to emergency care is paramount for acute conditions, time sensitivity is also critical for early cancer detection, which significantly enhances treatment efficacy, or for infectious diseases, in order to control epidemic growth.

Effectively reducing time to adequate treatment constitutes a significant public health challenge, shaped by both supply-side factors (e.g., geographical availability of healthcare facilities) and demand-side determinants (e.g., patients' ability to recognize first symptoms). Regarding the spatial allocation of facilities, regulators are faced with the trade-off between centralising specialised resources to enhance efficiency and safety while improving equity by reducing distance to facilities. Disparities in healthcare access, particularly in under-resourced regions, have been shown to exacerbate socioeconomic inequalities in health outcomes Bertoli and Grembi [2017], Turner et al. [2022]. Additionally, delays in symptom recognition among specific subpopulations contribute to adverse health outcomes. This phenomenon is particularly evident in cardiovascular diseases among women, where symptom presentation diverges from conventional clinical expectations, often resulting in under-recognition Lichtman et al. [2015].

Robust evidence on the value of further reducing time to adequate treatment is still lacking to guide policy. Existing studies have consistently documented strong associations between delays in access to treatment and patient health outcomes. Establishing causality in healthcare, which is the purpose of this paper, remains a challenge due to data constraints and the presence of unmeasured confounders Angrist et al. [2024]. Regarding the former, most of the data covers truncated care pathways, ignoring the exact starting point of the disease. Confounders, for their part, mostly relate to selection issues in relation to severity.

On the supply-side, clinicians' management of acute events relies on a severity-based patient prioritization, whether upon arrival at the healthcare facility or during pre-hospital decision-making (e.g. the selection of transportation modes tailored to severity levels). Best practices, clinical heuristics, and data increasingly inform these prioritization strategies Siciliani et al. [2015] and empirical evidence shows that official triage protocols tend to prioritize patients presenting with the most severe symptoms in emergency settings Zachariasse et al. [2019].

On the demand side, patients' ability to seek care is influenced by multiple factors. Two primary dimensions warrant consideration: the ability to interpret symptoms (some conditions sharing the same symptoms) and individuals' health literacy (e.g. their ability to recognize symptoms). Goff [1998] founds that individuals with prior experience of heart disease exhibited greater awareness of myocardial infarction symptoms, potentially facilitating more rapid medical attention. Similarly, in stroke management, patients exhibiting more overtly severe symptoms tend to present earlier for treatment.

The impact of selection on treatment heterogeneity has been extensively explored in the economics literature, particularly in education Angrist and Krueger [1991]. A parallel selection mechanism operates in acute healthcare settings, with two primary sources of selection bias: patients self-select based on symptom severity, and healthcare providers further prioritize patients whom they perceive as deriving the greatest benefit from early intervention. The interplay of supply and demand determinants results in heterogeneity in the returns to minimizing RT, contingent on the severity as perceived by both providers and patients.

In this paper, we address causality between time and health outcomes using Acute Ischemic Stroke (AIS) as a case study, in order to assess whether non-causal estimations lead to underestimate the benefits of reducing time to treatment. We use rich individual data with 10,250 hospital stays for Acute Ischemic Stroke (AIS) of the most severe type (Large Vessel Occlusion - LVO), a condition known to be extremely time sensitive [Mazighi et al., 2013], between March 2017 and November 2023. Compared to other conditions such as cancer, stroke symptoms are relatively easy to detect, allowing a precise collection of the time of symptom onset upon arrival at EMS. As a result, this study focuses on the full time segment, from symptom onset to access to care, hereafter defined as Response Time (RT). RT is multifactorial, since it includes demand-side factors, such as patients' ability to seek care upon symptom onset, and supply-side factors, such as the distance to the nearest EMS and its efficiency at handling cases, especially under strong capacity constraints.

Beyond mortality rates, this study uses measures of utility loss, leveraging a mapping function to predict utility scores from a neurological disability index Rivero-Arias et al. [2009]. To estimate the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT), we use an innovative instrumental variable (IV) approach within a partially non-parametric control function (CF) framework. Specifically, air temperature (measured in degrees Celsius) is utilized as an instrument. Air temperature is associated with various socio-behavioral factors that influence RT from both the demand and supply sides, exhibiting a negative correlation with RT. The validity of this instrument is rigorously tested and discussed D'Haultfœuille et al. [2024].

Our results show that, due to the negative correlation between RT and severity, previous results have underestimated the benefits of further reducing RT on health outcomes. Comparing the naive versus instrumented estimation yields important policy implications: the existing prioritisation of patients by EMS is already playing a crucial role in improving health outcomes of the most severe patients. Additional benefits would be derived from further reducing RT, whether by improving the territorial distribution of EMS, or alternatively, by dedicating fast-track pathways for ambulances, by improving existing prioritization practices (using predictive tools based on Artificial Intelligence) or by developing literacy and preventive skills in the population.

The remainder of this paper is structured as follows: Section 2 provides a comprehensive review of the relevant literature. Section 3 details the sample characteristics and empirical methodology. Section 4 presents the main results, followed by a discussion in Section 5.

2 Literature

First, this paper is greatly indebted to the vast literature that has uncovered the many biases in data analysis. Estimating the impact of time-varying exposure involving self-selection and heterogeneous benefits is a recurrent topic in empirical economics. This issue arises for example in the study of returns to schooling, where self-selection occurs on multiple levels based on expected returns and costs [Card, 2001]. In this context, thorough econometric analyses have disentangled the various mechanisms at play, revealing the true (structural) relationship between years of education and labour market outcomes (Angrist and Krueger [1991], among many others).

Secondly, this paper belongs to the set of health economics studies applying similar econometric techniques to data derived from both randomized controlled trials [Fransen et al., 2016, Goyal et al., 2016, Saver et al., 2016] and real-world data [Joundi et al., 2024, Al-Mufti et al., 2023, Alawieh et al., 2018, 2019, Mulder et al., 2018, Spiotta et al., 2013]. In these studies, time to treatment is significantly correlated with clinical outcomes and some medical research has also highlighted the inherent self-selection bias associated with the severity of the condition [Saver et al., 2013]. Although these studies have generally identified the direction of the effect Angrist et al. [2024] shows that a complementary data analysis using instrumental variables is relevant under a varying time of exposure.

More precisely, in the context of emergency care, this paper builds on the work of Wilde [2013], who found no statistically significant relationship between RT and mortality when the issue of endogeneity was not addressed. Two other studies emphasize the necessity of exploiting exogenous shocks to evaluate patients' sensitivity to delays. Bertoli and Grembi [2017] assess the effect of hospital proximity in emergency situations (road traffic accidents) by leveraging the exogenous variation in hospital proximity to cities, which is legally defined based on population size. Their results show that increasing the distance to the nearest hospital by 5 km raises the fatality rate by 13.84% at the sample average, corresponding to 0.92 additional deaths per 100 accidents. They also demonstrate that ordinary least squares and difference-in-differences estimates—commonly used approaches in the literature—tend to underestimate the true positive effect of hospital proximity on mortality. Similarly, Lucchese [2023] analyze patients who experienced a cardiac event in an Italian region in 2013 and 2014. By using hourly rainfall measurements at the time of the ambulance run as an instrument for RT, they find that its estimated effect on mortality approximately doubles compared to a naïve OLS estimation.

Compared to these studies, our paper offers two main advantages. First, our flexible estimation approach, which uses splines in a generalized additive model, allows for the estimation of the actual structural relationship between the two variables, interpreted as the average treatment effect (ATE), while the comparison with the average treatment effect on the treated (ATT) highlights policy implications related to these metrics. Second, our study employs a continuous utility index as the outcome variable, providing additional insightful results.

The heterogeneity of the effect has been studied in the literature: Jaldell et al. [2014] and Swan and Baumstark [2022] considered the severity of injuries (light or severe) and observed that the marginal effect of RT was greater for severe injuries than for mortality, using French and Thai data from rescue services, respectively. Additionally, Ma et al. [2019] provided evidence that the relationship between RT and health outcomes (mortality and morbidity measures) was non-monotonic. They suggest that the structural relationship between RT and outcomes is complex and that sophisticated econometric specifications are required to properly study these relationships. Building on this literature, our paper adds an analysis of injury severity while accounting for all unobserved factors associated with self-selection.

Lastly, this paper contributes to the growing economics literature that uses meteorological events as random exogenous shocks to identify treatment effects. Three weather variables are commonly used: rainfall, wind-speed and temperature. Rainfall has been used in several economic studies. Hodler and Raschky [2014] for instance find that economic shocks, indicated by night-time light intensity and instrumented by rainfall and drought, significantly increase the likelihood of civil conflicts. Miguel et al. [2004] is another example of how rainfall can be used as an instrument to measure the effect of economic activity on civil conflicts, showing that higher temperatures significantly increase their likelihood through reduced agricultural productivity. Similarly, wind-speed has been used in Bondy et al. [2020] to instrument pollution and assess its impact on health status and criminality. Atmospheric inversions have also been used as IVs for studying environmental outcomes, as in Sager [2019]. As for temperature variables, Deschênes and Greenstone [2007] analyse the effects of temperature on agricultural profits, using historical temperature variability as the instrument. Dell et al. [2012] use temperature as an exogenous variable to assess its impact on economic growth, finding that higher temperatures reduce output in poorer countries but have no significant impact in richer countries. Burke et al. [2015] evaluate temperature's role in explaining variations in conflict incidence among regions, showing that rising temperatures significantly increase interpersonal and intergroup conflict rates.

These studies also highlighted several limitations in using weather variables as IVs. For instance, the broad influence of temperature on various outcomes —such as social variables, labour productivity, conflict, health, migration, and institutional stability— challenges its ability to isolate a single causal effect, calling for a careful analysis of its validity in the specific context of its use. The main improvement of this paper regarding the use of an instrumental variable is the thorough analysis of the exclusion restriction hypothesis which is validated using a statistical test.

3 Data and Methods

3.1 Sample Description

3.1.1 Data sources

This study relies on two sources of data: the Endovascular Treatment in Ischemic Stroke Registry (ETIS) and weather observational data.

ETIS includes hospital stays between March 2017 and November 2023. It is a French multicentre observational study [1]. Although this database has been extensively used for clinical research and publications El Nawar et al. [2019], Douarinou et al. [2022], Bensoussan et al. [2023], Lambrou et al. [2024], to the best of our knowledge, no economic studies have yet been conducted using this rich material. It is the largest available database on AIS in France, including information on patients' RT, care pathways, and health outcomes.

The extraction of the database initially covers 19,884 hospital stays in 29 CSCs with MT performed by 186 different neurosurgeons. From this initial sample, 9,634 hospital

[1] Further information is available on this page: https://clinicaltrials.gov/study/NCT03776877

stays were excluded (detailed in Appendix A). Hospital stays with missing variables for RT (n=1,246) and the modified Rankin Scale (mRS) (n=5,916) had to be excluded for feasibility reasons. Stays in which MT was performed after 8 hours (not recommended) (n=1,465) and those involving in-hospital strokes (n=956) were also excluded, as they are very specific cases. Additionally, hospital stays of patients whose health-related quality of life improved after their stroke (n=51) were excluded for consistency reasons.

The final sample after data cleaning (detailed in Appendix A) contains 10,250 hospital stays. This filtering approach ensures a high-quality data set while preserving a sample size that is sufficient to maintain the statistical power of the results. Only two comorbidity variables (diabetes and high blood pressure) had a limited number of missing values (194 and 116 respectively). The missing values were replaced by imputing the most likely values, calculated using a logistic regression model with the other covariates as predictors.

The second source of data, the weather data, is publicly available[2]. It is collected at an hourly frequency through the extensive ground and altitude network of weather stations operated by the French National Weather Service (Météo-France). All weather observations —temperature, relative humidity, precipitation, pressure, cloud cover and wind-speed— are included in our database, although only temperature observations are used for the IV.

Each hospital stay is matched with the weather observations recorded at the station nearest to the hospitals where MT was performed at the time of the admission The maximum distance between the hospital and the nearest observation point is 1.3 km.

3.2 Variables

3.2.1 Outcome Variables

This study considers two outcome measures: mortality at 3 months post-stroke and utility losses.

Unlike mortality, utility losses are not directly observable in our data. Measures of pre-stroke and post-stroke utility were estimated from the reported pre-stroke and post-stroke levels of the modified Rankin Scale (mRS). The mRS is one of the most frequently used measures of health outcome in clinical trials involving AIS [Wilson et al., 2002], ranging from 0 to 6. Level 0 represents the absence of symptoms or limitations. Level 1 describes individuals who, despite minor symptoms (such as balance or speech difficulties), can perform all usual activities. Level 2 covers light disabilities, with no impairment for conducting daily activities independently but with limitations in social or work areas. Level 3 marks moderate disability, where some assistance is needed for household chores, though the person can walk without help. Level 4 represents moderately severe disability, where support is necessary for basic self-care activities. Level 5 is a severe disability, where the individual is bedridden, incontinent, and requires constant care from a caregiver, while Level 6 denotes death.

