DEMS WORKING PAPER SERIES

It could be worse...it could be raining : Ambulance response time and health outcomes

Elena Lucchese

No. 429 – February 2020

Department of Economics, Management and Statistics
University of Milano – Bicocca
Piazza Ateneo Nuovo 1 – 2016 Milan, Italy
http://dems.unimib.it/

It could be worse...it could be raining: Ambulance response time and health outcomes

Elena Lucchese*

Abstract

Ambulance response time to emergency calls is a key indicator of a health system's efficiency although its impact on health is not precisely known. This causal relation is identified by exploiting rainfall at the time of the ambulance run as a shock to responsiveness. The elasticity of the likelihood of a severe cardiovascular condition with respect to response time is 0.9 and that of the likelihood of death before reaching the hospital is 5. Finally, the economic value of time is quantified, and it is shown that improving the ambulance's ability to locate the scene would substantially increase efficiency.

JEL classification: C26, I18, R41

*Department of Economics, Management and Statistics, University of Milano-Bicocca, email: elena.lucchese@unimib.it. I wish to thank my supervisors, Margherita Fort and Matteo Lippi Bruni, to which I am indebted. I am thankful to Jerome Adda, Vincenzo Atella, Marco Bertoni, Jayhanta Bhattacharya, Decio Coviello, Amitabh Chandra, Joseph Doyle, Mark Duggan, Veronica Grembi, Sara Lazzaroni, Giacomo Pasini, Nicola Persico, Paolo Roberti and Jonathan Skinner, and seminar participants at EEA/ESEM, Brucchi Luchino, CRENoS, University of Bologna and Stanford University for helpful comments and suggestions. I acknowledge support from AIES, EIEF and the University of Bologna. A special thank goes to Francesco Bermano, Edoardo Berti Riboli and Domenico Gallo without whom this project would have not been possible and to Andrea Furgani for his constant support.

Keywords: efficiency, response time, health, policy

1 Introduction

Healthcare systems provide a wide range of services and consume about 10% of global GDP.[1] In order to foster the efficient use of resources and support policy-making it is important to investigate the effectiveness of medical care (Baicker et al., 2012).[2] However, the complexity of the setting and the limited availability of adequate data have hindered the effort to carry out causal analysis.

The aim of this research is to quantify the effect of ambulance response time on health outcomes. Response time is a crucial performance measure used by policymakers to evaluate the overall quality of a healthcare system.[3] Questionnaires distributed among discharged patients reveal that the responsiveness of healthcare providers is one of the core performance measures that offers the greatest margin for improvement.[4] In addition, learning more about the effect of response time on health will allow policymakers to better gauge the effect of healthcare policies. An example is the improvement in access to ambulance services as a result of the

[1] World Health Organization Global Health Expenditure database (apps.who.int/nha/database).

[2] This approach is recommended also by the standard economic model of healthcare due to Auster et al. (1972)

[3] The ambulance service in Europe must meet specific response-time targets which are usually defined by regulation or national law (ec.europa.eu/health). In the United States, local health care agencies contractually set response time levels together with ambulance providers (Ludwig, 2004). Response time is also important in other types of emergencies. For instance, Blanes i Vidal and Kirchmaier (2017) show that police response time is a crucial determinant of clearance rate.

[4] See, for instance, the summary results of the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey in the United States: https://hcahpsonline.org/en. The aim of this questionnaire is to support consumer choice and to create incentives for hospitals to improve their quality of care.

Affordable Care Act in the United States, which resulted in a 19% increase in wait times due to the increased strain on the system (Courtemanche et al., 2019). A prompt response, especially during a medical emergency, may substantially affect the likelihood of severe health conditions or death. However, in the absence of studies that precisely quantify the effect of response time on health it is difficult to evaluate the cost and benefit associated with different healthcare policies.

Quantifying the direct effect of response time on health is not straightforward and for a number of reasons. First, patients may sort across hospitals, resulting in different outcomes for similar treatments. In addition, hospital patients receive a number of compounding treatments thus affecting the external validity of results since such treatments might vary substantially across jurisdictions and over time.[5] Furthermore, such treatments may be influenced by the amount of time the patient waited for them, such that patients who wait longer may receive more intense treatment.[6] The speed of the medical response might also be endogenous with respect to the health condition, such that patients with a more severe condition might elicit a quicker response. This analysis deals with these issues and contributes to the literature by estimating the effect of response time on health where the effect is observed before medical treatment is provided, thus obtaining results with greater external validity. The final part of the analysis assigns a value to time in order to evaluate the cost-effectiveness of policies to reduce ambulance response time.

The ambulance setting is particularly appropriate for the analysis.[7] Doyle

[5]These issues have been discussed by Chandra et al. (2016), Finkelstein et al. (2016), Skinner and Staiger (2015) and Chandra and Staiger (2007), among others.

[6]Gruber et al. (2018) show that wait time might affect the intensity of medical treatment in the Emergency Department.

[7]A number of studies have recently focused on the emergency medical setting, which is one

et al. (2015), Doyle et al. (2019) and Hull (2018) also use the ambulance setting and in particular the preferences of different ambulance companies in dealing with the problem of patient sorting. One of the problems in these studies is that the level of performance of the ambulance companies may positively correlate with the quality of their preferred hospital. Ambulance performance does matter and Jena et al. (2017) show how marathons increase 30-day mortality by lengthening ambulance response times. They report no significant difference with respect to which hospital the patients are brought to.

The estimation makes use of an instrumental variable identification strategy that exploits rainfall at the time of the ambulance run as a shock to the speed of ambulance response. The findings suggest that response time may be one of the explanations for the heterogenous health outcomes observed across jurisdictions and over time, which have been extensively discussed in the literature. In particular, it is shown that a 10-percent increase in the time required to reach a patient experiencing a cardiovascular episode (20% of a standard deviation from a mean of 28 minutes) increases the probability of a life-threatening condition by 9 percent (8% of a standard deviation from a mean of 45%) and the likelihood of patient mortality by 49 percent (10% of a standard deviation from a mean of 4%).[8] In support of the idea that time matters, Gruber et al. (2018) show that reducing the time spent in the waiting room of an emergency department by 10 percent reduces mortality by 14 percent.

of the crucial nodes of the healthcare system (Berchet, 2015). The Emergency Department is responsible for providing medical treatment to patients with an immediate need and handles, on average, 33.5 visits per 100 population each year (OECD, 2011).

[8]The elasticity of life-threatening conditions with respect to response time is 0.9 and 5 in the case of mortality. Avdic (2016) shows that most deaths during cardiovascular emergencies take place before the patients are able to get in-hospital medical treatment.

The analysis makes use of administrative data on 30,149 ambulance runs that took place in the Italian region of Liguria over a two-year period (2013-2014) in response to patients experiencing a cardiovascular event.[9] The regional setting is particularly relevant in this context given that healthcare services are organized and managed at the regional level in most European countries. The data for Liguria provide a unique level of detail, which includes mission-level information on the health condition of the patients on arrival of the ambulance at the scene. This information provides external validity of the results by making it possible to estimate the link between response time and health outcome before the administration of other treatments, specific to a given time, region or country. Furthermore, Liguria has installed a software program that automates data collection in real time, thus minimizing the likelihood of mistakes and misreporting that might occur when information is self-reported by care providers following the event, as in the case of non-automated data collection.

The identifying assumption of the instrumental variable procedure is that rainfall affects the *severity* of cardiovascular problems at a given point in time and in a given municipality only because it lengthens response time. The medical literature provides support for this assumption. In particular, Phillips et al. (2004) show that controlling for public holidays eliminates the seasonality of cardiovascular events. Additionally, the analysis includes a rich set of covariates, such as municipality fixed effects and priority dispatch of the ambulance, among others. The estimated effects reported in this analysis are robust to included covariates and alternative

[9] Cardiovascular problems account for about 1/3 of all urgent ambulance missions and 70% of out-of-hospital deaths. For further details about the relevance of cardiovascular problems to morbidity and mortality rates in the population, see the report by the World Health Organization, Mendis et al. (2011).

specifications.

After quantifying the impact of time on health, it is possible to assign an economic value to time and formulate sound policy recommendations. In particular, this value can be used to establish a benchmark for formulating cost-effective policies. One policy option would be to improve the ability of ambulance drivers to navigate to the scene. Indeed, precise directions for the ambulance are crucial for a quick response. However, in many instances, this information may be unknown to the caller or – because of the shock of the moment – may not be communicated correctly. Adopting a technology that obtains this information directly from the caller's phone can substantially improve performance at very little cost. The policy proposed here makes use of smartphone technology to convey GPS or WiFi-based location data to emergency service providers. This information is sent directly to the call center without any active involvement of the caller and does not require any previous download of applications.[10] It is conservatively estimated that implementing this policy in Liguria would reduce the average response time by one minute at a cost of about 2,500 euros. All patients would benefit from quicker response times, but even in the case of cardiovascular patients alone an average reduction of one minute would lower the proportion of patients developing a severe condition by 1.5 percentage points and would reduce mortality by 0.7 percentage points, which translates into 226 patients and 105 lives, respectively, each year.

The rest of the paper is organized as follows: Section 2 describes the setup and the data. Section 3 introduces the empirical methodology. Section 4 presents the results and discusses their robustness and sensitivity to alternative specifications.

[10]For further details, see the technical report DTR/EMTEL-00035 by the European Telecommunications Standards Institute (ETSI).

Section 5 discusses policy recommendations. Section 6 concludes.

2 Data and Setup

2.1 Institutional setting

Emergency medical calls are received at a centralized call center where trained nurses collect information from the caller and assess the presumed pathology and degree of urgency. During the call, the nurse asks a predetermined set of questions, specifically designed to maximize the quality of information collected.[11] The nurse fills in a form with the information collected and conveys it to the ambulance crew that is dispatched to the scene. The ambulance crew is in continual contact with the call center which can provide it with further support if needed. The call center monitors all the ambulances and their locations and dispatches the closest one to the scene. The ambulance crew is composed of trained paramedics. They communicate with the call center the moment they arrive on the scene and provide an assessment of the patient's situation. They then diagnose the pathology and provide first aid. Finally, the patient is transported to the nearest hospital that can provide appropriate treatment. The choice of the hospital is made by the call center, based on the situation of each hospital at that moment. The service is provided free of charge, although a fee of 25 euros is charged if the situation was

[11] The entire procedure is described in a manual called *Dispatch*, which has been adopted by all developed healthcare systems. In Liguria, nurses attend specific courses that train according to the procedure described in this manual. The procedure involves a hierarchy of questions that enables a diagnosis of symptoms as one of 17 classes of pathology and according to 4 levels of urgency.

not an emergency.[12]

2.2 Measurement of performance and outcomes

The standard measure of performance in the emergency framework is response time, i.e. the time required to reach the scene once a request is made. The dataset covers two years: 2013 and 2014. In 2012, Liguria introduced an innovative informative system that records data in real time rather than it being self-reported by the ambulance crew after the incident. Thus, the data are highly reliable, and information on response time, which is the variable of interest, is precisely recorded. The severity of the patient's condition is recorded at three points in time: at the time of the emergency call; upon arrival at the patient's location, but before medical treatment; and upon arrival at the hospital.

2.3 Cardiovascular problems

Cardiovascular problems are the leading cause of death in developed countries.[13] Accordingly, the response to emergency cardiac events has become a topic of interest in the recent literature, partly because rapid response in this setting may be of crucial importance (see, for example, Pons et al. (2005); Wilde (2013); Avdic (2016)).[14] Ambulance responsiveness is a critical factor in this context. Indeed, when the emergency medical systems were developed during the 90s, they were designed to provide a response to cardiac events and this included the introduc-

[12]There are some groups, namely pregnant women, children under the age of 14, the disabled, and low-income individuals, that are exempt from paying the fee.

[13]For a detailed discussion, see Mendis et al. (2011).

[14]See also Nichol et al. (1996) for a review of less recent research.

tion of response time standards for ambulances.[15] An additional feature of cardiac events that make them particularly attractive to researchers is that they are usually characterized by standard and well-documented symptoms. Therefore, correctly identifying the problem and the degree of severity is relatively simple and there is little chance of error that might confound the analysis.[16]

2.4 Descriptive statistics

The analysis makes use of administrative data on 30,149 ambulance dispatches to answer emergency calls related to cardiac events in the Italian region of Liguria in 2013 and 2014. The use of these data was authorized under a data-use agreement with the regional health authority. The regional setting is particularly relevant given that the health care system in Italy, as well as in most European countries, is managed at the regional level. Focussing at this geographical level is associated with a high degree of homogeneity in the characteristics of the service provided. In addition, Liguria is a setting of particular interest because of the unique quality of the data. The data were recorded using an innovative management system that collects data in real time. As a result, the data is more precise than that generated by the previous system of self-reporting by the ambulance crew upon arrival at the hospital. The data is recorded for each ambulance dispatch and includes, among other things, information about the pathology, age and gender of the patient, date and time of the call and when the ambulance crew arrived

[15]For further details, see the guidelines issued by the American Heart Association (1992) and the European Resuscitation Council (1992).

[16]It is relatively easy to identify the severity of a cardiac event and the nurse is able to identify the type of problem and degree of severity quite accurately during the emergency call. It is not as simple in the case of other types of injuries or car accidents, in which it is very difficult to establish the nature of the problem and the severity during the call.

at the patient's side, the type of ambulance (advanced vs basic life support), the municipality in which the patient is located, the priority of the ambulance dispatch and the severity of the patient both on the ambulance's arrival on the scene – but before medical treatment – and on admission to hospital. Exogenous shocks to ambulance response time are provided by the amount of rainfall at the time and in that location (municipality).

The main variable of interest is ambulance response time (RT) which is also the main measure of performance for policymakers. RT is the number of minutes from the start of the emergency call to the arrival of the ambulance on the scene.[17] Figure 4 presents the distribution of RT. The distribution is left-skewed and has an average of 28 minutes and a median of 25.6 minutes, as reported in Table 1.

[Figure 4 and Table 1 about here]

Data on rainfall were provided by the Regional Agency for the Environment of Liguria (ARPAL). The hourly data are collected by 213 land-based weather stations (represented by red dots in Figure 3 in the Appendix), and each station covers an average area of 20 km^2. We look at 242 municipalities, which is the smallest administrative unit within the region and which have an average area of about 30 km^2.[18] The amount of rainfall is expressed in millimeters (1 millimeter

[17]RT is the performance measure typically adopted in the literature (see for instance Blackwell and Kaufman (2002); Pons and Markovchick (2002); Swor and Cone (2002); Pons et al. (2005); Hollenberg et al. (2009)). Others have instead made use of the ambulance's driving time (from the time of the ambulance dispatch to its arrival on the scene), without taking into account the time required to manage the call and to dispatch the ambulance. This is typically due to data limitations, as in Wilde (2013). The measure of performance adopted here considers the total time required to reach the patient once an ambulance is requested. It is claimed that this is the best measure of time elapsed between the initial health shock and the administration of first aid – a period in which the health condition of the patient deteriorates quickly and a prompt response by the health system matters.

[18]There are 233 municipalities in Liguria. The municipality of Genoa, the largest city in the

= 0.04 inches). The average rainfall is 1.4 mm and 14% of ambulance dispatches take place during rainfall.

The outcomes of interest are the degree of severity observed at two points in time: the ambulance's arrival on the scene (H1) and arrival at the hospital (H2). The degree of severity is ranked from 1 to 5, where 4 is the highest level of severity and indicates a severe health impairment and that the patient is in imminent danger of dying, and 5 for the case that the patient dies.[19] An interesting feature of H1 is that it is recorded before performing any medical treatment. This is particularly useful since it offers the possibility of quantifying the direct effect of RT on observed severity. Table 1 reports the distribution of H1 and H2 in the sample. The main outcomes adopted in the analysis are M1, a dummy equal to one when H1 assumes the highest degree of severity and zero otherwise (which is the case for 45% of patients in the sample, a statistic consistent with data on cardiac events), and M2, a dummy equal to one when the patient dies out-of-hospital and zero otherwise (which is the case for 4% of the patients, a statistic that is consistent with data on timely treatment in cardiac events).[20] Severity and mortality rates in the sample are comparable to those reported in the literature. The analysis focuses on the highest degrees of severity – indexed by M1 and M2 – in order to reduce concerns about potential misjudgments by the medical personnel (since

region, has an area of 240 km^2, as opposed to other municipalities which are all in the vicinity of about 30 km^2. To achieve comparability, Genoa is split into its 9 neighborhoods, each of which has an area of about 27 km^2. The hourly amount of rainfall recorded by the land-based weather stations is interpolated by adopting the inverse distance weighting ratio, as suggested by Agrillo and Bonati (2013). In this way, the average level of rainfall at a given time and in a given municipality is obtained.

[19]There are 5 cases in which the patient dies in H1. Those observations are omitted in the process of sample construction. The final results were not affected. Details about sample construction are contained in the Appendix.

[20]In the discussion that follows, "M" refers to the health outcome in general, while M1 or M2 refer to a specific health outcome.

there is limited possible ambiguity associated with these conditions). Section 4.2 reports the results also for alternative outcome classifications and the estimated effect is shown to be robust and coherent.