The pre-stroke level of mRS is calculated from patients' functional status before the AIS, as declared retrospectively by the patient or the family. The mRS post-stroke is collected three months after discharge by the medical team during a follow-up phone

[2]https://donneespubliques.meteofrance.fr/donnees_libres/Static/listeStations_Metro-OM_PackRadome.csv

call or medical visit. These two measures are then converted into utility scores using the mapping function of Rivero-Arias et al. [2009], with estimations using the French tariffs, described in Table 1. This mapping function was built from a dataset in which measures of mRS and answers to the EQ5D-5L questionnaire were collected at the same time for the same individuals three months after having an AIS.

Table 1: Mapping mRS - Utility scores

mRS score	**Utility Score**
0	0.942
1	0.867
2	0.67
3	0.413
4	0.104
5	-0.215

After converting the pre-stroke and the 3-month post-stroke mRS into utility levels, we calculate the utility loss (Y_L) for a given patient as a result of the stroke and the subsequent episode of care using the following formula:

$$Y_L = U_{3month} - U_{prestroke}$$

Table 2 shows the contingency table of pre AIS- and post AIS-mRS scores.

Table 2: Contingency table mRS pre AIS and mRS post AIS

mRS before stroke **mRS after stroke**	0	1	2	3	4
0	1169	-	-	-	-
1	1608	140	-	-	-
2	1216	134	82	-	-
3	1195	168	106	94	-
4	894	129	91	77	48
5	483	98	63	65	42
6	1435	331	256	236	89

Although considered continuous in the study, due to the mapping exercise, the variable for loss of utility takes 26 possible values as there were no patients with a pre-stroke mRS greater than 4 (moderately severe disability).

3.2.2 Covariates

We consider four types of control variables in our study :

- *Patient Characteristics.* Being diabetic and having high blood pressure are predictors of poor health outcomes after AIS. Age is also a good predictor of poor outcomes, partly due to the positive association found between age and multimorbidity.
- *Stroke Severity* The National Institute of Health Stroke Scale (NIHSS) is a measure of stroke severity, commonly used by healthcare providers to quantify the level of impairment caused by a stroke Originally designed for acute stroke trials, it is composed of 11 items, each of which scores a specific ability on a scale from 0 to 4. A score of 0 typically indicates normal function for that ability, while higher

scores indicate varying levels of impairments. Individual item scores are summed to calculate a patient's total NIHSS score, ranging from 0 to 42. This variable was collected in the dataset for medical purposes, ensuring high-quality measurement.

- *Patient pathway* Patients' mode of entry is recorded as being either direct or through another hospital (a patient was admitted to an initial hospital, then referred to a more specialised centre for MT). Direct admission is positively associated with better health outcomes and lower RT.
- *Hospital and year fixed effects* Hospitals fixed effects are used as well as years fixed effects, especially to ensure the exogeneity of the instrumental variable. In addition, these two sets of variables allow us to account for possible differences in technology across hospitals and years.

3.2.3 Treatment Variable

The treatment variable, RT, represents the time elapsed from symptom onset to the initiation of MT, measured in hours. Emergency services systematically record the date and time of symptoms' onset, as the French Health Authority guidelines recommend that certain treatments be administered only within a given time window, following symptoms' onset. As with the NIHSS, the exact time of MT for symptom onset and the initiation of MT are recorded for medical purposes, ensuring high-quality and reliable data collection.

3.2.4 Instrument

Air temperature, measured in degrees Celsius, is used as an instrument for the treatment variable. The justification for using this instrument is detailed in the following section. It is measured at the nearest weather station from the hospital facility at the time of hospital admission.

3.3 Empirical Setting

3.3.1 Rationale for the method

Understanding the data-generating process provides valuable intuition for the empirical strategy. Unlike many other conditions, stroke symptoms are typically identifiable with relative ease, although they can occasionally be confused with other pathologies. This specific nature of stroke symptoms allows for precise collection of the time of symptom onset. Thus, we assume that this measure of RT is observed with reasonable accuracy. The primary aim of this study is to examine the relationship between RT, the time to treatment, and Y, the post-stroke utility change (capturing the before-and-after difference in health outcomes). To ease the understanding of the data generating process, let's consider that utility loss is a function of two key elements: the severity (s) of and the response time (t)[3]. The exact shape of the relationship is unknown, so a simple and flexible representation would be that the utility loss for a given patient (y) can be any function of t and s:

$$y = h(t, s)$$

[3] t designates the time for a single individual between symptoms onset and in-hospital treatment, while we keep RT as the designation for the variable in our data. Following the literature (for example Florens et al. [2008], t refers to the potential treatment while RT is the observed (realised) treatment.

The objective of this paper is to estimate $\frac{\partial h(t,s)}{\partial t}$. Yet, there is reason to believe that t impacts s. First, it can be assumed that the more severe the symptoms, the easier to identify. Secondly, it is assumed that healthcare services prioritize the most severe cases of AIS. It seems reasonable to suggest that some individuals may exhibit a steeper slope in the relationship between response time and outcomes, a scenario commonly described in the literature as having a higher return on treatment and similarly recognised in the context of returns to schooling (e.g., Garen [1984]). As a consequence, one can assume that:

$$t = g_1(s) + D$$

where D accounts for all factors impacting t except the severity; the most important being the distance to the nearest facility but also some aspects of individual responsiveness that are independent of the severity. $g_1(.)$ is a decreasing function.

A last aspect should be considered for the picture to be complete, s is the true severity at the moment the AIS is discovered, while the observed severity (s_1) is recorded at hospital entry. For medical reasons, the severity of the conditions worsens in the absence of any treatment. As a consequence, if we define t_1 as the time elapsed before s_1 is recorded, and $g_2(.)$ the function modulating the s_1 as a consequence of t, we also have that:

$$s_1 = s + g_2(t_1)$$

As a consequence, the observed severity cannot be used as a relevant control for patient severity because it is likely impacted by t. In addition, as:

$$y = h\left(D + g_1(s), s\right)$$

estimating h using cross-section data requires an identification strategy which isolates some variations in D that are independent of s.

3.3.2 Identification

The identification strategy relies on an exogenous variation in response time induced by air temperature (Z). More precisely, it is assumed that $E(D|Z) \neq 0$. While air temperature itself is not the direct cause of increased RT during colder conditions, it correlates with various social behaviours, which are presented in the subsequent section. Consequently, in the data, it is possible to express RT as:

$$RT = E(RT|Z) + \eta_i$$

Where Z is expected to be independent of the severity, and any other possible unobserved variable and η_i contains the remaining variance, including the part associated with the severity. In addition, Z should not impact s by any means. The validity of air temperature as a relevant exogenous variation has occasionally been questioned in the literature [Schultz and Mankin, 2019]. The following paragraphs provide theoretical justifications and outline the tests to be conducted to ensure the robustness of the procedure in our context.[4]

[4]The CF framework is appealing here for its flexibility, allowing especially non-parametric estimation. In addition, Guo and Small [2016] show that, when the instrumental variables are valid, the control function method is more efficient than the usual two-stage least square, sometimes more than 10 times more efficient.

Relevance. Several justifications exist for the negative relationship between air temperature and RT. First, cold weather influences individual behaviour, such as spending less time outdoors and interacting less with others, which jeopardises early stroke detection. Secondly, hospital congestion, which is known to be greater during cold weather events (whether caused by cold-related illnesses or injuries due to weather conditions), slows the delivery of acute treatments necessary for timely stroke management, significantly impacting patient outcomes [Rizmie et al., 2022]. This slower responsiveness of healthcare services may result from reduced availability of emergency vehicles and phone operators. Third, in addition to slowing healthcare services, cold weather can disrupt transportation networks, delay access to care, and strain emergency response systems. Finally, stroke symptoms—which require immediate recognition and response—may be misinterpreted by patients due to environmental conditions. For example, heatstroke during a heatwave or hypothermia in freezing conditions might mask or mimic the clinical presentation of a stroke, delaying calls for emergency medical services (EMS) [Bakradze and Liberman, 2018]. Statistically speaking, the relevance of Z is measured by comparing a model predicting RT with all covariates and a model with the air temperature as an additional variable using an ANOVA test which provides a joint F-statistics for the additional predictors used in the regression model.

Independence. The independence hypothesis asserts that the unobserved components in both the first and second stages are unrelated to the instrument. One possible limitation to this hypothesis, identified in this study, relates to differences in geographical location (e.g., North vs. South or altitude), which can simultaneously impact RT and Z and also influence patient outcomes, potentially due to variations in the level of hospital specialization or the characteristics of the local population. Annual temperature variation is also a possible limitation to the independence of Z, as air temperature tends to increase from one year to another, while the average RT decreases. To address these potential issues, geographical variables capturing local characteristics (hospital dummies) and year dummies are included in the analysis. Although this test does not directly assess the independence of the instrument (a hypothesis required in the CF framework), a J-test is performed to test the exogeneity of the instrument in a 2SLS framework by assessing the correlation between the instrumented RT and the second-stage residual of a 2SLS.

Exclusion restriction. A growing literature is concerned with the violation of this hypothesis especially when using weather instruments. For example, Mellon [2024] review 195 papers using meteorological instrumental variables highlighting numerous cases of violation. More concerning for this study, recent articles published in medical journals argued that air temperature could directly impact the probability of AIS onset [Zhu et al., 2024] and the number of deaths by AIS per day [Alahmad et al., 2024]. While these papers identify the mechanisms coming from the probability of getting ill, some possible vectors also involve a greater severity of the disease and could lessen the quality of our empirical strategy. Therefore, the method developed by D'Haultfœuille et al. [2024] is applied to the data. It involves identifying segments of Z for which the instrument is considered "Locally Irrelevant" (*Assumption 4* in the paper). An instrument is locally irrelevant if it doesn't perfectly separate the population into two groups for which a stochastic dominance is observed in the distribution of the treatment. In other words, there exists a value of response time (RT) where the cumulative distribution functions (CDFs) conditional on the instrument are equal. At this precise point, Z assumes two distinct values (the average of each segment), yet x remains unchanged. The method then assesses whether the outcome variable demonstrates a significant change at this specific point. This change is defined as the Kolmogorov-Smirnov (KS) distance between the CDFs of the outcome, conditional on x and Z, for the given point and both segments. Since our instrument is continuous, but the test requires a binary variable, the test is repeated across different sub-samples of the database. The results provide the KS distance and a p-value derived from the method outlined in the paper. The code used to perform the test is available in Appendix C.

Although some indications are available about the structural form of the response curve - it is likely bounded, at worst by the death of the patient- , and it is also expected to be heterogeneous, with some patients benefiting more from a reduced response time - we do not no much about its actual shape. A simple specification might assume a linear relationship between patient outcomes and response time, but this could be misleading. In such a context, a partially non-parametric approach becomes appealing. This situation is typical in econometrics, where understanding the structural relationship between dependent variables is critical to address key questions, but there is no prior knowledge about that structure [Newey et al., 1999, Newey and Powell, 2003]. This approach differs from standard non-parametric regression because the goal is to estimate the structural model rather than merely the conditional expectation. It combines flexibility in estimation with the capacity to define interpretable parameters, which is particularly useful for tasks such as defining counterfactual scenarios.

The estimation procedure works as follows: first, the response time is regressed on the air temperature and all other control variables. In this first regression, the equation does not use the air temperature directly. Instead, we prefer using the natural spline of the temperature with 3 degrees of freedom (chosen by cross-validation). This method allows to account for the non-linear relationship between the air temperature and the RT: the relationship between the two variables is negative but the slope is steeper as the temperature becomes colder. Secondly, the residuals from the initial regression are incorporated as a variable in a subsequent regression. Unlike the standard control function (CF) approach, we do not rely on a simple ordinary least squares (OLS) method in this second stage of estimation. Instead, we employ a generalised additive model (GAM), which utilises a back-fitting interactive smoothing algorithm. A spline is chosen as the functional form for response time, allowing for a flexible representation of the structural relationship. This approach is advantageous because the actual structural form of the relationship between RT and Y is not predefined but is determined through the smoothing process (for details, see Hastie [1992]).

Furthermore, since the utility loss is bounded, we specify the a quasi-binomial link between the two variables. This choice is particularly suitable because it fixes a bounded support of the utility ensures that predicted values remain within the feasible range, thus preventing estimates from falling outside these limits [Gómez–Déniz et al., 2020]. For the estimation involving the binary outcome (mortality), a standard logit model is employed within the GAM framework, incorporating a binomial link function. The adoption of the quasibinomial link is logical, as it provides a coherent framework to understand both the binary relationship between survival and death, as well as the continuum of health states leading to the worst outcome, namely death. When using utility loss as the outcome, the estimation is extended to account for this continuum, enabling the representation of varying levels of disability status.

On average, the CF term – the residual from the first regression – controls for the function's slope. However, we would expect this slope to vary across individuals. To test this hypothesis, we introduce an interaction term between response time and the CF in the model. This term is also included non-linearly, but also estimated parametrically in robustness checks (see Appendix B). This approach, which is standard in control function methods (random correlated coefficient [Wooldridge, 2015]), allows for capturing subgroup effects in the presence of heterogeneity and also accounts for heteroscedasticity – situations where the bias increases or decreases with the treatment. In our case, we expect the difference between early and late entrants to widen over time, as low-severity patients may reach their maximum utility loss relatively quickly. In contrast, high-severity patients continue to experience utility losses, potentially leading to more significant utility reductions, including the ultimate loss associated with death.