The analysis includes a rich set of covariates, which include controls for characteristics of the ambulance dispatch, including when and where it took place and who was rescued; a fixed effect for the call center that managed the call (there are five such centers in Liguria which are reached by dialing a designated emergency number); a dummy variable for high-priority ambulance dispatches; a dummy for an advanced type of ambulance (Advanced Life Support (ALS) vehicle); and distance travelled and its square. Time controls include year, day of the week fixed effects and a dummy for weekend and public holidays. Patient's characteristics are controlled for using a dummy for males and an indicator for age category (50 $\geq$ age $\leq$ 79 years and $>$ 79 years, where $<$ 50 is the excluded category). Finally, population density in the municipality (low, medium or high) is controlled for and municipality fixed effects are included.[21] According to the descriptive statistics reported in Table 1, 92% of the ambulance dispatches are high priority and over 18% are performed by ALS vehicles. The average distance travelled is 20 kilometers. The sample is balanced between males and females. The average age is 70 and half of the sample is between the ages of 50 and 79. Almost half of the dispatches are in densely populated areas.[22]

Finally, Table 7 reports the balancing test of the covariates. For simplicity, the sample is split into two groups, rain vs no-rain (as opposed to the degree of

[21]By including fixed effects at the municipality level, the source of variation in the first stage, i.e. the effect of rainfall on response time, is the variation in rainfall over time, net of the effect of year, day of the week and holidays.

[22]Information on population density at the municipality level was provided by the Italian National Institute of Statistics (ISTAT).

variation adopted in the identification strategy, where the amount of rainfall is also exploited in order to instrument response time).[23] Column (1) reports the means and standard deviations of the covariates when it wasn't raining while column (2) reports them for when it was. Columns (3) and (4) show the difference between the means and the p-value of the difference, respectively. As expected, we observe a statistically significant difference in response times. Rainfall also correlates more or less with population density; including municipality fixed effects should capture this source of heterogeneity. The table does not report significant differences in patient demographics and the characteristics of the dispatch, such as distance driven and type of ambulance.

[Table 7 about here]

3 Empirical methodology

During medical emergencies, it is reasonable to expect that the production function of health crucially depends on time and in this section an attempt is made to quantify this relationship. To accomplish this, the baseline model presents a linear probability framework, which was chosen for ease of discussion. The regression results presented in Section 4 also report the estimates for non-linear models, which are in line with those obtained with the linear probability model.

The standard OLS model used to estimate the correlation between ambulance response time (RT) and health (M) is the following:

[23] Increasing the number of groups by rainfall intensity would further reduce observable differences across groups since, at the margin, the groups become virtually identical.

$$M_{ipt} = \alpha_0 + \mathbf{X}_{ipt}\alpha_1 + \alpha_2 RT_{ipt} + \epsilon_{ipt}, \tag{1}$$

where M is the health outcome of the patient rescued during ambulance dispatch i, which took place in municipality p and was initiated by a call made at time t. M is a dummy variable for the patient's health condition. M1 is the severity of the patient's condition observed when the ambulance crew reaches the patient's side and it is strictly related to the time it took to respond to the call. It takes a value of one when the severity is judged to be 4 on a scale of 1 to 4, which implies imminent risk of death. The outcome M2, on the other hand, reflects whether the patient has died by the time of arrival at the hospital.[24] $\mathbf{X}$ is the vector of controls that includes time and characteristics of the individual and the location, as described in the previous section. RT is the number of minutes from the start of the call until the arrival of the ambulance crew at the patient's side.

There are several potential issues when empirically attempting to isolate the effect of RT on M. First, it is difficult to fully control for factors that may characterize specific groups of events. For example, given that driving fast is a risk for the ambulance crew, they might adjust their behavior given all the information they have collected about the event, which may not be entirely observable and reflected in the data record. It is reasonable to assume that the ambulance crew will respond fastest to the most critical cases. In this case, the OLS estimator will lead to downwardly biased estimates due to reverse causality.[25] As a result,

[24]For ease of discussion and given the relevance of severe health outcomes in the emergency medical context, the general discussion focuses on the probability of observing the most severe outcome, as captured by M1 and M2. In Section 4.2, results are presented for the other degrees of severity.

[25]The error term reported in Eq. 1 is the sum of two components: (i) an idiosyncratic factor, i.e. a part of the severity that is not observed by the ambulance driver nor by the researcher;

the specification reported in Equation 1 would not lead to correct estimates of the coefficient of interest. An instrumental variable identification strategy is used to address this problem. In particular, hourly changes in rainfall in the relevant municipality, given by Z in Equation 2, are used to instrument RT. The first stage of the estimation will then be:

$$RT_{ipt} = \beta_0 + \mathbf{X}_{ipt}\beta_1 + \beta_2 Z_{pt} + v_{ipt}, \tag{2}$$

where Z is millimeters of rainfall at time t of the emergency call in municipality p. The identifying assumption is that Z affects M only through its effect on RT. In other words, a patient has the same ex-ante probability of being in a severe health condition when the ambulance reaches the scene and of dying before reaching hospital whether or not it is raining and the only channel that explains a greater probability of a severe condition is longer RT due to Z. In particular, rainfall has a positive effect on RT because in order to drive safely the ambulance must proceed at a slower speed.[26] Even if weather can be forecasted, its effect cannot be completely mitigated. For this reason, weather shocks have been widely adopted in the literature as instrumental variables for endogenous regressors.[27] The fulfillment of the exclusion restrictions is well supported by the medical literature which has demonstrated the relationship between weather conditions and

and (ii) a factor that reflects characteristics of the patient that are observed by the ambulance crew but not by the researcher and that influence their response time. If this is negatively correlated with RT, i.e., RT is shorter for more severe cases, then OLS leads to downwardly biased estimates.

[26]See the Manual on Highway Capacity (2015) and Agarwal et al. (2005) for a discussion of the effect of rain on road conditions and traffic. See Thornes et al. (2014) for a discussion of the effect of weather conditions on ambulance performance.

[27]See Dell et al. (2014) for a review.

the severity of cardiovascular problems.[28] Furthermore, the granularity of the data tends to support the instrument's validity, since it is unlikely that variations in the *amount* of rainfall within the same municipality have a direct effect on the severity of cardiovascular problems. The results remain unchanged whether rainfall is included as a dummy equal to one for rainfall or as a variable measured in millimeters, as in the baseline specification. The analysis also includes a rich set of controls for characteristics of the incident, its timing, where it took place, and patient demographics. In Section 4, it is shown that the results are robust to the covariates included. Finally, using the rule of thumb proposed by Stock and Yogo (2002) that the instrument is strong enough if the first-stage F-statistic exceeds 10, the estimation of Equation 2 returns an F-statistic of 15.9.[29]

Estimating the following regression makes it possible to quantify the causal effect of RT on M:

$$M_{ipt} = \gamma_0 + \mathbf{X}_{ipt}\gamma_1 + \gamma_2 \hat{RT}_{ipt} + u_{ipt}, \tag{3}$$

where RT is augmented by Equation 2. Results are discussed in the following section. The estimates are obtained by clustering standard errors at the same level of variation as the instrument, i.e. at the time, date and municipality levels, as suggested by Abadie et al. (2017).[30]

[28] Phillips et al. (2004) show that the correlation between weather conditions and the severity and mortality rates of patients affected by cardiovascular problems disappears after controlling for holidays. In particular, cardiovascular problems are observed to spike during Thanksgiving, Christmas and New Year's Eve. In the analysis, I include fixed effects for public holidays.

[29] For completeness, Section 7.2 in the Appendix discusses the condition of monotonicity for internal validity. Suggestive evidence is presented on the monotonic effect of rainfall on RT, following the approach of Angrist and Imbens (1995).

[30] In particular, Abadie et al. (2017) discuss the effect of clustering at the aggregate level, arguing that this can lead to unnecessarily conservative standard errors, even in large samples. They discuss the motivation for the adjustment of standard errors and the appropriate level of

4 Results

The principal outcome of the analysis is M1, i.e. the health condition of the patient on arrival of the ambulance. Quantifying the relationship between RT and M1 is the best available measure of how fast health deteriorates as a function of time. Ambulance response provides the setting for a natural experiment to estimate this relation due to three key factors: the patient needs medical treatment but is forced to wait for it and rainfall is used to assign them quasi random waiting times; the moment the health shock began is observed quite precisely and is given by the moment in which the emergency call comes; and the moment that medical treatment could first be provided, i.e. when the ambulance reached the scene, is also known. It is difficult to think of a better setting in which to quantify the importance of medical responsiveness on health outcomes. Furthermore, the fact that M1 is observed before medical treatment is provided reinforces the validity of the results. This is due to the fact that treatment might differ across regions and countries, confounding the estimate of the direct effect of time on health.

The second outcome is M2, the out-of-hospital patient mortality observed up to arrival at the hospital. Quantifying the relationship between RT and M2 provides insight into whether the effect of longer RT is mitigated by medical treatment provided on the scene by the ambulance crew or whether it persists. By considering mortality, we can estimate the effect of RT on a condition that cannot be reversed, as opposed to other health conditions. While M1 is observed on arrival at the scene, such that the direct relationship between RT and M1 is straightforward, the link between RT and M2 may not be as direct given that there is an additional

clustering.

phase, namely the time required to drive back to the hospital. If the instrument adopted for RT, i.e. rainfall at the time and at the municipality level, also affects the duration of the return trip, this could result in an overestimate of the effect of RT on M2 since part of the effect would be due to that return trip.

The effect of rainfall on RT is subject to the fact that the ambulance is responding to an emergency call and generally tries to reach the patient as fast as possible, while rainfall decreases the maximum speed at which the ambulance can safely drive. Operating guidelines prescribe that ambulances should reach the scene of an accident as quickly as possible in order to provide first aid but should drive slowly on the way back in order to avoid the potential adverse effects of abrupt breaking and acceleration on the patient's condition. If so, then rainfall should not substantially affect the time required to drive back since the ambulance is driving slowly in any case and the identification strategy would rely mainly on the time till arrival on the scene, i.e. RT. Table 2 reports the effect of rainfall on RT and on the duration of the return trip. The estimates reported in the table show that rainfall has a large and statistically significant effect on RT but a small and not significant effect on the return trip. However, given the low precision of the estimates of rain's effect on the return trip it still may be that some of the effect of RT on M2 is due to slower return trips.

Tables 3 and 4 report the results for M1 (the probability of a patient being at the highest degree of severity at the time of the ambulance's arrival on the scene) and M2 (patient mortality before arrival at the hospital), respectively. Column (1) reports the estimates of the simple OLS model represented by Equation 1. Column (2) reports the first-stage estimates (FS) based on Equation 2 and column (3) reports the reduced-form intention-to-treat (ITT) estimates, i.e. the change in the probabilities of M1 and M2 when it is raining. Finally, column (4) reports the second-stage least squares (IV) estimates for Equation 3. In all cases, the full set of controls described in Section 2 was included and standard errors were clustered at the time, date, and municipality levels, as explained in Section 3. The tables report the estimated parameters for all the covariates included in the regressions. The main covariates of interest, namely response time (RT) and rainfall (Z), are reported in the first two rows.

[Tables 3 and 4 about here]

Column (1) of the two tables reports the correlation between RT and M (M1 in Table 3 and M2 in Table 4). According to these estimates, a one-minute increase in RT raises the probability of M1 by 0.9 percentage points (1.8% in terms of the standard deviation) and raises M2 by 0.3 pp (1.5% in terms of the standard deviation). As discussed in Section 3, the OLS estimates may be downward-biased because information known to the ambulance driver about the condition of a patient may result in faster ambulance responses in the case of patients in more severe condition. The effect of the instrumental variable on RT reported in column (2) shows that a millimeter of hourly rainfall increases RT by 1/3 of a minute (FS F-statistic of 15.9). The average amount of rainfall is 1.4 millimeters and the standard deviation is 2.5.[31] Column (3) shows that a millimeter of rainfall increases the probability that a patient will be in the most severe condition by 0.5 percentage points and the probability of mortality by 0.2 percentage points.

[31] Alternatively, we can redefine the instrument as a dummy equal to one for rain. In that case, RT during rain increases by 0.8 minutes. The results remain unchanged. Our preferred specification is rain in millimeters, which provides greater insight into the sensitivity of RT to rain, and the results are more easily comparable to other regions or countries where the average amount of hourly rainfall might be different.

Finally, column (4) quantifies the effect of response time on health after exploiting rainfall to address the problem of endogeneity. It is found that a one-minute increase in RT increases the probability of the most severe condition (M1) by 1.5 percentage points and the probability of death (M2) by 0.7 percentage points (3% and 3.5% respectively, in terms of the standard deviation).

The use of instrumental variables approximately doubles the estimated effect of RT relative to OLS. This is in line with the expected downward bias described above. Previous studies that did not address reverse causality reported little to no effect.[32] The OLS estimate obtained here is larger than that reported previously in the literature, and although smaller in size than the IV estimates, it is nonetheless significantly different from zero (but statistically smaller that the IV estimates). The relatively large OLS estimate reported here can be attributed to the data-recording method used to collect the data. Indeed, the aim of this analysis is to tackle the problem of endogeneity that arises from systematic differences in the response times of care providers due to the *ex-ante* health condition of a patient, thus biasing the estimated effect of response time on health. On the other hand, ambulance operators are evaluated on the basis of their response time and for most of the data currently available this information is self-reported by ambulance crews on their return to the hospital.[33] Given the complexity of emergency situations, it is reasonable to assume that the likelihood of systematic mistakes and misreporting correlate with the severity of a patient's condition. In contrast, the data collected for this study were automatically recorded in real time at all stages of the incident.

[32] See Blackwell and Kaufman (2002), Pons et al. (2005), Swor and Cone (2002), and the review of the literature by Nichol et al. (1996).

[33] In some countries, such as in the United States, the reimbursement of the ambulance crew for service depends on performance.

In this way, the likelihood of error and manipulation of data is minimized.

The magnitude of the IV effect reported here is, on the other hand, comparable to previous results in the literature, despite the focus here is on health outcomes before hospital treatment whether other settings look health outcomes observed after that. This is in line with the idea that time matters and that, as shown by Avdic (2016), most deaths during cardiovascular emergencies take place before the patients reach the hospital. Jena et al. (2017) report an increase in 30-day mortality rate for a cardiovascular episode due to ambulance delay similar to this setting (4.4 minutes longer responses on marathon days, and 3.3 percentage points risk difference in mortality). Wilde (2013) reports a growing effect over time following the event. In particular, the effect appears not to be significant within 24 hours from the event and then increases subsequently. The effect on 30-day mortality is similar to the estimate presented here, suggesting that they converge over time due to the fact the compounding effect of other confounding factors has been addressed.[34]

4.1 Sensitivity to included covariates

The sensitivity of the results to the included covariates is reported in Table 5 and 6, which relate to the outcomes M1 and M2, respectively. The first column of the tables reports the baseline results while columns (2)-(5) report the estimates as covariates are gradually excluded. Thus, column (2) excludes the set of controls related to the ambulance dispatch (contact center that received the call, priority dispatch of the ambulance, type of vehicle and distance driven). Column (3)

[34] In particular, Wilde (2013) exploits distance from the hospital. If individuals are not randomly allocated geographically, this might lead to underestimation of the short-term effect of ambulance responsiveness on health and an increase in the estimated effect over time.

further excludes time fixed effects (day of the week, holiday and year). Column (4) also excludes patient demographics (age group and gender) and finally column (5) also excludes fixed effects of the location's characteristics (population density and municipality fixed effects). The point estimate of the effect of RT on the outcomes of interest is stable across all the specifications. The first-stage F-statistic is reported at the bottom of the tables and is consistently above the threshold of 10. By excluding controls, the effect of RT on M2 loses some statistical power, especially in the unconditional specification reported in column (5), suggesting that the covariates help to clean the noise due to other factors, and particularly fixed effects for time and location, from the effect of interest.

[Tables 5 and 6 about here]

4.2 Alternative Specifications

Table 8 and 9 correspond, respectively, to outcome H1 (health condition on arrival on the scene) and H2 (health condition on arrival at the hospital) and report the results for various outcome specifications and for an ordered probit model instead of the linear probability model.[35]

Column (1) of Table 8 reports the baseline results where the outcome, M1, is a dummy equal to 1 for the highest degree of severity (4 on a scale of 1 to 4). Column (2) shows the effect of RT on conditions of severity 3 or 4, which describe over 90% of patients experiencing a cardiac event. The small amount of variation that results from grouping the two highest degrees of severity limits the possibility of identifying the effect. The magnitude of the effect is small and statistically

[35] As in the main specification, the results are obtained by including the full set of controls and clustering standard errors at the time, day, and municipality levels.

indistinguishable from zero. This suggests that ambulance response time tends to affect the probability of a shift from a severity of 3 to a severity of 4, rather than from 1 or 2 to 3 or 4. This is in line with the pattern of severity usually observed among cardiac patients.

Column (3) reports the estimates for the effect of RT on H1, the *level* of severity on arrival at the scene. A one-minute increase in RT increases H1 by 0.017 points; the average response time is 28 minutes. This result is almost identical to the baseline result, suggesting that the dummy variable M1 is a relevant classification of the outcome to highlight which kind of health condition is determined by RT. Finally, the outcome is redefined across three levels – 1 or 2, 3 and 4 – and the ordered probit model is then estimated. The parameters are designed to sum to zero and appear in the last two columns.[36] The marginal effect reported in column (5) is a 1.6-percentage point increase in the probability of severity of degree 4 for every additional minute, which is similar to the baseline result reported in the same column.[37]

[Table 8 about here]

Column (1) of Table 9 reports the baseline results when the outcome, M2, is a dummy equal to 1 in the case of death, i.e. severity of degree 5. Column (2) reports the results for the case when the dummy is equal to 1 for severity of degree

[36]The severities of degree 1 and 2 were combined because of the small number of observations of degree 1. To obtain convergence in the estimation of the model, some controls were excluded. The included controls were: fixed effects for priority dispatch, type of vehicle, distance driven and its square, patient's gender, age category, and fixed effect for day of the week. The standard errors are clustered at the usual level. The same applies to the estimates reported in Table 9

[37]The effect reported in column (4), i.e. for severity of degree 3, is -1.5. It almost cancels out the effect reported in column (5), indicating that longer RT shifts patients from 3 to 4 which is in line with the results reported in column (2). The ordered probit model is estimated at the sample average. The average values of the relevant regressors, namely average RT, patient's age, and distance travelled by the ambulance, are reported at the bottom of the table.