The model estimated is:

$$\begin{cases} \log\left(\dfrac{Y}{1-Y}\right) = f_1(RT) + f_2(v) + f_3(v, RT) + X\beta_1 + u \\ \qquad\qquad\qquad\qquad\qquad\qquad RT = s(Z) + X\beta_2 + v \end{cases} \tag{1}$$

Where $f_1(.)$, $f_2(.)$ are splines of RT and v, and $f_3(.)$ is the tensor products of the two. These three functions are penalized splines so that their smoothness results from a tradeoff between predictability and degrees of freedom[5]. $s(.)$ is a natural spline with three degrees of freedom.

The ATE is obtained by comparing the value of $\hat{f}_1(.)$ at different points in time as follows: $\Delta^{ATE}(t_2 - t_1) = \hat{f}_1(t_2) - \hat{f}_1(t_1)$ or by linear approximation at a single point as $\Delta^{ATE}(t) = \frac{d}{dt}\hat{f}_1(t)$. Due to self-selection, one might expect the response time to be correlated with the return on such response time (e.g., marginal cost of time). People for whom the return to response time is higher will, on average, arrive earlier. As a consequence, the ATE is not necessarily the policy-relevant treatment effect as defined by [Heckman and Vytlacil, 2001]; in the context of our data, it is likely that the ATT is a better predictor of any policy aiming to reduce the time to treatment because it would impact individuals as they are currently sorted into the treatment. Under continuous treatment, Florens et al. [2008] suggest using the local average response parameter, which is the derivative with respect to the treatment variable and includes the random coefficient (e.g., the interaction between the treatment and the residual from the first stage). Following this suggestion, it is possible to define the ATT as $\Delta^{ATT}(t) = \frac{d}{dt}\hat{f}_1(t) + \frac{d}{dt}\hat{f}_3(t)$. Finally, the approach can also be extended by defining the ATE for population subgroups, where subgroups are defined with respect to the instrument, as shown by Wooldridge [2015], by setting the residuals to a specific value: $\Delta_g^{ATE} = \frac{d}{dt}\hat{f}_1(t) + \frac{d}{dt}\hat{f}_3(t,v)\Big|_{v=g}$.

[5] We chose to use the mgcv package in R for estimation, as it provides generalized additive modelling functions that are very similar to those of Hastie [1992], with some extensions. The main difference lies in the fact that mgcv is based on a penalized regression spline approach with automatic smoothness selection instead of the standard back-fitting algorithm.

The procedure's two stages are bootstrapped using the standard resampling method with one thousand replications to obtain robust confidence intervals. More precisely, as the estimation is non-parametric, each sample is randomly drawn before the first estimation and used to recover a prediction from the model fitted to this subsample. Once the 1,000 estimations are completed, the 0.975 and 0.025 quantiles are recovered to construct the confidence intervals.

Robustness checks are performed by testing alternative specifications for the model, and the results are presented in Appendix B. The most significant modifications introduced in these robustness checks include the parametric inclusion of v and $v \cdot RT$ as linear predictors. Additionally, $v \cdot RT^2$ was introduced to ensure that the shape of the curve is sensitive to the unobservable. A fully parametric approach was employed in another robustness check, incorporating RT, RT^2, and RT^3 as predictor variables.

4 Results

4.1 Descriptive statistics

4.1.1 Univariate Descriptive Statistics

Table 3: Descriptive Statistics

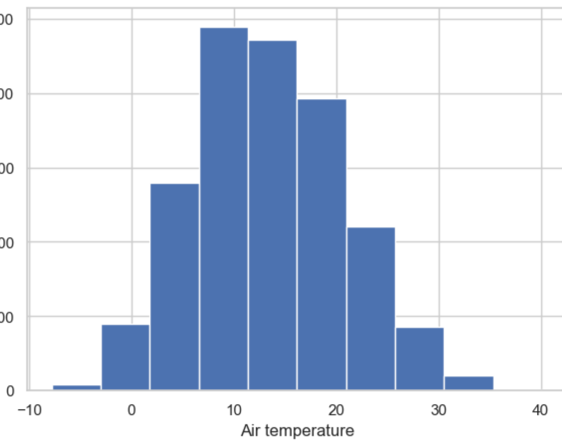
Sample size: 10,250 **N hospitals:** 29			**First observation** 2017.01.02 **Last observation** 2023.10.26			
Variable	**Mean**	**SD**	**Median**	**Min**	**Max**	**Skew**
RT	4.34	1.44	4.20	1.05	7.98	0.37
Utility loss	0.47	0.38	0.45	0.00	1.16	0.22
Mortality	0.23	0.42	0.00	0.00	1.00	1.30
Air temperature	13.39	7.30	12.84	-7.77	40.14	0.22
Diabetes	0.17	0.38	0.00	0.00	1.00	1.72
High blood pr.	0.60	0.49	1.00	0.00	1.00	-0.39
Age	70.76	14.73	73.00	17.00	102.00	-0.75
Mod. Admission	0.50	0.50	0.00	0.00	1.00	0.02
Severity	15.03	7.15	16.00	0.00	42.00	0.04

Table 3 provides univariate descriptive statistics. The final sample includes 10,250 observations.

The average utility loss is 0.47, with a maximum slightly above 1, which is possible because extreme disabilities (mRS 5) result in negative utility values. The rate of death following a stroke in this sample reaches 23%. Figure 2 provides a more detailed description of the two main variables distribution. The utility loss seems evenly distributed across patients except for no utility loss which is overrepresented in the distribution. The discrete nature of the variable from which utility scores are retrieved is visible in this data. The mean time to receive treatment for AIS is 4.33 hours (around 4 hours and 20 minutes), ranging from 1 hour to under 8 hours by study design (see section "Sample Description").

The air temperature instrument has an average of 13.4 °C, with a minimum of −7.8 °C and a maximum of 40.1 °C. The distribution of the instrument is normal with a mean of 13.40 °C, which is close to the average temperature in France during this period (13.8

Figure 1: Distribution of air temperature



°C)[6]

Regarding comorbidities and individual characteristics, we find that approximately 60% of individuals in the sample have high blood pressure, while 0.17% have diabetes. The mean age is 70.8 years, with the first quartile at 62 years and the third quartile at 82 years. Stroke severity, as measured by NIHSS, averages 15.4.

Figure 2: Distribution: Utility loss (left), Response Time in hours (right)

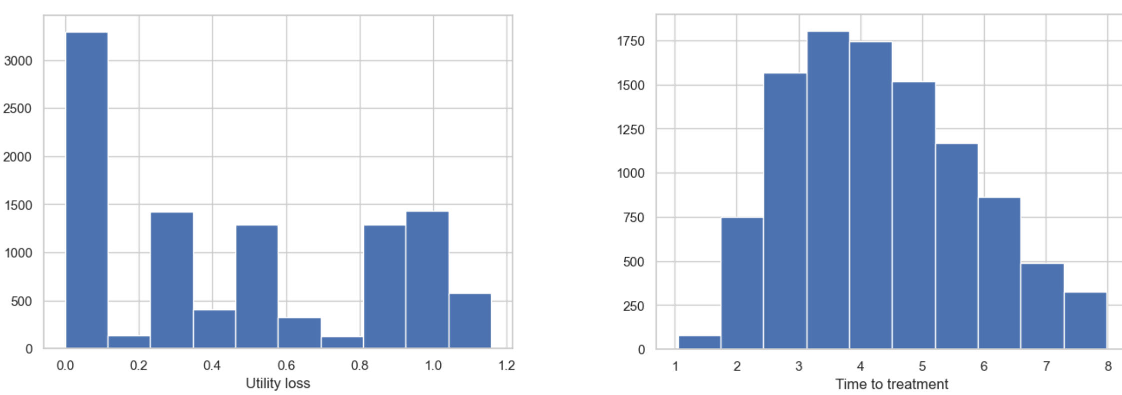


Table 4: Patients per year

2017	2018	2019	2020	2021	2022	2023
622	1053	1163	1783	2211	2268	1150

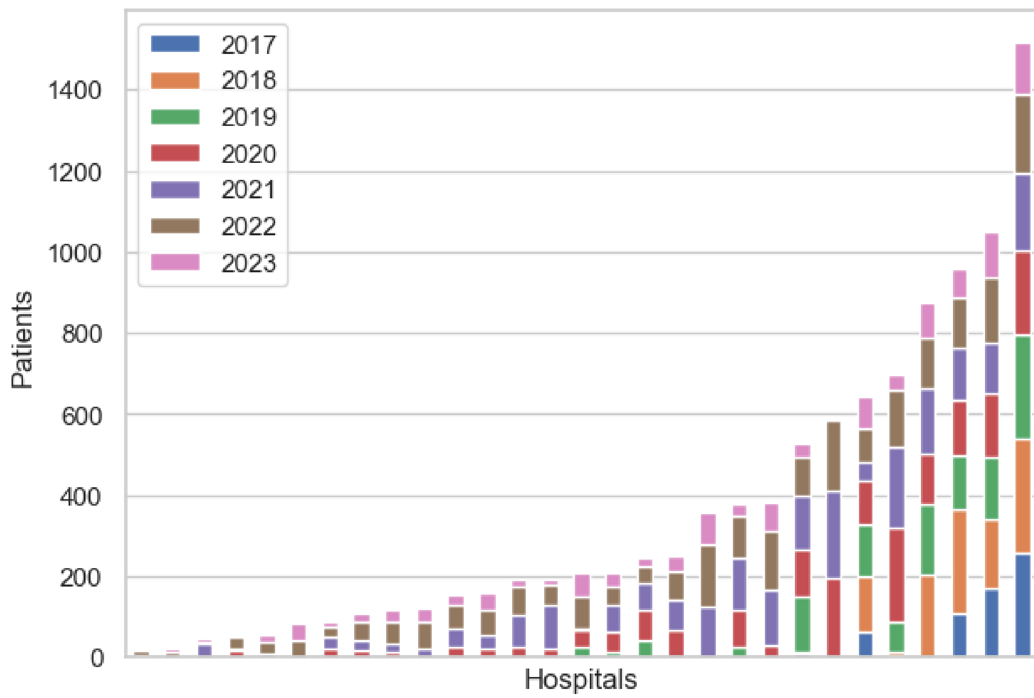
Patients are treated in 29 hospitals over seven years. The number of cases increases over time (Table 4), mostly due to the integration of new hospitals in the data collection program and the inclusion of more patients in each participating hospital. Overall around 50% of the patients were admitted through another hospital (5101) while the others were admitted directly (5200).

[6]https://meteofrance.fr/actualite/publications

4.1.2 Bivariate Descriptive Statistics

Among the 29 hospitals involved in the data collection, 5 provide half of the patients available in the database, both because they were involved earlier in the data collection and also because treat (or include) more patients per year (Figure 3). Figure 4a represents the relationship between RT and patient outcome. The figure on the top plots the relationship between RT and patient outcome, measured with a utility score. It shows an increase of around 0.20 unit of utility loss when RT increases by 4 hours and 30 minutes. Similarly, when removing patients who did not survive, the outcome increases by around 0.13. Eventually, the number of patients who did not survive the AIS increases over time: while patients arriving before three hours have more than 80% chances to survive, after five more hours it reaches 70%.

Figure 3: Distribution of Patients by Year and Hospitals

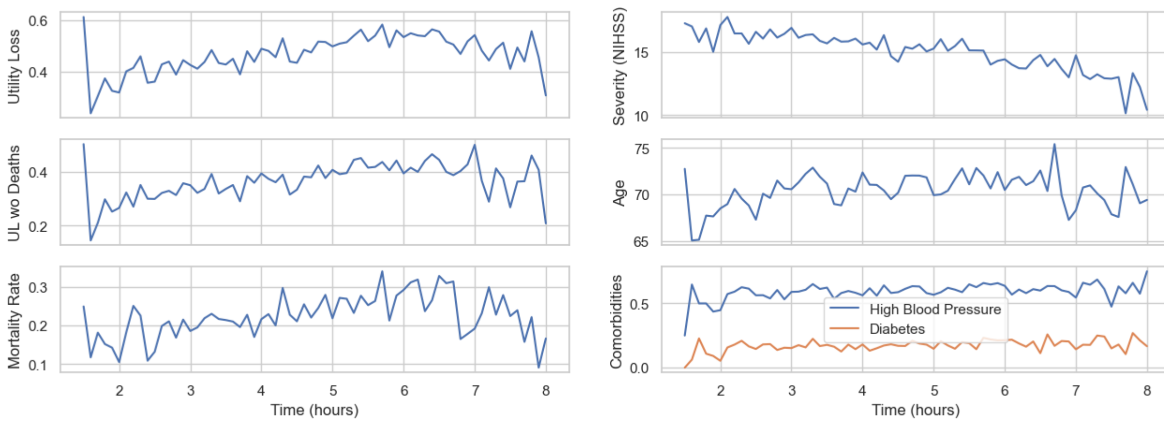


Surprisingly, the utility loss and the likelihood of death decrease after 6 hours, which is unlikely from a medical point of view and is explained by the sources of endogeneity detailed in Section 3.3.2.

The descriptive statistics support prior intuitions, showing that the drop in the outcome curve appears in both “utility with no death” and “death only” categories. This suggests that data selection effects and individual sorting likely occur simultaneously.