4 or 5. In column (3), the outcome is a dummy equal to 1 when the outcome is 3, 4, or 5. The magnitude of the effect reported in column (2) is larger, suggesting that RT negatively affects not just the likelihood of dying but also that of being in an extremely severe condition, which is consistent with the previous discussion. Column (4) shows the results for H2 as a level, such that a one-minute increase in RT raises H2 by 0.036 points. Finally, columns (5)-(7) report the marginal effects on H2 when it is equal to 3, 4 or 5 in the ordered probit model. (Given that the parameters sum to zero, the results for outcomes of 1 or 2 are omitted, as before.) A one-minute increase in RT increases the probability of the most severe condition (H2 of 4, column 6) by 1 percentage point and increases the probability of dying (H2 of 5, column 7) by 0.6 percentage points. The results of the ordered probit estimation are similar to those of the baseline estimation (column 1).

[Table 9 about here]

5 Policy discussion

In this section, an attempt is made to estimate the value of one minute of response time, with the goal of identifying the frontier at which an investment in reducing response time is cost-effective, thus providing a sort of indifference curve for policymakers. This is followed by a discussion of the cost-effectiveness of increasing ambulance performance by improving the quality of information conveyed during the emergency call. It is worth mentioning that the estimates presented in what follows are based on conservative back-of-the-envelope calculations.

5.1 The value of one minute

The economic value of reducing response time can be assessed looking at the number of lives that would be saved with a faster response time and place a value on them. The Value of a Statistical Life (VSL) approach provides an estimate of the economic value of a life.[38] Viscusi and Aldy (2003) present a comprehensive review of the literature that developed VSL measures and discuss the main differences between the various approaches and their application in public policy decisions. In what follows, the approach proposed by Murphy and Topel (2006) is adopted since it provides a way to estimate the value of *one year* of life. Other approaches relate to an entire lifetime but they may lead to overestimates in our context in view of the fact that the average age in the sample is 70 (standard deviation of 17).[39] It is assumed that someone undergoing a cardiac event who calls an ambulance will survive for at least one year if the hospital is reached in time. This assumption is supported by Avdic (2016) who shows that the mortality rate of emergency patients is similar at the time of admission to the hospital and one year later.

The sample consists of 30,149 ambulance dispatches for cardiac events over a period of two years. In 4% of the cases the patients died before reaching the hospital. The estimations show that, on average, one additional minute of response time increases the probability of mortality by 0.7 percentage points. Multiplying this by the number of ambulance patients per year provides the expected number of lives that would be saved by a one-minute reduction in ambulance response

[38] VSL measures are intended to provide policymakers with a reference point in assessing the benefit of risk-reduction efforts.

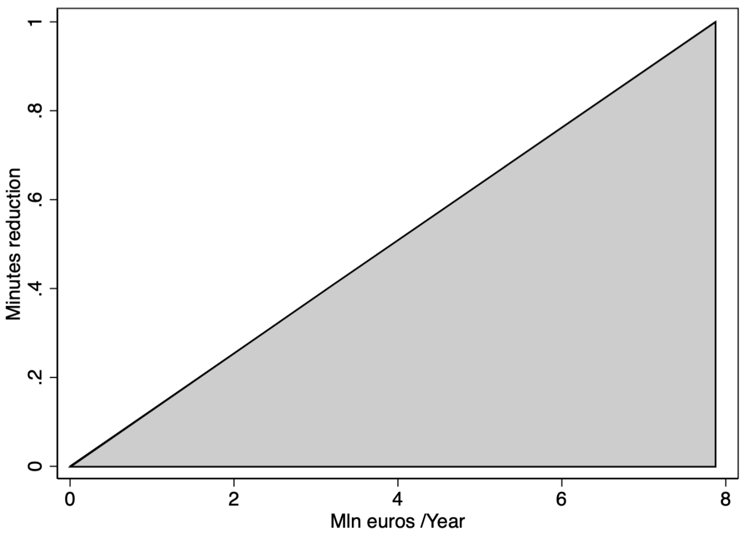
[39] The average life expectancy for a 65 year-old in Liguria is 18 years (ISTAT).

time, yielding a result of 105 lives (30,149*0.007/2).

According to Murphy and Topel (2006), the value of one year of life is equal to four times the individual's annual income. Given that over 75% of individuals in the sample are over 60 and the retirement age in Italy in 2013 was 62, the median annual retirement income in Liguria in 2013 can be used in the calculation. Thus, the value of one year of life in this setting is 74,640 euro.

The implied economic value of a one-minute reduction in response time is 7.8 million euro (105*74,640). This is essentially the maximum investment that a policymaker should be willing to make each year to reduce RT by one minute.[40] The frontier for cost-effective policies is presented in Figure 1.

Figure 1: Value of one minute



NOTES: The vertical axis measures the reduction in response time while the horizontal axis measures the annual cost. The solid diagonal line represents the policymaker's willingness-to-pay in order to obtain a given response time reduction in the case of patients experiencing a cardiac event. The white area indicates cost-effectiveness.

[40] This is likely to be a conservative estimate of the potential benefit since it is calculated only for cardiac patients. A reduction in RT would also benefit patients with other conditions.

The solid line represents a policymaker's willingness-to-pay for a given improvement in average response time. Policy proposals can be mapped onto this graph and their cost-effectiveness measured by the vertical distance between it and the frontier. Points below the frontier are obviously not worthwhile.

5.2 The ability to locate patients

The aim of an ambulance crew is to reach the patient as quickly as possible. Therefore, the quality of the directions provided to the ambulance driver is of crucial importance in determining response time. For example, the caller may not know the address or in the panic of the moment the address may not be communicated accurately. In this section, an attempt is made to estimate the magnitude of this problem.

Each ambulance run is composed of two segments: from the dispatch to the scene and from the scene back to the hospital. Each has a different likelihood of difficulty in locating the patient. On the way to the scene, the ambulance driver's ability to locate the patient depends on the quality of directions provided during the emergency call. On the way back, the driver knows the address of the destination hospital precisely and therefore this segment is probably not affected by this issue. Therefore, the magnitude of the problem is calculated as the difference between the driving time for each segment. This is accomplished by means of a regression in which the dependent variable is the duration of the trip back to the hospital (which is not affected by the problem of locating the patient) and is regressed onto a constant and the driving time of the trip to the scene. The constant term represents the average driving time of the return trip while the coefficient represents the difference in driving time between the two segments. The standard errors indicate how well these measures predict driving time. Descriptive statistics for driving time (in minutes) are presented in Table 10 and the regression results are presented in Table 11. Data on driving time is provided by a software application that records when the ambulance departs and arrives for each segment. The difference between the two driving times is not affected by factors that are fixed for a particular run, such as characteristics of the driver, the ambulance and the patient. On average, the ambulance takes 15 minutes to drive to the scene and 11.5 minutes to return. The estimation results show that driving time on the way to the scene is 3.7 minutes longer than on the way back.

[Tables 10 and 11 about here]

Using the difference in driving time as a proxy for the location problem is likely to yield a conservative estimate for two reasons. First, with the patient onboard the maximum speed that the ambulance can attain is lower on the return trip. Second, the distance driven by the ambulance on the return trip is often not shorter than the trip to the scene since the call center dispatches the closest ambulance to the patient who, in turn, is not always transported to the nearest hospital.

5.3 The cost-effectiveness of improving the ambulance's ability to locate the patient

Precise directions for the ambulance are crucial for a quick response. However, this information may be unknown to the caller or might not be communicated correctly.