Figure 4b illustrates this composition bias using observables: a clear downward relationship emerges between stroke severity and RT, highlighting the link between disease severity and the responsiveness of both individuals and EMS. The second graph shows a slight increase in age with RT, followed by a drop after 6 to 7 hours. Finally, a small but positive correlation exists between RT and comorbidities, likely due to the spatial age distribution. These statistics highlight the correlation between RT and other observed confounders, suggesting that other unobserved confounders may introduce biases into

Figure 4: Comparison of Response Time Outcomes and Covariates



(a) Moving Average (range = 6 minutes): Outcomes Conditional on Observed Response Time

(b) Moving average: Rounded Response Time and Covariates

the relationship between RT and health outcomes.

Several complementary factors explain the pattern between observables and treatment. First, hospitals involved in data collection are located in cities with generally younger populations, meaning the age increase could reflect regional age distribution. Second, younger individuals tend to have less severe acute ischemic stroke (AIS), as indicated by a significant positive relationship between age and stroke severity, which may explain the overrepresentation of younger individuals after 6 to 7 hours.

4.2 Prediction of RT using the air temperature

Figure 5 presents $E(RT|Z)$, estimated using the natural spline of the air temperature with 3 degrees of freedom. The curve exhibits a steep slope until the temperature reaches 10 degrees whereas temperatures above 10 degrees impact less the RT. Overall, the average response time varies by more than 0.5 (30 minutes) when the temperature goes from its minimum to its maximum. To establish whether this is enough variation for Z to be valid as an instrument, we used joint F-statistics for the parameters of the instruments in this regression.

The number of degrees of freedom for the natural spline, chosen by cross-validation also offers the model with the straightest relationship between the air temperature and RT as detailed in table 5. The F-statistic of roughly 13.9 is enough to use this variable as an instrument. The p-value is highly significant indicating a strong statistical relationship between the included terms and the dependent variable in the first stage. This is promising for relevance. In addition, the difference in the sum of square between the two models confirms the explanatory power of the air temperature.

Especially because temperature (notably extreme heat) is known to cause emergency conditions, the exogeneity of the instrument would be questioned. Under a direct causal relationship between the air temperature and the unobserved variables of the structural equation, especially the severity, one could expect that the instrument introduces new

Figure 5: Natural spline regression of response time on air temperature

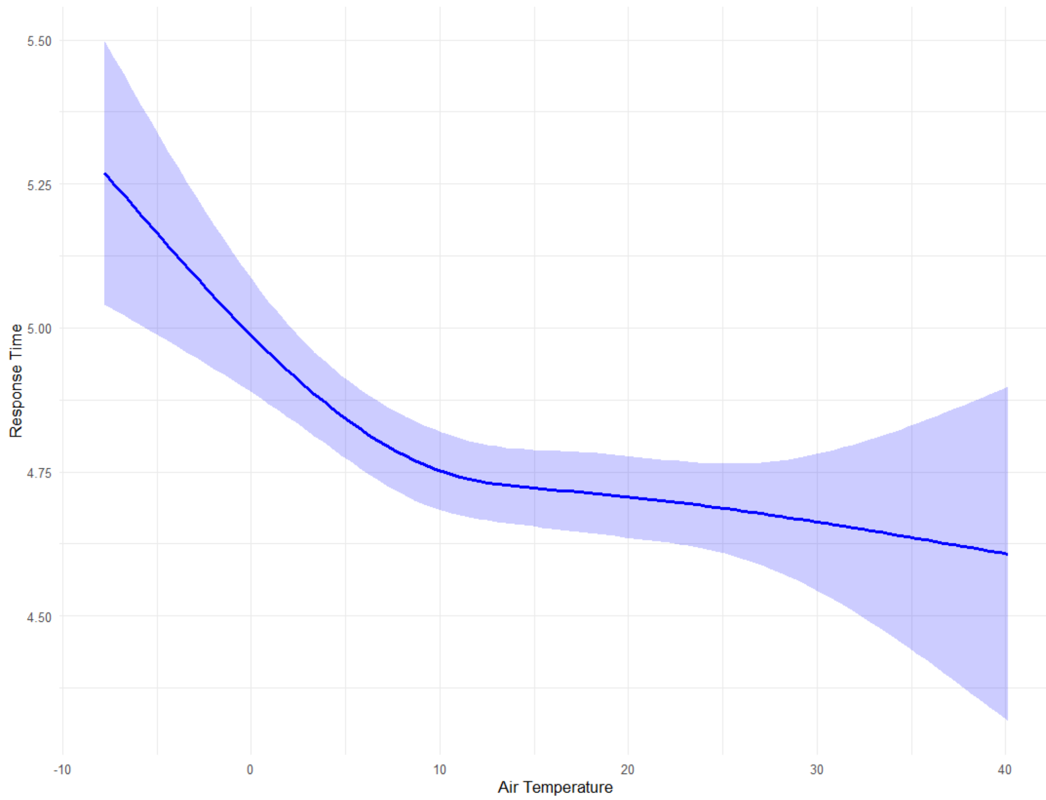


Table 5: Comparison of Model 1 (without Air Temperature) and Model 2 (with Air Temperature)

Model	**Res.Df**	**RSS**	**Df**	**Sum of Sq**	**F**	**p-value**
Model 1	10219	14650	-	-	-	
Model 2	10216	14590	3	59.464	13.879	4.989×10^{-9} ***

factors of endogeneity in the structural equation. We conducted a J-test, by testing the hypothesis that the instrument cannot predict the residuals of the reduced form. The F-statistic of the test is 0.3345 with a p-value of 0.8004. This result suggests that the instrument is exogenous. As we partially observed the severity of the condition in our data (although the measure is endogenous of time) it is possible to directly test the relationship between the instrument and one of the factors of endogeneity; results are in line with those of the J-test [see Appendix XXX TBC].

Eventually, the test developed by Guo and Small [2016] to compare the control function estimate to the standard 2SLS is used. It takes the form of a Hausman test comparing the coefficients of the 2SLS and the control function (estimated parametrically, see Appendix B. If the p-value of the test is less than 0.05, then there is evidence that the control function estimator is inconsistent and the usual two-stage least squares estimator should be used. Applying the methods to our data gives a test statistic of 1.087921 and a p-value of 0.2969323, supporting the use of the control function method.

The last aspect to be tested to ensure that the instrument chosen is valid is the fact that the instrument chosen has no direct impact on patient outcomes (exclusion restriction hypothesis). Table 6 presents the results of the test developed by D'Haultfœuille

Table 6: Test of exclusion restriction

Subset boundaries $Z \in$	]-2, 18 [	]2, 22[	]6, 26[	]10, 30[	]14, 34[	]18, 38[
Mean Z	10.12	12.27	14.56	17.14	19.98	22.69
x^*	2.75	2.62	2.55	2.57	2.5	2.5
KS statistic	3.34	3.93	3.01	1.86	2.27	2.76
p-value	0.35	0.3	0.5	0.87	0.72	0.55
N	7432	8414	8130	6622	4549	2774

Note: the table presents the results of the test developed by D'Haultfœuille et al. [2024]. H0: the instrument is valid (e.g. has no direct impact on the dependant variable). The test has been performed on several subsets of the dataset defined by the minimum and the maximum value of the air temperature. For each of these subsets, the air temperature is divided into two categories based on its mean value. x^* is the value of RT for which the instrument is locally irrelevant. The p-value gives the location of the KS-statistic in the bootstrap distribution of the estimator. See annexe [XX TBC] for more details about the method.

et al. [2024]. As the method requires a binary instrument, we performed several tests by choosing different thresholds for the air temperature to be binarised. Since the variable is strictly decreasing as shown in Figure 5, any threshold could work. One can see from Table 6 that the hypothesis of exclusion restriction seems to hold regardless of the threshold chosen: it is not possible to reject the exclusion restriction hypothesis at the 0.1 level in the different subsets used of the test. This means that the difference in the outcome induced by the instrument is not significantly different from zero. Therefore, it is possible to state, partially at odds with the medical literature - see section discussion - that the impact of extreme temperature is not significantly associated with patient outcomes except through the mean of the response time.

4.2.1 RT/outcome relationship

Figure 6: RT/Utility loss relationship

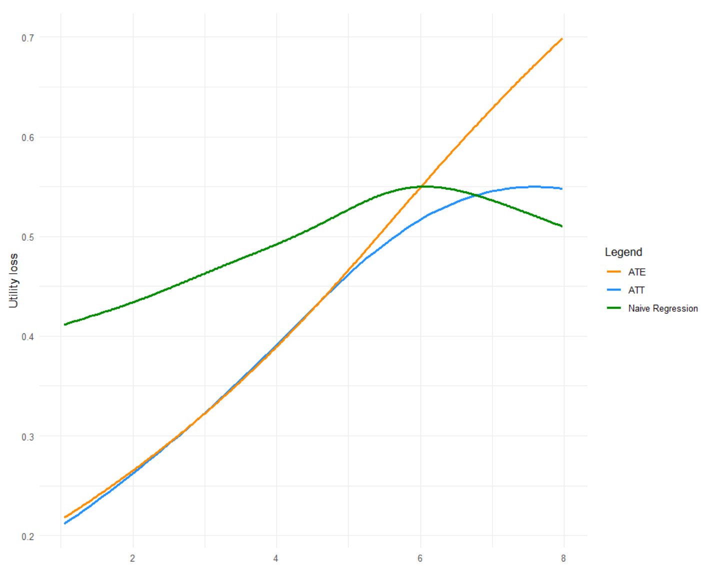


Table 6 presents the main result of this paper. In particular, the orange curve presents the average treatment effect. Its curve represents the utility loss for an additional delay before being treated in a CSC. This curve is almost a straight line, it exhibits a slope of 0.05 to 0.09 additional utility loss for an additional hour before treatment. It is interesting to compare the ATE to the results of the naive regression. Because of patient sorting, the curve of the naive regression exhibits a much less steep slope. The curve even decreases after 6 hours because the effect of patient sorting becomes larger than the effect of response time.

The ATT completes the picture as it provides an average treatment effect conditional on the type of patients treated on average after a certain response time. The slope of this curve should be understood as the sensibility to RT of patients treated at time t. Unsurprisingly, the slope of the curves starts to flatten when the naive regression begins to decrease: patients treated after five to six hours are less sensitive to RT, these results are consistent with the intuition regarding patient sorting. In contrast, the ATT and the ATE are quite similar under five hours whereas one could also have expected that patients arriving earlier have a greater time sensibility - even if the ATT line has a slighter steeper slope as predicted.

These comparisons are informative regarding the mechanism of selection at stake. While selection mostly occurs on levels before five hours - which is in line with a selection based on the evidentness of the symptoms - it occurs also on trend after five hours, where the ATT begins to flatten.

Figure 7: ATE comparison: early and late entrants

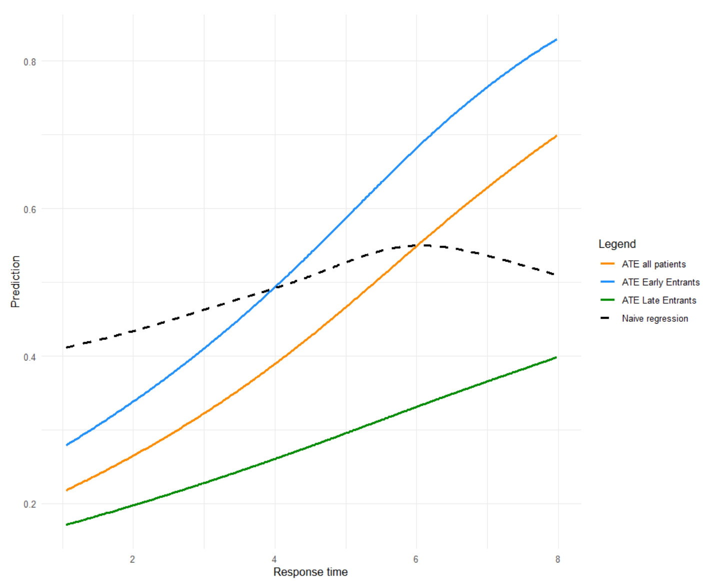
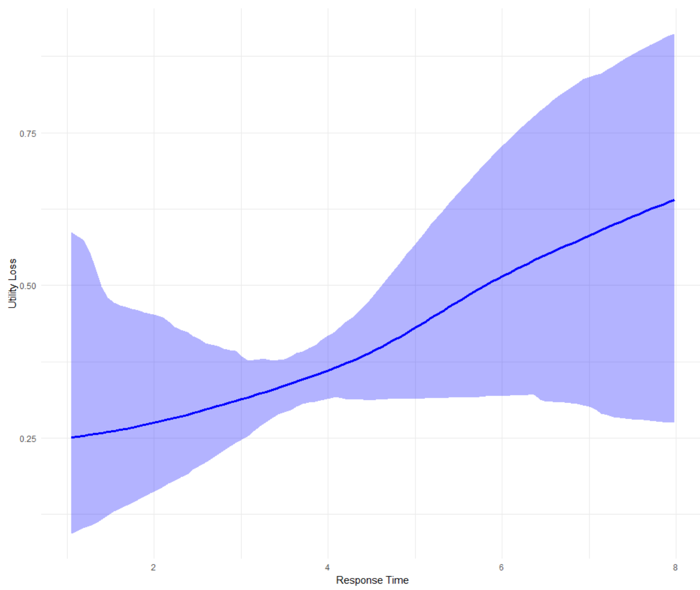


Figure 7 illustrates the subgroup ATEs, with groups categorised as either late or early entrants. Late entrants are those who, on average, arrived at the hospital later than the mean defined by the instrument, while early entrants arrived earlier than this mean. The figure uses the 0.01 and 0.99 quantiles of the first-stage equation residuals to define these groups. Due to patient prioritisation, late entrants can be interpreted as representing the least severe cases, while early entrants correspond to the most severe.

The figure also displays the overall ATE and the results of a naive regression to facilitate comparison. The findings align with those presented in Figure 6: the overall ATE and ATE_{early} exhibit a similar average slope, while the ATE_{late} is considerably less steep. Comparing the ATE_{late} with the naive regression provides a revealing insight:

Figure 8: RT/Utility loss relationship with 95% CI



the similarity in their slopes suggests that, due to patient prioritisation, the observed utility loss aligns closely with that of late entrants, who are typically the least severe patients.

Figure 9: RT/mortality relationship

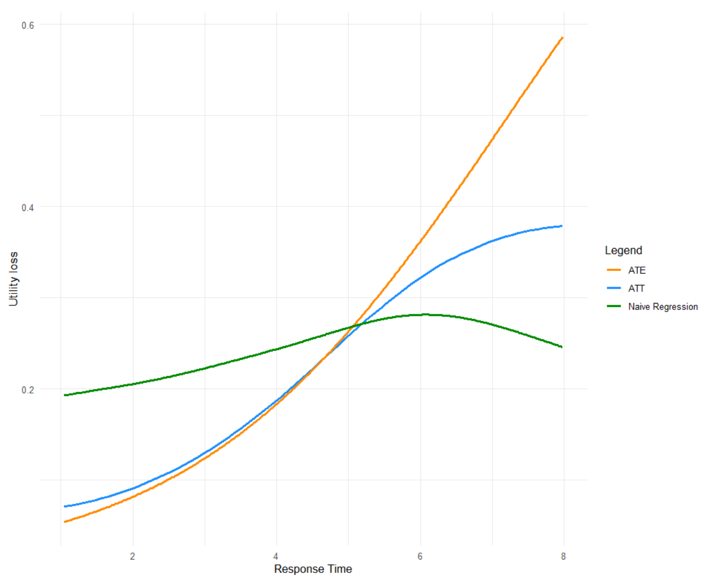
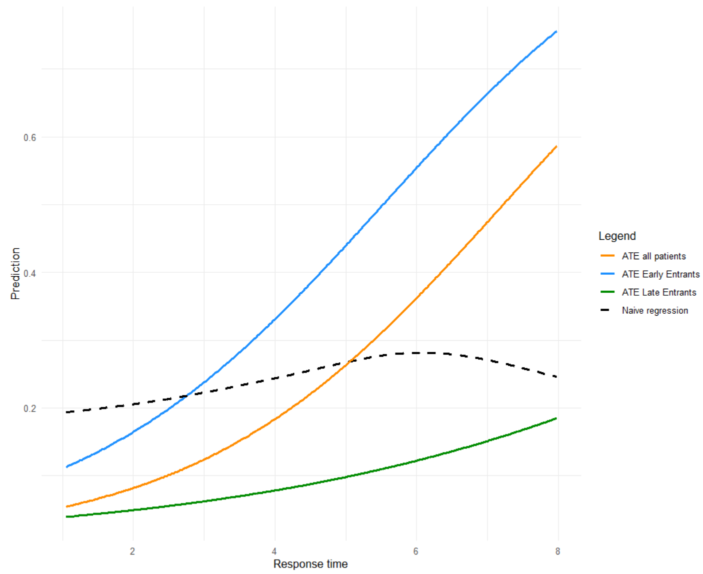


Figure 8 presents the inference investigation for the ATE, using utility loss as the outcome. The confidence interval is quite large due to the two-step procedure and the heterogeneity of the impact.

Figures 9 and 10 present the same results using mortality as the outcome. These results are similar to those for utility loss. The lines are steeper in the middle of the graph, reflecting the binary nature of the variable. The interpretation of the different lines is similar.

Figure 10: ATE mortality comparison: early and late entrants



5 Discussion and Conclusion

This study revisits the relationship between response time and patients' health outcomes in emergency care, focusing on severe AIS patients in France. Response time is defined as the time elapsed from symptoms' onset to treatment, and covers both demand-side determinants (patients' ability to seek emergency care in due time) and supply-side determinants (emergency care operational processes and physical distance). It uses a large, detailed and robust dataset for an extensive period of time. The study relies on an innovative empirical strategy to overcome two biases which have affected previous results: severity is frequently measured imperfectly in available data, resulting in omitted variable bias that underestimates the impact of RT on health outcomes; hospital data about acute conditions face a censoring issue as some individuals do not reach the emergency department before dying. The instrumented estimation results show that naïve assessments have strongly underestimated the impact of response time on patients' health outcomes, calling for corrective emergency care policies.

Subsequently, this paper demonstrates that relying solely on existing medical scales to control for observed severity is insufficient to ensure a robust ceteris paribus analysis. These scales are inherently correlated with both the time taken to access healthcare services and health outcomes. Intuitively, for any given patient, an increased response time worsens their condition. This dual correlation introduces additional endogeneity challenges, for which an explanation is provided in this paper: while there is a negative relationship between the time of arrival and severity, the relationship between response time and severity is actually positive. A causal link between treatment delay and health outcomes is identified, alongside a negative association between severity and the likelihood of reaching healthcare services promptly. These findings can be generalised to other conditions where severity is measured at the point of entry into the healthcare system.

Building on this new evidence, and considering both supply and demand determinants, healthcare policies aimed at reducing response time are needed that extend beyond geographical adjustments in the distribution of emergency care. Focusing on patient-centred care, streamlining operational processes within EMS, and strategically allocating resources can provide a solid foundation for sustainable improvements in emergency care delivery. Key actions include educating patients, enhancing EMS processes and improving the spatial distribution of emergency care, considering innovations such mobile care units and Point-of-Care solutions to ensure that specialized care units are available in underserved areas. This comprehensive strategy emphasizes equity and scalability, ensuring that the benefits reach both urban centres and remote communities. Such a balanced focus not only strengthens public health infrastructure but also maximizes the impact of investments, paving the way for a more efficient and inclusive healthcare system.

A Data cleaning process

Database extraction: 19,884 hospital stays in 29 CSCs
MT performed by 186 neurosurgeons

↓

Documented time for both
stroke onset and groin puncture
(18,638 cases remain)

↓

Recorded mRS scores both
pre-stroke and 3 months post-stroke
(12,722 cases remain)

↓

Onset-to-groin puncture time <8 hours
(11,257 cases remain) (Note: Delays
> 8 hours are too rare for conclusions)

↓

Exclusion of in-hospital strokes
(Time recording differs)
(10,301 cases remain)

↓

Exclusion of patients whose
health improved post-AIS
(Recording errors likely)
(Final dataset: 10,250 cases)

↓

Final dataset:
10,250 cases for analysis

B Results using a two-step parametric estimation

B.1 First stage estimation (parametric polynomial)

This appendix presents the prediction of RT by the air temperature. After conditioning on hospitals and year fixed-effects, in addition to covariates. Air temperature appears to be negatively associated with RT, with the slope being steeper for cold weather. While the square of the air temperature is positively associated with RT, suggesting a U-shaped relationship between the two variables, reaching a minimum when the temperature is 23.2 Celsius degrees.

Table 7: First Stage Ordinary Least Square Estimation

Dep. variable: **Response time**	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
Air temperature	-0.0205626	0.0061164	-3.362	0.000777
Air temperature squared	0.0004378	0.0002044	2.141	0.032261
High blood pressure	0.0321	0.0267	1.201	0.2299
Diabetes	0.0703	0.0323	2.176	0.0296
Age	0.0025	0.0009	2.836	0.0046
Mode of admission	1.5681	0.0249	63.099	$< 2 \times 10^{-16}$
Regression includes year and hospital fixed effects.				
Residual standard error: 1.393 on 10209 degrees of freedom				
Multiple R-squared: 0.9075, Adjusted R-squared: 0.9072				
F-statistic: 2444 on 41 and 10209 DF, p-value: ¡ 2.2e-16				

B.2 Reduced form (parametric polynomial)

Table 8 illustrate the utility loss associated with an additional hour of RT delay. The relationship between RT and utility loss is concave. In the naive model, utility loss peaks at an RT value of 6.7, confirming the descriptive relationship between these variables after accounting for covariates. In contrast, the control function model predicts the maximum utility loss at an RT value of 12.5, which lies outside the data's observed range. These findings align with prior expectations regarding the direction of the bias. Figure 11 compares the estimation provided by the two regressions showing a steeper curve after using the instrument.

The results presented in table 9 illustrate the probability of death associated with an additional hour of RT delay. The relationship between RT and the probability of death is concave, as for the utility loss. The parameters of the logit model show, again, a steepening of the curve once an instrument is used in our control function setup. Results are more easily sizable in figure 12: while the non-instrumented curve shows a probability of death of 25%, it is around 40% when using the control function after 6 hours from symtpoms onset. The results suggest that reducing the time to treatment by one hour for patients arriving in four hours may reduce the probability of death by 10%. These findings are also in line with prior expectations regarding the direction of the bias.

Table 8: Utility Loss Regression Results: Naive vs. Control Function

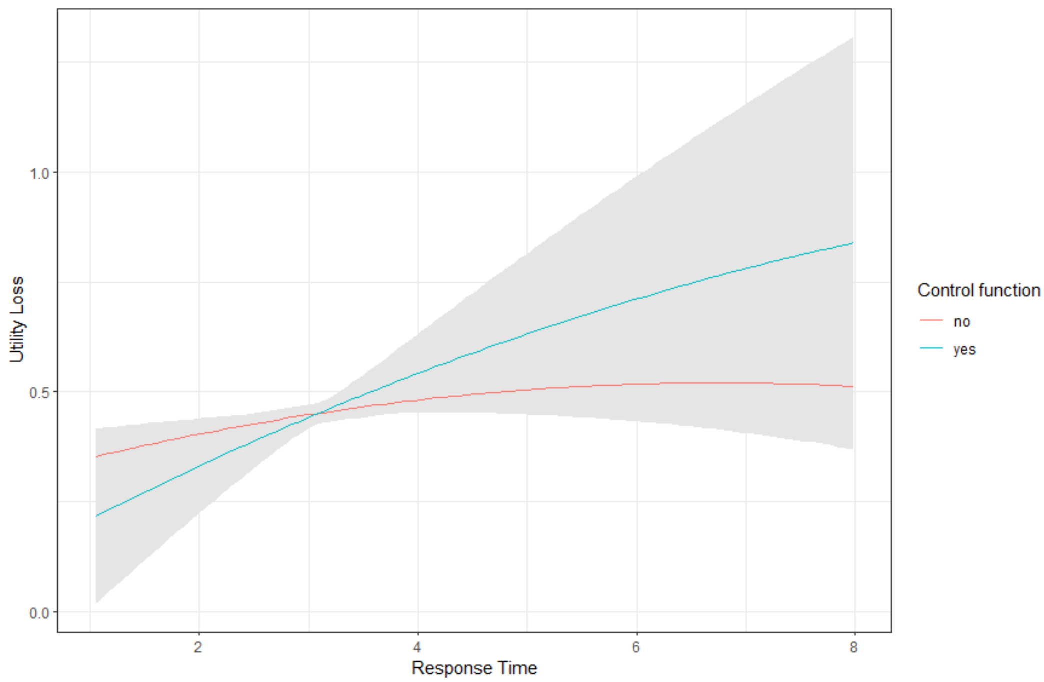
	Naive Model				**Control Function Model**			
	Est.	SE	t	*P*	Est.	SE	t	*P*
Response time	0.0701	0.0150	4.66	3.1e−6	0.1323	0.0514	2.57	0.010
Resp. time^2	−0.0052	0.0016	−3.32	0.0009	−0.0052	0.0016	−3.30	0.0010
High BP	0.0212	0.0081	2.62	0.0087	0.0192	0.0082	2.33	0.020
Diabetes	0.0750	0.0095	7.92	2.6e−15	0.0711	0.0099	7.15	9.1e−13
Age	0.0054	0.0003	20.5	< 2e−16	0.0053	0.0003	18.2	< 2e−16
Mode adm.	0.0220	0.0090	2.44	0.015	−0.0757	0.0776	−0.98	0.330
Res. 1st stage	—	—	—	—	−0.0628	0.0496	−1.27	0.206

Naive Model
RSE: 0.3632 (10,253 df)
R^2: 0.638, Adj. R^2: 0.637
F-stat: 440.8 ($p < 2.2e{-16}$)

Control Function Model
RSE: 0.3632 (10,252 df)
R^2: 0.6381, Adj. R^2: 0.637
F-stat: 430.3 ($p < 2.2e{-16}$)

Both regressions include year and hospital fixed effects.