Starting in 2016, a new technology was developed for the purpose of collecting information on location directly from the caller's smartphone and without the direct engagement of the caller. This innovation is promising for at least two reasons: first, it is effective for an overwhelming majority of callers given the popularity of smartphones; second, it is built into the phone and works without the direct engagement of the caller (as opposed to applications that have to be downloaded beforehand and activated by the caller when needed). The technology discussed here locates the caller automatically when an emergency number is dialed. This technology was installed in Android phones in 2016 and in Apple phones in 2018.[41] To take advantage of this tool, the call center's management system must install a software application that receives the information conveyed by the caller's phone. This tool can be installed at a very low cost and requires only minimal training of personnel. (It is assumed that 50 hours of operator training is needed, at a cost of 50 euros per hour, for a total of 2,500 euros.) If we estimate that half of the calls will be improved by the installation of this software and that the time wasted on locating the scene is halved, then average response time will be reduced by 0.9 minutes.[42]

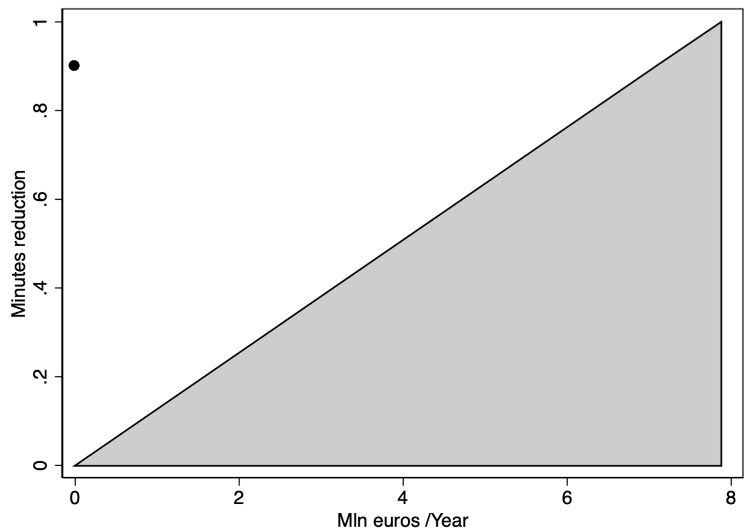
Figure 2 shows the cost-effectiveness frontier and the point representing the new technology. Clearly, this solution is highly cost-efficient. Furthermore, such technology would result in a more efficient use of already existing resources, such as ambulances, operators, technicians, paramedics, and physicians.[43]

[41]For further information, see technical report DTR/EMTEL-00035 by the European Telecommunications Standards Institute (ETSI).

[42]That is, 0.5*(3.7/2).

[43]A policy often discussed in the literature is the acquisition of additional ambulances. It is difficult to estimate the cost and benefit of this type of solution due to the shorter distance that will be driven by all the ambulances and the question of where to station it and even the cost of a parking spot. For completeness, Section 7.3 in the Appendix presents a cost-benefit analysis

Figure 2: Value of one minute and cost-benefit of improving location abilities



NOTES: The vertical axis measures the reduction in response time while the horizontal axis measures the annual cost. The solid diagonal line represents the policymaker's willingness-to-pay in order to obtain a given response time reduction in the case of patients experiencing a cardiac event. The dot represents the cost-benefit ratio associated with introducing a technology that improves communication during the emergency call.

6 Conclusions

We have investigated the contribution of ambulance response time (RT) to patients' health with the goal of gaining a better understanding of the importance of medical responsiveness and the implications of policies that reduce waiting time. A significant proportion of all hospital admissions involve a medical emergency.[44] Therefore, it is of crucial importance to understand the factors that affect the health outcome of these patients.

for acquiring an additional ambulance that follows the standard approach in the literature.

[44] In the United States, for instance, Emergency departments are responsible for about 1/3 of all acute medical visits and 1/2 of all hospital admissions: https://news.brown.edu/articles/2013/04/emergency (accessed on March 2019).

Medical responsiveness is widely considered to be one of the most critical determinants of patients' health outcomes. Nonetheless, there are challenges in precisely measuring response time and quantifying its importance. In the case of ambulance response, it is possible to observe the moment at which the need for medical treatment becomes known, i.e. the time of the emergency call, and the time that elapses until first aid is administered. Another issue is that patients in more critical conditions may receive more intense medical treatment, thus confounding the effect of response time *per se*. The dataset used in the analysis includes information about the patient's health condition on arrival of the ambulance but before medical treatment is provided. This is useful in quantifying the direct effect of response time on health outcomes, net of other confounding factors such as specific medical treatments, which would otherwise compromise the external validity of the findings. Finally, ambulance runs that have been assigned a higher priority may be characterized by an *ex-ante* higher or lower likelihood of saving the patient, thus affecting the speed of response. This problem has been addressed by exploiting the amount of rainfall at the time and in the municipality of the ambulance run in order to instrument response time.

The empirical analysis contributes to the literature in several ways. First, a novel instrument in this context is used to address the problem of endogeneity. Second, the effect of RT on the severity of a patient's condition on the ambulance's arrival at the scene is quantified and it is shown that a one-minute increase in RT (3.6% at the mean or 7% of a standard deviation) raises the probability of observing an extremely serious condition by 1.5 percentage points (3% at the mean or 3% of a standard deviation). It is calculated that the elasticity of the level of severity with respect to response time is 0.9. Thus, it can be concluded that RT has a sizable effect on a patient's condition, which is immediately visible and measurable. Third, RT also affects the likelihood that the patient dies before arrival at the hospital. A one-minute increase in RT raises the mortality rate by 0.7 percentage points (17.5% at the mean or 3.6% of a standard deviation). It is calculated that the elasticity of out-of-hospital mortality with respect to response time is equal to 5 at the sample mean. The results are robust across specifications and to included covariates. Thus, improving the performance of emergency medical services appears to generate potentially sizable rewards in terms of health.

Quantifying the effect of medical responsiveness on health outcomes makes it possible to estimate the economic value of response time and to evaluate the cost-effectiveness of proposed policies. Given the estimated effect of RT on mortality, a conservative back-of-the-envelope calculation indicates that the value of reducing the average RT by one minute in Liguria would be 7.8 million euros per year, just in the case of cardiac events. It is then shown that efficiency in locating the scene may be a crucial determinant of RT, most likely because of the poor quality of directions provided during the emergency call. Introducing a designated software system to improve the transmission of information during the emergency call should lead to a reduction of 0.9 minutes in response time at a cost of only 2,500 euros, a highly cost-effective investment.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? *NBER Technical Working Paper*, (w24003).

Agarwal, M., Maze, T. H., and Souleyrette, R. (2005). *Impacts of weather on urban freeway traffic flow characteristics and facility capacity.* Proceedings of the Mid-Continent Transportation Research Symposium, Ames, Iowa.

Agrillo, G. and Bonati, V. (2013). *Atlante climatico della Liguria.* ARPAL-Centro Funzionale, Liguria, Italia.

American Heart Association, A. H. A. (1992). Guidelines for cardiopulmonary resuscitation emergency cardiac care. *Journal of the American Medical Association*, 268(16):2184–2198.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442.

Auster, R., Leveson, I., and Sarachek, D. (1972). The production of health, an exploratory study. In *Essays in the economics of health and medical care*, pages 135–158. NBER.

Avdic, D. (2016). Improving efficiency or impairing access? Health care consolidation and quality of care: Evidence from emergency hospital closures in Sweden. *Journal of Health Economics*, 48:44–60.

Baicker, K., Chandra, A., and Skinner, J. S. (2012). Saving money or just saving lives? Improving the productivity of us health care spending. *Annual Reviews of Economics*, 4(1):33–56.

Berchet, C. (2015). Emergency care services: Trends, drivers and interventions to manage the demand. *OECD Health Working Papers*, (83).

Blackwell, T. H. and Kaufman, J. S. (2002). Response time effectiveness: Comparison of response time and survival in an urban emergency medical services system. *Academic Emergency Medicine*, 9(4):288–295.

Blanes i Vidal, J. and Kirchmaier, T. (2017). The effect of police response time on crime clearance rates. *The Review of Economic Studies*, 85(2):855–891.

Chandra, A., Finkelstein, A., Sacarny, A., and Syverson, C. (2016). Health care exceptionalism? Performance and allocation in the US health care sector. *American Economic Review*, 106(8):2110–44.

Chandra, A. and Staiger, D. O. (2007). Productivity spillovers in health care: Evidence from the treatment of heart attacks. *Journal of Political Economy*, 115(1):103–140.

Courtemanche, C., Friedson, A., Koller, A. P., and Rees, D. I. (2019). The affordable care act and ambulance response times. *Journal of Health Economics*, 67:102213.

Dell, M., Jones, B. F., and Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, 52(3):740–798.

Doyle, J. J., Graves, J., and Gruber, J. (2019). Evaluating measures of hospital quality: Evidence from ambulance referral patterns. *Review of Economics and Statistics*.

Doyle, J. J., Graves, J. A., Gruber, J., and Kleiner, S. A. (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy*, 123(1):170–214.

European Resuscitation Council, E. R. C. (1992). Guidelines for advanced life support. *Resuscitation*, 24:111–121.

Finkelstein, A., Gentzkow, M., and Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics*, 131(4):1681–1726.

Gruber, J., Hoe, T. P., and Stoye, G. (2018). Saving lives by tying hands: The unexpected effects of constraining health care providers. *NBER Working Paper Series*, (w24445).

Hollenberg, J., Riva, G., Bohm, K., Nordberg, P., Larsen, R., Herlitz, J., Pettersson, H., Rosenqvist, M., and Svensson, L. (2009). Dual dispatch early defibrillation in out-of-hospital cardiac arrest: The SALSA-pilot. *European Heart Journal*, 30(14):1781–1789.

Hull, P. (2018). Estimating hospital quality with quasi-experimental data. *Working Paper*, (SSRN 3118358).

Jena, A. B., Mann, N. C., Wedlund, L. N., and Olenski, A. (2017). Delays in emergency care and mortality during major us marathons. *New England Journal of Medicine*, 376(15):1441–1450.

Ludwig, G. G. (2004). EMS response time standards. *Emergency Medical Services*, 33(4):44.

Mendis, S., Puska, P., and Norrving, B. (2011). *Global atlas on cardiovascular disease prevention and control*. World Health Organization, Geneva, Swizerland.

Murphy, K. M. and Topel, R. H. (2006). The value of health and longevity. *Journal of Political Economy*, 114(5):871–904.

Nichol, G., Detsky, A. S., Stiell, I. G., O'Rourke, K., Wells, G., and Laupacis, A. (1996). Effectiveness of emergency medical services for victims of out-of-hospital cardiac arrest: A meta-analysis. *Annals of Emergency Medicine*, 27(6):700–710.

Phillips, D. P., Jarvinen, J. R., Abramson, I. S., and Phillips, R. R. (2004). Cardiac mortality is higher around Christmas and New Year's than at any other time. *Circulation*, 110(25):3781–3788.

Pons, P. T., Haukoos, J. S., Bludworth, W., Cribley, T., Pons, K. A., and Markovchick, V. J. (2005). Paramedic response time: Does it affect patient survival? *Academic Emergency Medicine*, 12(7):594–600.

Pons, P. T. and Markovchick, V. J. (2002). Eight minutes or less: Does the ambulance response time guideline impact trauma patient outcome? *The Journal of Emergency Medicine*, 23(1):43–48.

Skinner, J. and Staiger, D. (2015). Technology diffusion and productivity growth in health care. *Review of Economics and Statistics*, 97(5):951–964.

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear IV regression. *NBER Technical Working Paper*, (w0284).

Swor, R. A. and Cone, D. C. (2002). Emergency medical services advanced life support response times: Lots of heat, little light. *Academic Emergency Medicine*, 9(4):320–321.

Thornes, J. E., Fisher, P. A., Rayment-Bishop, T., and Smith, C. (2014). Ambulance call-outs and response times in Birmingham and the impact of extreme weather and climate change. *Emergency Medicine Journal*, 31(3):220–228.

Transportation Research Board, T. R. B. (2015). *Highway capacity manual.* The National Academies of Sciences, Engineering and Medicine, Washington DC, Washington.

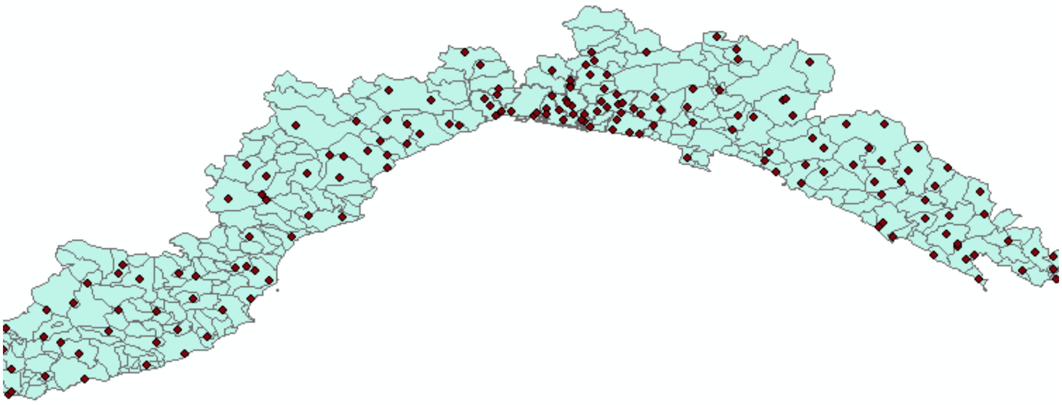
Viscusi, W. K. and Aldy, J. E. (2003). The value of a statistical life: A critical review of market estimates throughout the world. *Journal of Risk and Uncertainty*, 27(1):5–76.

Wilde, E. T. (2009). Do emergency medical system response times matter for health outcomes? *Working paper, doi: 10.1.1.688.2602.*

Wilde, E. T. (2013). Do emergency medical system response times matter for health outcomes? *Health Economics*, 22(7):790–806.

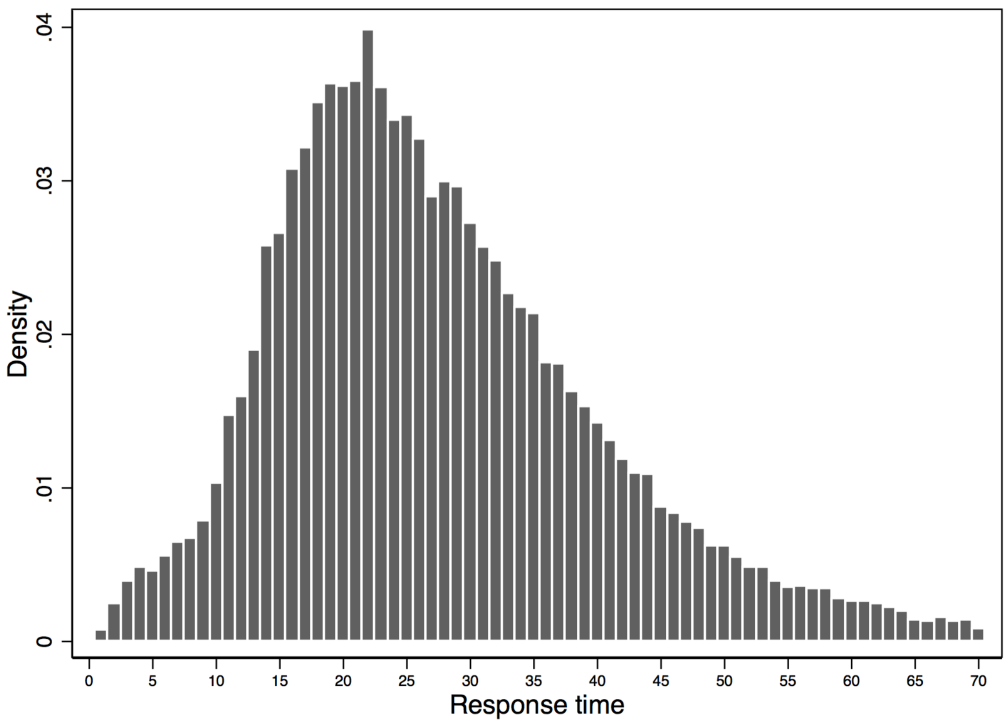
Figures and Tables

Figure 3: Map of the municipalities in Liguria and locations of the weather stations.



NOTES: There are 233 + 9 municipalities in Liguria. The additional 9 municipalities are obtained by splitting the municipality of Genova in its 9 districts, in order to obtain similarly-sized municipalities. The red dots indicate the locations of the 213 weather stations in Liguria.

Figure 4: Response time distribution



Notes: Empirical probability distribution function of response time in the sample. Response time is expressed in minutes and is the time from the start of the emergency call to the arrival of the ambulance at the site.

Table 1: Descriptive Statistics

Variables	Mean	Std. Dev.
H1: Level 1 (%)	0.03	1.62
H1: Level 2 (%)	7.48	26.30
H1: Level 3 (%)	47.38	49.93
H1: Level 4 (%) = Morbidity (M1)	45.11	49.76
H2: Level 1 (%)	0.34	5.84
H2: Level 2 (%)	14.38	35.09
H2: Level 3 (%)	65.09	47.67
H2: Level 4 (%)	16.14	36.79
H2: Level 5 (%) = Mortality (M2)	4.04	19.69
Response time	28.22	14.24
Rainfall (%)	14.10	34.80
Rainfall (mm)	1.43	2.50
High Priority Dispatch (%)	91.98	27.16
Patient age: 50-79 years (%)	49.74	50.00
Patient age: 80+ (%)	37.14	48.32
Patient Gender: Male (%)	49.82	50.00
Distance (km)	20.80	22.09
Type of Ambulance: Advanced Life Support	18.68	38.97
Population density: High (%)	48.36	49.97
Population density: Medium (%)	43.99	49.64
Population density: Low (%)	7.64	26.57
Day of the week: Monday (%)	14.85	35.56
Day of the week: Tuesday (%)	14.70	35.41
Day of the week: Wednesday (%)	14.12	34.82
Day of the week: Thursday (%)	14.11	34.82
Day of the week: Friday (%)	14.58	35.29
Day of the week: Saturday (%)	13.67	34.36
Day of the week: Sunday (%)	13.96	34.66
Number of Observations	30,149	

Notes: The outcomes of interest are the degree of severity observed on the ambulance's arrival on the scene (H1) and on the arrival at the hospital (H2). Response time is expressed in minutes.