Figure 11: Average utility loss by response time



C Python code for the method of D'Haultfœuille et al. [2024]

```python

import numpy as np
from scipy.stats import norm

def kernel(u):
    """Kernel function."""
    return 0.75 * (1 - u**2) * ((u > -1) & (u < 1))

def bandwidth(y, X, Z, numboot=1000):
    """
    Test and Relax the Exclusion Restriction in the Control Function Approach.
```

Table 9: Binomial Regression: Probability of Death as a Function of Response Time

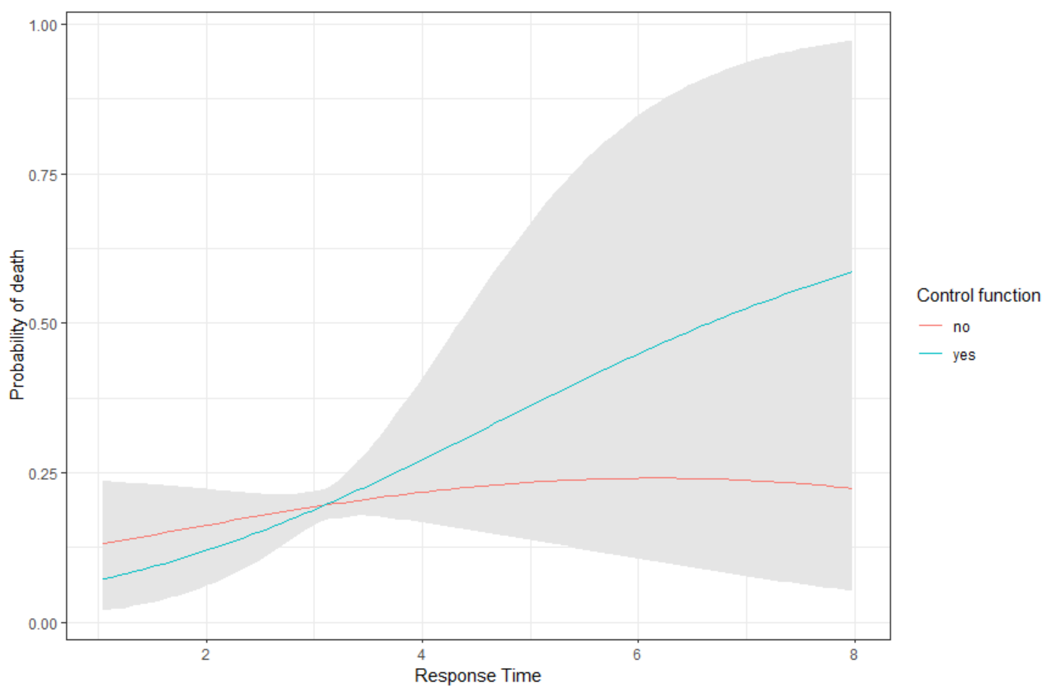
	Naive model				Control Function			
	Est.	SE	z	P	Est.	SE	z	P
Intercept	−5.73	0.37	−15.49	< 2e−16	−6.58	1.13	−5.82	5.8e−9
Resp. time	0.35	0.11	3.28	0.001	0.62	0.36	1.75	0.08
Resp. time^2	−0.03	0.01	−2.56	0.01	−0.03	0.01	−2.55	0.01
High BP	0.11	0.06	1.99	0.046	0.10	0.06	1.80	0.072
Diabetes	0.68	0.06	11.51	< 2e−16	0.67	0.06	10.59	< 2e−16
Age	0.045	0.002	20.86	< 2e−16	0.045	0.002	19.36	< 2e−16
Mode adm.	0.02	0.06	0.37	0.71	−0.40	0.54	−0.75	0.45
Res. (iv)	—	—	—	—	−0.27	0.34	−0.80	0.42

Naive model
Deviance: 9942.4 (10,253 df)
AIC: 10024

Control Function
Deviance: 9941.7 (10,252 df)
AIC: 10026

Both regressions include year and hospital fixed effects.

Figure 12: Probability of death by response time




```
    Parameters:
        y : ndarray
            Dependent variable.
        xz : ndarray
            Independent variables (X and Z as columns).
        numboot : int
            Number of bootstrap samples.
    Returns:
        results : dict
            KS statistic, p-value, bandwidth, and additional outputs.
    """
    n = len(y)
```

```
    x = X
    z = Z
    mean_z = np.mean(z)
    z = (z > np.mean(z)).astype(int)  # Threshold Z
    mean_z = np.mean(z)
    # Generate xlist and ylist
    x1, x2 = np.quantile(x, [0.01, 0.99])
    xlist = np.linspace(x1, x2, 1001)

    y1, y2 = np.quantile(y, [0.01, 0.99])
    ylist = np.linspace(y1, y2, 1001)

    # Compute FXZ0 and FXZ1
    fxz0 = np.array([np.mean((x <= x_star) & (z == 0)) for x_star in xlist])
    fxz1 = np.array([np.mean((x <= x_star) & (z == 1)) for x_star in xlist])

    xstar_idx = np.argmin(np.abs(fxz1 - fxz0))
    xstar = xlist[xstar_idx]

    if abs(fxz1[xstar_idx] - fxz0[xstar_idx]) > 0.01:
        print("Warning: No CDF crossing at the tolerance level of 1%.")

    # Bandwidth selection
    hlist = np.arange(1, 21) / 10 * np.std(x)
    medy = np.median(y)
    cv_sse = []
    print(hlist)
    for h in hlist:
        errors = []
        for i in range(n):
            others = np.delete(np.arange(n), i)
            kout = kernel((x[others] - x[i]) / h)
            prediction = np.mean((y[others] <= medy) * kout * (z[others] == z[i])) / \
                             (np.mean(kout * (z[others] == z[i])) + 1e-6)
            errors.append((int(y[i] <= medy) - prediction) ** 2)
        cv_sse.append(np.sum(errors))

    return hlist[np.argmin(cv_sse)] * (n ** (1 / 5 - 5 / 12))

def test_exclusion(y, X, Z, h,  numboot=1000):
    n = len(y)
    x = X
    z = Z
    mean_z = np.mean(z)
    z = (z > np.mean(z)).astype(int)  # Threshold Z


    # Generate xlist and ylist
    x1, x2 = np.quantile(x, [0.1, 0.9])
    xlist = np.linspace(x1, x2, 1001)

    y1, y2 = np.quantile(y, [0.1, 0.9])
    ylist = np.linspace(y1, y2, 1001)

    # Compute FXZ0 and FXZ1
    fxz0 = np.array([np.mean((x <= x_star) & (z == 0)) for x_star in xlist])
    fxz1 = np.array([np.mean((x <= x_star) & (z == 1)) for x_star in xlist])

    xstar_idx = np.argmin(np.abs(fxz1 - fxz0))
    xstar = xlist[xstar_idx]

```

```
    if abs(fxz1[xstar_idx] - fxz0[xstar_idx]) > 0.01:
        print("Warning: No CDF crossing at the tolerance level of 1%.")
        message = "Warning: No CDF crossing at the tolerance level of 1%."
    else :
        print("CDF crossing ok")
        message = "CDF crossing ok"
    h = h

    # Compute FYXZ0, FYXZ1, and KS statistic
    fyxz0, fyxz1 = [], []

    kout = kernel((x - xstar) / h)

    mean_x_z0 = np.mean(x*kout*(z==0))
    mean_x_z1 = np.mean(x*kout*(z==1))

    mean_y_z0 = np.mean(y*kout*(z==0))
    mean_y_z1 = np.mean(y*kout*(z==1))

    for yval in ylist:
        fyxz0.append(np.mean((y <= yval) * kout * (z == 0)) / (np.mean(kout *
            (z == 0)) + 1e-6))
        fyxz1.append(np.mean((y <= yval) * kout * (z == 1)) / (np.mean(kout *
            (z == 1)) + 1e-6))

    fyxz0, fyxz1 = np.array(fyxz0), np.array(fyxz1)
    ks_statistic = np.sqrt(n * h) * np.max(np.abs(fyxz1 - fyxz0))

    # Multiplier bootstrap
    ifz0 = np.zeros((n, len(ylist)))
    ifz1 = np.zeros((n, len(ylist)))

    x_indices = np.array([np.argmin(np.abs(x_val - xlist)) for x_val in x])

    #print(fyxz0[x_indices])

    for i, yval in enumerate(ylist):

        ifz0[:, i] = (np.sqrt(n * h) * ((y <= yval) - fyxz0[x_indices]) * kout
             * (z == 0)) / \
                     (np.sum(kout * (z == 0)) + 1e-6/n)
        #plt.hist(ifz0[:, i])
        #plt.show()
        ifz1[:, i] = (np.sqrt(n * h) * ((y <= yval) - fyxz1[x_indices]) * kout
             * (z == 1)) / \
                     (np.sum(kout * (z == 1)) + 1e-6/n)
    #print(ifz0)
    mb = [np.max(np.abs(np.sum((np.random.normal(size=n)[:, None]) * (ifz1 -
        ifz0), axis=0))) for _ in range(numboot)]
    #print(mb)
    #plt.hist(mb)
    p_value = np.mean(np.array(mb) > ks_statistic)

    # Results
    results = {
        'KS_statistic': ks_statistic,
        'p_value': p_value,
        'bandwidth': h,
        'num_bootstrap': numboot,
        'n': n,
        'xstar': xstar,
        'CI+' : np.quantile(mb, 0.95),
```

```
        'message' : message,
        'mean_z' : mean_z,
        'mean_var' : [mean_x_z0, mean_x_z1, mean_y_z0, mean_y_z1]
    }
    return results

--------------------------------------------------------------------------------

bandwidth(df.Y.reset_index(drop=True),
          df.T.reset_index(drop=True),
          df.Z.reset_index(drop=True), numboot=1000)

--------------------------------------------------------------------------------

RES = []
for t in [-2, 2, 6, 10, 14, 18] :
    res = test_exclusion(df[(df.temperature_2m>t) & (df.temperature_2m<t+20)].ut_loss.reset_index(drop=True),
                   df[(df.temperature_2m>t) & (df.temperature_2m<t+20)].OTT.reset_index(drop=True),
                   df[(df.temperature_2m>t) & (df.temperature_2m<t+20)].temperature_2m.reset_index(drop=True),
                   0.15586483562757023,
                   numboot=1000)
    RES.append(res)
    #print(res)

dviz = pd.DataFrame({'Mean Z' : [RES[i]['mean_z'] for i in range(len(RES))],
            'xstar' : [RES[i]['xstar'] for i in range(len(RES))],
            'KS statistic' : [RES[i]['KS_statistic'] for i in range(len(RES))],
            'p-value' : [RES[i]['p_value'] for i in range(len(RES))], }).round(2).astype(str).set_index('Mean Z').T

dviz = pd.DataFrame({'N' : [RES[i]['n'] for i in range(len(RES))],
            'Mean Z' : [RES[i]['mean_z'] for i in range(len(RES))],
            'xstar' : [RES[i]['xstar'] for i in range(len(RES))],
            'KS statistic' : [RES[i]['KS_statistic'] for i in range(len(RES))],
            'p-value' : [RES[i]['p_value'] for i in range(len(RES))], }).round(2).astype(str).set_index('Mean Z').T
```

References

Fawaz Al-Mufti, Mohamed Elfil, Hazem S Ghaith, Sherief Ghozy, Ahmed Elmashad, Ashutosh P Jadhav, Chirag D Gandhi, and Stephan Mayer. Time-to-treatment with endovascular thrombectomy in patients with large core ischemic stroke: the 'late window paradox'. *Journal of NeuroInterventional Surgery*, 15(8):733–734, May 2023. ISSN 1759-8486. doi: 10.1136/jnis-2023-020493. URL http://dx.doi.org/10.1136/jnis-2023-020493.

Barrak Alahmad, Haitham Khraishah, Meghana Kamineni, Dominic Royé, Stefania I. Papatheodorou, Ana Maria Vicedo-Cabrera, Yuming Guo, Eric Lavigne, Ben Armstrong, Francesco Sera, Aaron S. Bernstein, Antonella Zanobetti, Eric Garshick, Joel Schwartz, Michelle L. Bell, Fahd Al-Mulla, Petros Koutrakis, Antonio Gasparrini, Achilleos Souzana, Fiorella Acquaotta, Shih-Chun Pan, Micheline Sousa Zanotti Stagliorio Coelho, Valentina Colistro, Tran Ngoc Dang, Do Van Dung, Francesca K. De' Donato, Alireza Entezari, Yue-Liang Leon Guo, Masahiro Hashizume, Yasushi Honda, Ene Indermitte, Carmen Íñiguez, Jouni J.K. Jaakkola, Ho Kim, Whanhee Lee, Shanshan Li, Joana Madureira, Fatemeh Mayvaneh, Hans Orru, Ala Overcenco, Martina S. Ragettli, Niilo R.I. Ryti, Paulo Hilario Nascimento Saldiva, Noah Scovronick, Xerxes Seposo, Susana Pereira Silva, Massimo Stafoggia, and Aurelio Tobias. Extreme temperatures and stroke mortality: Evidence from a multi-country analysis. *Stroke*, 55(7):1847–1856, July 2024. ISSN 1524-4628. doi: 10.1161/strokeaha.123.045751. URL http://dx.doi.org/10.1161/STROKEAHA.123.045751.

Ali Alawieh, Jing Zhao, and Wuwei Feng. Factors affecting post-stroke motor recovery: Implications on neurotherapy after brain injury. *Behavioural Brain Research*, 340: 94–101, March 2018. ISSN 0166-4328. doi: 10.1016/j.bbr.2016.08.029. URL http://dx.doi.org/10.1016/j.bbr.2016.08.029.

Ali Alawieh, Jan Vargas, Kyle M. Fargen, E. Farris Langley, Robert M. Starke, Reade De Leacy, Rano Chatterjee, Ansaar Rai, Travis Dumont, Peter Kan, David McCarthy, Fábio A. Nascimento, Jasmeet Singh, Lukas Vilella, Aquilla Turk, and Alejandro M. Spiotta. Impact of procedure time on outcomes of thrombectomy for stroke. *Journal of the American College of Cardiology*, 73(8):879–890, March 2019. ISSN 0735-1097. doi: 10.1016/j.jacc.2018.11.052. URL http://dx.doi.org/10.1016/j.jacc.2018.11.052.

Joshua Angrist, Bruno Ferman, Carol Gao, Peter Hull, Otavio L. Tecchio, and Robert W. Yeh. Instrumental Variables with Time-Varying Exposure: New Estimates of Revascularization Effects on Quality of Life, December 2024. URL https://www.nber.org/papers/w33296.

Joshua D. Angrist and Alan B. Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, November 1991. ISSN 0033-5533. doi: 10.2307/2937954. URL http://dx.doi.org/10.2307/2937954.

Ekaterina Bakradze and Ava L. Liberman. Diagnostic error in stroke—reasons and proposed solutions. *Current Atherosclerosis Reports*, 20(2), February 2018. ISSN 1534-6242. doi: 10.1007/s11883-018-0712-3. URL http://dx.doi.org/10.1007/s11883-018-0712-3.

G. S. Becker. Health as human capital: synthesis and extensions. *Oxford Economic Papers*, 59(3):379–410, April 2007. ISSN 1464-3812. doi: 10.1093/oep/gpm020. URL http://dx.doi.org/10.1093/oep/gpm020.

Sacha Bensoussan, Stephanos Nikolaos Finitsis, Bertrand Lapergue, Gaultier Marnat, Igor Sibon, Solène Moulin, Jean-Marc Olivot, Sébastien Richard, Charlotte Rosso, Benjamin Gory, and Frédéric Clarençon. Endovascular treatment of distal-m2 segment occlusions: A clinical registry and meta-analysis. *Journal of Stroke*, 25(2):299–302, May 2023. ISSN 2287-6405. doi: 10.5853/jos.2022.03692. URL http://dx.doi.org/10.5853/jos.2022.03692.

Paola Bertoli and Veronica Grembi. The life-saving effect of hospital proximity. *Health Economics*, 26(S2):78–91, 2017. ISSN 1099-1050. doi: 10.1002/hec.3571. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3571. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.3571.

Malvina Bondy, Sefi Roth, and Lutz Sager. Crime is in the air: The contemporaneous relationship between air pollution and crime. *Journal of the Association of Environmental and Resource Economists*, 7(3):555–585, 2020. doi: 10.1086/707127.

Marshall Burke, Solomon M. Hsiang, and Edward Miguel. Global non-linear effect of temperature on economic production. *Nature*, 527(7577):235–239, October 2015. ISSN 1476-4687. doi: 10.1038/nature15725. URL http://dx.doi.org/10.1038/nature15725.

David Card. Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69(5):1127–1160, 2001. ISSN 1468-0262. doi: 10.1111/1468-0262.00237. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00237. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0262.00237.

Melissa Dell, Benjamin F Jones, and Benjamin A Olken. Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, 4(3):66–95, July 2012. ISSN 1945-7715. doi: 10.1257/mac.4.3.66. URL http://dx.doi.org/10.1257/mac.4.3.66.

Olivier Deschênes and Michael Greenstone. The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review*, 97(1):354–385, February 2007. ISSN 0002-8282. doi: 10.1257/aer.97.1.354. URL http://dx.doi.org/10.1257/aer.97.1.354.

Marian Douarinou, Benjamin Gory, Arturo Consoli, Bertrand Lapergue, Maeva Kyheng, Julien Labreuche, Mohammad Anadani, Raphael Blanc, Gaultier Marnat, Romain Bourcier, Igor Sibon, François Eugène, Stéphane Vannier, Gérard Audibert, Gioia Mione, Sébastien Richard, Michel Piotin, Hocine Redjem, Simon Escalard, Jean-Philippe Dessilles, François Delvoye, Stanislas Smajda, Benjamin Maier, Hebert Solène, Mikael Mazighi, Mikael Obadia, Candice Sabben, Seners Pierre, Raynouard Igor, Ovide Corabianu, Thomas de Broucker, Eric Manchon, Guillaume Taylor, Malek Ben Maacha, Laurie-Anne Thion, Lecler Augustin, Savatovsjy Julien, Adrien Wang, Serge Evrard, Maya Tchikviladze, Nadia Ajili, David WeisenburgerLile, Lucas Gorza, Géraldine Buard, Oguzhan Coskun, Federico Di Maria, Georges Rodesh, Sergio Zimatore, Morgan Leguen, Julie Gratieux, Fernando Pico, Haja Rakotoharinandrasana, Philippe Tassan, Roxanna Poll, Sylvie Marinier, Florent Gariel, Xavier

Barreau, Jérôme Berge, Patrice Menegon, Ludovic Lucas, Stéphane Olindo, Pauline Renou, Sharmila Sagnier, Mathilde Poli, Sabrina Debruxelles, François Rouanet, Thomas Tourdias, Jean-Sebastien Liegey, Pierre Briau, Nicolas Pangon, Lili Detraz, Benjamin Daumas-Duport, Pierre-Louis Alexandre, Monica Roy, Cédric Lenoble, Hubert Desal, Benoît Guillon, Solène de Gaalon, Cécile Preterre, Serge Bracard, René Anxionnat, Marc Braun, Anne-Laure Derelle, Romain Tonnelet, Liang Liao, François Zhu, Emmanuelle Schmitt, Sophie Planel, Lisa Humbertjean, Jean-Christophe Lacour, Nolwenn Riou-Comte, Marcela Voicu, Lionel Alb, Marie Reitter, Madalina Brezeanu, Agnès Masson, Adriana Tabarna, Iona Podar, Francisco Macian-Montoro, Suzanna Saleme, Charbel Mounayer, Aymeric Rouchaud, Vincent Costalat, Caroline Arquizan, Cyril Dargazanli, Grégory Gascou, Pierre-Henri Lefèvre, Imad Derraz, Carlos Riquelme, Nicolas Gaillard, Isabelle Mourand, Lucas Corti, Federico Cagnazzo, Adrien ter Schiphorst, Jean-Christophe Ferre, Hélène Raoult, Thomas Ronziere, Maria Lassale, Christophe Paya, Jean-Yves Gauvrit, Clément Tracol, and Sophie Langnier-Lemercier. Impact of strategy on clinical outcome in large vessel occlusion stroke successfully reperfused: Etis registry results. *Stroke*, 53(1), January 2022. ISSN 1524-4628. doi: 10.1161/strokeaha.121.034422. URL http://dx.doi.org/10.1161/STROKEAHA.121.034422.

Xavier D'Haultfœuille, Stefan Hoderlein, and Yuya Sasaki. Testing and relaxing the exclusion restriction in the control function approach. *Journal of Econometrics*, 240 (2):105075, March 2024. ISSN 0304-4076. doi: 10.1016/j.jeconom.2020.09.012. URL https://www.sciencedirect.com/science/article/pii/S0304407621000439.

Rody El Nawar, Bertrand Lapergue, Michel Piotin, Benjamin Gory, Raphael Blanc, Arturo Consoli, Georges Rodesch, Mikael Mazighi, Frederic Bourdain, Maéva Kyheng, Julien Labreuche, Fernando Pico, Michel Piotin, Raphaël Blanc, Hocine Redjem, Simon Escalard, Jean-Philippe Desilles, Gabriele Ciccio, Stanislas Smajda, Mikael Mazighi, Robert Fahed, Michael Obadia, Candice Sabben, Ovide Corabianu, Thomas de Broucker, Didier Smadja, Sonia Alamowitch, Olivier Ille, Eric Manchon, Pierre-Yves Garcia, Guillaume Taylor, Malek Ben Maacha, Frédéric Bourdain, Jean-Pierre Decroix, Adrien Wang, Serge Evrard, Maya Tchikviladze, Bertrand Lapergue, Oguzhan Coskun, Arturo Consoli, Federico Di Maria, Georges Rodesch, Morgan Leguen, Marie Tisserand, Fernando Pico, Haja Rakotoharinandrasana, Philippe Tassan, Roxanna Poll, Benjamin Gory, Paul Emile Labeyrie, Roberto Riva, Francis Turjman, Norbert Nighoghossian, Laurent Derex, Tae-Hee Cho, Laura Mechtouff, Anne-Claire Lukaszewicz, Frédéric Philippeau, Serkan Cakmak, Karine Blanc-Lasserre, and Anne-Evelyne Vallet. Higher annual operator volume is associated with better reperfusion rates in stroke patients treated by mechanical thrombectomy. *JACC: Cardiovascular Interventions*, 12(4):385–391, February 2019. ISSN 1936-8798. doi: 10.1016/j.jcin.2018.12.007. URL http://dx.doi.org/10.1016/j.jcin.2018.12.007.

J. P. Florens, J. J. Heckman, C. Meghir, and E. Vytlacil. Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects. *Econometrica*, 76(5):1191–1206, 2008. ISSN 1468-0262. doi: 10.3982/ECTA5317. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA5317. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA5317.

Puck S. S. Fransen, Olvert A. Berkhemer, Hester F. Lingsma, Debbie Beumer, Lucie A. van den Berg, Albert J. Yoo, Wouter J. Schonewille, Jan Albert Vos, Paul J.

Nederkoorn, Marieke J. H. Wermer, Marianne A. A. van Walderveen, Julie Staals, Jeannette Hofmeijer, Jacques A. van Oostayen, Geert J. Lycklama à Nijeholt, Jelis Boiten, Patrick A. Brouwer, Bart J. Emmer, Sebastiaan F. de Bruijn, Lukas C. van Dijk, L. Jaap Kappelle, Rob H. Lo, Ewoud J. van Dijk, Joost de Vries, Paul L. M. de Kort, J. S. Peter van den Berg, Boudewijn A. A. M. van Hasselt, Leo A. M. Aerden, René J. Dallinga, Marieke C. Visser, Joseph C. J. Bot, Patrick C. Vroomen, Omid Eshghi, Tobien H. C. M. L. Schreuder, Roel J. J. Heijboer, Koos Keizer, Alexander V. Tielbeek, Heleen M. den Hertog, Dick G. Gerrits, Renske M. van den Berg-Vos, Giorgos B. Karas, Ewout W. Steyerberg, H. Zwenneke Flach, Henk A. Marquering, Marieke E. S. Sprengers, Sjoerd F. M. Jenniskens, Ludo F. M. Beenen, René van den Berg, Peter J. Koudstaal, Wim H. van Zwam, Yvo B.W. E. M. Roos, Robert J. van Oostenbrugge, Charles B. L. M. Majoie, Aad van der Lugt, and Diederik W. J. Dippel. Time to reperfusion and treatment effect for acute ischemic stroke: A randomized clinical trial. *JAMA Neurology*, 73 (2):190, February 2016. ISSN 2168-6149. doi: 10.1001/jamaneurol.2015.3886. URL http://dx.doi.org/10.1001/jamaneurol.2015.3886.

John Garen. The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica*, 52(5):1199–1218, 1984. ISSN 0012-9682. doi: 10.2307/1910996. URL https://www.jstor.org/stable/1910996. Publisher: [Wiley, Econometric Society].

D. C. Goff. Knowledge of heart attack symptoms in a population survey in the united states: The react trial. *Archives of Internal Medicine*, 158(21):2329–2338, November 1998. ISSN 0003-9926. doi: 10.1001/archinte.158.21.2329. URL http://dx.doi.org/10.1001/archinte.158.21.2329.

Mayank Goyal, Bijoy K Menon, Wim H van Zwam, Diederik W J Dippel, Peter J Mitchell, Andrew M Demchuk, Antoni Dávalos, Charles B L M Majoie, Aad van der Lugt, Maria A de Miquel, Geoffrey A Donnan, Yvo B W E M Roos, Alain Bonafe, Reza Jahan, Hans-Christoph Diener, Lucie A van den Berg, Elad I Levy, Olvert A Berkhemer, Vitor M Pereira, Jeremy Rempel, Mònica Millán, Stephen M Davis, Daniel Roy, John Thornton, Luis San Román, Marc Ribó, Debbie Beumer, Bruce Stouch, Scott Brown, Bruce C V Campbell, Robert J van Oostenbrugge, Jeffrey L Saver, Michael D Hill, and Tudor G Jovin. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *The Lancet*, 387(10029):1723–1731, April 2016. ISSN 0140-6736. doi: 10.1016/s0140-6736(16)00163-x. URL http://dx.doi.org/10.1016/S0140-6736(16)00163-X.

Michael Grossman. *The Demand for Health: A Theoretical and Empirical Investigation*. Columbia University Press, September 1972. ISBN 9780231544528. doi: 10.7312/gros17900. URL http://dx.doi.org/10.7312/gros17900.

Zijian Guo and Dylan S. Small. Control Function Instrumental Variable Estimation of Nonlinear Causal Effect Models. 2016.

E. Gómez–Déniz, D. I. Gallardo, and H. W. Gómez. Quasi-binomial zero-inflated regression model suitable for variables with bounded support. *Journal of Applied Statistics*, 47(12):2208–2229, September 2020. ISSN 0266-4763. doi: 10.1080/02664763.2019.1707517. URL https://doi.org/10.1080/02664763.2019.1707517. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02664763.2019.1707517.

Trevor J. Hastie. Generalized Additive Models. In *Statistical Models in S*. Routledge, 1992. ISBN 978-0-203-73853-5. Num Pages: 59.

James J. Heckman and Edward Vytlacil. Policy-Relevant Treatment Effects. *The American Economic Review*, 91(2):107–111, 2001. ISSN 0002-8282. URL https://www.jstor.org/stable/2677742. Publisher: American Economic Association.

Roland Hodler and Paul A. Raschky. Economic shocks and civil conflict at the regional level. *Economics Letters*, 124(3):530–533, September 2014. ISSN 0165-1765. doi: 10.1016/j.econlet.2014.07.027. URL http://dx.doi.org/10.1016/j.econlet.2014.07.027.

Henrik Jaldell, Prachaksvich Lebnak, and Anurak Amornpetchsathaporn. Time Is Money, But How Much? The Monetary Value of Response Time for Thai Ambulance Emergency Services. *Value in Health*, 17(5):555–560, July 2014. ISSN 1098-3015. doi: 10.1016/j.jval.2014.05.006. URL https://www.sciencedirect.com/science/article/pii/S1098301514018786.

Raed A. Joundi, Michael D. Hill, Jillian Stang, Dana Nicol, Amy Ying Xin Yu, Moira K. Kapral, James A. King, Mary-Lou Halabi, and Eric E. Smith. Association between time to treatment with endovascular thrombectomy and home-time after acute ischemic stroke. *Neurology*, 102(12), June 2024. ISSN 1526-632X. doi: 10.1212/wnl.0000000000209454. URL http://dx.doi.org/10.1212/WNL.0000000000209454.

Victoria Lambrou, Benjamin Gory, Bertrand Lapergue, Gaultier Marnat, Raoul Pop, Lionel Calviere, Mohammad Anadani, Raphael Blanc, and Stephanos Nikolaos Finitsis. Number of passes and outcome of endovascular treatment of anterior circulation large core ischemic stroke: insights from the endovascular treatment in ischemic stroke (etis) registry. *Journal of NeuroInterventional Surgery*, pages jnis–2024–021857, May 2024. ISSN 1759-8486. doi: 10.1136/jnis-2024-021857. URL http://dx.doi.org/10.1136/jnis-2024-021857.

Judith H. Lichtman, Erica C. Leifheit-Limson, Emi Watanabe, Norrina B. Allen, Brian Garavalia, Linda S. Garavalia, John A. Spertus, Harlan M. Krumholz, and Leslie A. Curry. Symptom Recognition and Healthcare Experiences of Young Women with Acute Myocardial Infarction. *Circulation. Cardiovascular quality and outcomes*, 8(2 Suppl 1):S31–S38, March 2015. ISSN 1941-7713. doi: 10.1161/CIRCOUTCOMES.114.001612. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4801001/.

Elena Lucchese. How important are delays in treatment for health outcomes? The case of ambulance response time and cardiovascular events. *Health Economics*, n/a(n/a), 2023. ISSN 1099-1050. doi: 10.1002/hec.4791. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.4791. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.4791.

Lu Ma, Hao Zhang, Xuedong Yan, Jiangfeng Wang, Ziqi Song, and Hui Xiong. Smooth associations between the emergency medical services response time and the risk of death in road traffic crashes. *Journal of Transport & Health*, 12:379–391, March 2019. ISSN 2214-1405. doi: 10.1016/j.jth.2018.08.011. URL https://www.sciencedirect.com/science/article/pii/S2214140518300641.

Mikael Mazighi, Saqib A. Chaudhry, Marc Ribo, Pooja Khatri, David Skoloudik, Maxim Mokin, Julien Labreuche, Elena Meseguer, Sharon D. Yeatts, Adnan H. Siddiqui, Joseph Broderick, Carlos A. Molina, Adnan I. Qureshi, and Pierre Amarenco. Impact of Onset-to-Reperfusion Time on Stroke Mortality. *Circulation*, 127(19): 1980–1985, 2013. doi: 10.1161/CIRCULATIONAHA.112.000311. URL https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.112.000311. _eprint: https://www.ahajournals.org/doi/pdf/10.1161/CIRCULATIONAHA.112.000311.

Jonathan Mellon. Rain, rain, go away: 194 potential exclusion-restriction violations for studies using weather as an instrumental variable. *American Journal of Political Science*, August 2024. ISSN 1540-5907. doi: 10.1111/ajps.12894. URL http://dx.doi.org/10.1111/ajps.12894.

Edward Miguel, Shanker Satyanath, and Ernest Sergenti. Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112(4): 725–753, August 2004. ISSN 1537-534X. doi: 10.1086/421174. URL http://dx.doi.org/10.1086/421174.

Maxim J.H.L. Mulder, Ivo G.H. Jansen, Robert-Jan B. Goldhoorn, Esmee Venema, Vicky Chalos, Kars C.J. Compagne, Bob Roozenbeek, Hester F. Lingsma, Wouter J. Schonewille, Ido R. van den Wijngaard, Jelis Boiten, Jan Albert Vos, Yvo B.W.EM. Roos, Robert J. van Oostenbrugge, Wim H. van Zwam, Charles B.L.M. Majoie, Aad van der Lugt, and Diederik W.J. Dippel. Time to endovascular treatment and outcome in acute ischemic stroke: Mr clean registry results. *Circulation*, 138(3): 232–240, July 2018. ISSN 1524-4539. doi: 10.1161/circulationaha.117.032600. URL http://dx.doi.org/10.1161/CIRCULATIONAHA.117.032600.

Météo-France. Observations du réseau sol de france. URL https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=93&id_rubrique=32. Accessed: 2024-11-25.

Whitney K. Newey and James L. Powell. Instrumental Variable Estimation of Nonparametric Models. *Econometrica*, 71(5):1565–1578, 2003. ISSN 0012-9682. URL https://www.jstor.org/stable/1555512. Publisher: [Wiley, Econometric Society].

Whitney K. Newey, James L. Powell, and Francis Vella. Nonparametric Estimation of Triangular Simultaneous Equations Models. *Econometrica*, 67(3):565–603, 1999. ISSN 0012-9682. URL https://www.jstor.org/stable/2999545. Publisher: [Wiley, Econometric Society].

Oliver Rivero-Arias, Melissa Ouellet, Alastair Gray, Jane Wolstenholme, Peter M. Rothwell, and Ramon Luengo-Fernandez. Mapping the modified rankin scale (mrs) measurement into the generic euroqol (eq-5d) health outcome. *Medical Decision Making*, 30(3):341–354, October 2009. ISSN 1552-681X. doi: 10.1177/0272989x09349961. URL http://dx.doi.org/10.1177/0272989X09349961.

Dheeya Rizmie, Laure de Preux, Marisa Miraldo, and Rifat Atun. Impact of extreme temperatures on emergency hospital admissions by age and socio-economic deprivation in England. *Social Science & Medicine*, 308:115193, September 2022. ISSN 0277-9536. doi: 10.1016/j.socscimed.2022.115193. URL https://www.sciencedirect.com/science/article/pii/S0277953622004993.

Lutz Sager. Estimating the effect of air pollution on road safety using atmospheric temperature inversions. *Journal of Environmental Economics and Management*, 98: 102250, 2019. doi: 10.1016/j.jeem.2019.102250.

Jeffrey L. Saver, Gregg C. Fonarow, Eric E. Smith, Mathew J. Reeves, Maria V. Grau-Sepulveda, Wenqin Pan, DaiWai M. Olson, Adrian F. Hernandez, Eric D. Peterson, and Lee H. Schwamm. Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke. *JAMA*, 309(23):2480, June 2013. ISSN 0098-7484. doi: 10.1001/jama.2013.6959. URL http://dx.doi.org/10.1001/jama.2013.6959.

Jeffrey L. Saver, Mayank Goyal, Aad van der Lugt, Bijoy K. Menon, Charles B. L. M. Majoie, Diederik W. Dippel, Bruce C. Campbell, Raul G. Nogueira, Andrew M. Demchuk, Alejandro Tomasello, Pere Cardona, Thomas G. Devlin, Donald F. Frei, Richard du Mesnil de Rochemont, Olvert A. Berkhemer, Tudor G. Jovin, Adnan H. Siddiqui, Wim H. van Zwam, Stephen M. Davis, Carlos Castaño, Biggya L. Sapkota, Puck S. Fransen, Carlos Molina, Robert J. van Oostenbrugge, Ángel Chamorro, Hester Lingsma, Frank L. Silver, Geoffrey A. Donnan, Ashfaq Shuaib, Scott Brown, Bruce Stouch, Peter J. Mitchell, Antoni Davalos, Yvo B. W. E. M. Roos, and Michael D. Hill. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: A meta-analysis. *JAMA*, 316(12):1279, September 2016. ISSN 0098-7484. doi: 10.1001/jama.2016.13647. URL http://dx.doi.org/10.1001/jama.2016.13647.

Kenneth A. Schultz and Justin S. Mankin. Is temperature exogenous? the impact of civil conflict on the instrumental climate record in sub-saharan africa. *American Journal of Political Science*, 63(4):723–739, March 2019. ISSN 1540-5907. doi: 10.1111/ajps.12425. URL http://dx.doi.org/10.1111/ajps.12425.

Luigi Siciliani, Valerie Moran, and Michael Borowitz. WHAT WORKS? WAITING TIME POLICIES IN THE HEALTH SECTOR. 2015.

Alejandro M Spiotta, Jan Vargas, Raymond Turner, M Imran Chaudry, Holly Battenhouse, and Aquilla S Turk. The golden hour of stroke intervention: effect of thrombectomy procedural time in acute ischemic stroke on outcome. *Journal of NeuroInterventional Surgery*, 6(7):511–516, September 2013. ISSN 1759-8486. doi: 10.1136/neurintsurg-2013-010726. URL http://dx.doi.org/10.1136/neurintsurg-2013-010726.

David Swan and Luc Baumstark. Does Every Minute Really Count? Road Time as an Indicator for the Economic Value of Emergency Medical Services. *Value in Health*, 25(3):400–408, March 2022. ISSN 1098-3015. doi: 10.1016/j.jval.2021.09.009. URL https://www.sciencedirect.com/science/article/pii/S1098301521017502.

Alex J Turner, Igor Francetic, Ruth Watkinson, Stephanie Gillibrand, and Matt Sutton. Socioeconomic inequality in access to timely and appropriate care in emergency departments. *Journal of Health Economics*, 85:102668, September 2022. ISSN 0167-6296. doi: 10.1016/j.jhealeco.2022.102668. URL https://www.sciencedirect.com/science/article/pii/S0167629622000844.

Elizabeth Ty Wilde. DO EMERGENCY MEDICAL SYSTEM RESPONSE TIMES MATTER FOR HEALTH OUTCOMES? *Health Economics*, 22(7):790–806, July 2013. ISSN 1057-9230, 1099-1050. doi: 10.1002/hec.2851. URL https://onlinelibrary.wiley.com/doi/10.1002/hec.2851.

J.T. Lindsay Wilson, Asha Hareendran, Marie Grant, Tracey Baird, Ursula G.R. Schulz, Keith W. Muir, and Ian Bone. Improving the Assessment of Outcomes in Stroke. *Stroke*, 33(9):2243–2246, September 2002. doi: 10.1161/01.STR.0000027437.22450.BD. URL https://www.ahajournals.org/doi/10.1161/01.STR.0000027437.22450.BD. Publisher: American Heart Association.

Jeffrey M. Wooldridge. Control Function Methods in Applied Econometrics. *The Journal of Human Resources*, 50(2):420–445, 2015. ISSN 0022-166X. URL https://www.jstor.org/stable/24735991. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System].

Joany M Zachariasse, Vera van der Hagen, Nienke Seiger, Kevin Mackway-Jones, Mirjam van Veen, and Henriette A Moll. Performance of triage systems in emergency care: a systematic review and meta-analysis. *BMJ Open*, 9(5):e026471, May 2019. ISSN 2044-6055. doi: 10.1136/bmjopen-2018-026471. URL http://dx.doi.org/10.1136/bmjopen-2018-026471.

Xinlei Zhu, Renjie Chen, Jing Yuan, Yang Liu, Yong Wang, Xunming Ji, Haidong Kan, and Jing Zhao. Hourly Heat Exposure and Acute Ischemic Stroke. *JAMA Network Open*, 7(2):e240627, February 2024. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2024.0627. URL https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2815560.