Table 2: Effect of rainfall on out-of-hospital response time measured from the time of the call to the ambulance's arrival on the scene and from departure from the scene to arrival at the hospital.

	Way there (1)	Way back (2)
Rainfall (mm)	0.34***	0.12
	(0.09)	(0.12)
Mission characteristics	✓	✓
Time controls	✓	✓
Patient characteristics	✓	✓
Location characteristics	✓	✓
Observations	30,149	30,149

NOTES: The results are for the full set of covariates: ambulance dispatch characteristics (contact center that managed the call, priority, dummy for Advanced Life Support (ALS) ambulance, number of kilometers driven by the ambulance and its square), time fixed effects (weekday, holiday, year), individual demographics (gender and age group) and location fixed effects (municipality and population density). Column (1) presents the effect of rainfall on response time (i.e., the time required to reach the patient). Column (2) shows the effect of rainfall on the duration of the return trip (i.e., to get the patient to the hospital). Clustered standard errors at the time, day and municipality level appear in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3: Effect of ambulance response time on patient's severity (M1)

Dependent variable:	OLS (M1)	FS (RT)	ITT (M1)	IV (M1)
Response Time	0.009***			0.015**
	(0.0002)			(0.0073)
Rainfall (mm)		0.341***	0.005*	
		(0.0856)	(0.0027)	
High Priority Dispatch	0.383***	2.421***	0.405***	0.368***
	(0.0051)	(0.2616)	(0.0051)	(0.0186)
Type of Ambulance: ALS	0.238***	7.789***	0.309***	0.190***
	(0.0076)	(0.2465)	(0.0076)	(0.0576)
Distance (km)	-0.001**	0.170***	0.001***	-0.002
	(0.0002)	(0.0093)	(0.0002)	(0.0013)
Distance (km)^2	0.000***	-0.000***	0.000	0.000**
	(0.0000)	(0.0001)	(0.0000)	(0.0000)
Gender: Male	0.078***	0.576***	0.084***	0.075***
	(0.0051)	(0.1539)	(0.0053)	(0.0067)
Age: 50-79	0.056***	4.140***	0.094***	0.031
	(0.0077)	(0.2338)	(0.0080)	(0.0313)
Age: 80+	0.015*	5.656***	0.066***	-0.020
	(0.0081)	(0.2390)	(0.0083)	(0.0421)
Day of the week: Tuesday	0.004	-0.239	0.001	0.005
	(0.0092)	(0.2835)	(0.0097)	(0.0094)
Day of the week: Wednesday	0.016*	-0.046	0.015	0.016*
	(0.0094)	(0.2895)	(0.0099)	(0.0095)
Day of the week: Thursday	0.019**	-0.084	0.018*	0.020**
	(0.0094)	(0.2827)	(0.0098)	(0.0095)
Day of the week: Friday	-0.007	-0.138	-0.008	-0.006
	(0.0093)	(0.2893)	(0.0097)	(0.0094)
Day of the week: Saturday	0.005	0.360	0.008	0.003
	(0.0094)	(0.2965)	(0.0099)	(0.0100)
Day of the week: Sunday	-0.007	-0.563	-0.011	-0.003
	(0.0180)	(0.5810)	(0.0186)	(0.0190)
Public Holiday	0.018	1.047**	0.027	0.011
	(0.0159)	(0.5193)	(0.0165)	(0.0185)
Year	-0.004	-0.685***	-0.010*	0.000
	(0.0051)	(0.1559)	(0.0053)	(0.0071)
Population density: Medium	-0.912***	-10.289***	-0.560***	-0.402***
	(0.2024)	(1.1275)	(0.0384)	(0.0848)
Population density: Low	-0.744***	-1.409*	-0.310***	-0.289***
	(0.2007)	(0.8365)	(0.0252)	(0.0277)
Contact Center 2	-0.028	12.497	0.086	-0.105
	(0.1706)	(17.4182)	(0.3224)	(0.1240)
Contact Center 3	0.165	-6.695	0.105	0.207
	(0.1593)	(5.6771)	(0.1932)	(0.1512)
Contact Center 4	-0.677***	7.713	-0.605***	-0.723***
	(0.1990)	(6.6581)	(0.2191)	(0.2023)
Contact Center 5	0.533***	5.611***	0.584***	0.498***
	(0.0197)	(0.6572)	(0.0203)	(0.0458)
Municipality FE	✓	✓	✓	✓
Observations	30,149	30,149	30,149	30,149
F statistic				15.92

NOTES: The results are for the full set of covariates: ambulance dispatch characteristics (contact center that managed the call, priority, dummy for Advanced Life Support (ALS) ambulance, number of kilometers driven by the ambulance and its square), time fixed effects (weekday, holiday, year), individual demographics (gender and age group) and location fixed effects (municipality and population density). The columns show, respectively, the simple ordinary least square estimates (OLS), the first stage (FS), the intention to treat (ITT) and the instrumented (IV) results. The first-stage F-statistic is reported at the bottom of the table. Clustered standard errors at the time, day and municipality level appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Effect of ambulance response time on patient's mortality rate (M2)

Dependent variable:	OLS (M2)	FS (RT)	ITT (M2)	IV (M2)
Response Time	0.003***			0.007**
	(0.0001)			(0.0037)
Rainfall (mm)		0.341***	0.002**	
		(0.0856)	(0.0012)	
High Priority Dispatch	0.026***	2.421***	0.034***	0.016*
	(0.0019)	(0.2616)	(0.0019)	(0.0093)
Type of Ambulance: ALS	0.043***	7.789***	0.067***	0.010
	(0.0040)	(0.2465)	(0.0041)	(0.0290)
Distance (km)	-0.004***	0.170***	-0.003***	-0.004***
	(0.0002)	(0.0093)	(0.0002)	(0.0007)
Distance (km)^2	0.000***	-0.000***	0.000***	0.000***
	(0.0000)	(0.0001)	(0.0000)	(0.0000)
Gender: Male	0.014***	0.576***	0.016***	0.012***
	(0.0022)	(0.1539)	(0.0022)	(0.0032)
Age: 50-79	0.001	4.140***	0.014***	-0.016
	(0.0023)	(0.2338)	(0.0024)	(0.0155)
Age: 80+	0.027***	5.656***	0.044***	0.003
	(0.0029)	(0.2390)	(0.0029)	(0.0210)
Day of the week: Tuesday	-0.004	-0.239	-0.005	-0.003
	(0.0039)	(0.2835)	(0.0040)	(0.0041)
Day of the week: Wednesday	-0.002	-0.046	-0.003	-0.002
	(0.0040)	(0.2895)	(0.0041)	(0.0042)
Day of the week: Thursday	-0.001	-0.084	-0.002	-0.001
	(0.0040)	(0.2827)	(0.0041)	(0.0042)
Day of the week: Friday	-0.001	-0.138	-0.001	-0.000
	(0.0039)	(0.2893)	(0.0040)	(0.0041)
Day of the week: Saturday	-0.003	0.360	-0.002	-0.004
	(0.0041)	(0.2965)	(0.0042)	(0.0045)
Day of the week: Sunday	-0.004	-0.563	-0.005	-0.001
	(0.0084)	(0.5810)	(0.0086)	(0.0092)
Public Holiday	0.002	1.047**	0.005	-0.003
	(0.0077)	(0.5193)	(0.0079)	(0.0092)
Year	-0.007***	-0.685***	-0.009***	-0.004
	(0.0022)	(0.1559)	(0.0022)	(0.0033)
Population density: Medium	0.048	-10.289***	0.155***	0.230***
	(0.1126)	(1.1275)	(0.0168)	(0.0421)
Population density: Low	-0.094	-1.409*	0.041***	0.051***
	(0.1121)	(0.8365)	(0.0119)	(0.0135)
Contact Center 2	-0.046	12.497	-0.008	-0.098
	(0.5595)	(17.4182)	(0.6125)	(0.4885)
Contact Center 3	-0.152	-6.695	-0.172	-0.123
	(0.1084)	(5.6771)	(0.1153)	(0.1058)
Contact Center 4	-0.240**	7.713	-0.216*	-0.272**
	(0.1113)	(6.6581)	(0.1165)	(0.1137)
Contact Center 5	-0.151***	5.611***	-0.134***	-0.175***
	(0.0093)	(0.6572)	(0.0094)	(0.0232)
Municipality FE	✓	✓	✓	✓
Observations	30,149	30,149	30,149	30,149
F statistic				15.92

NOTES: The results are for the full set of covariates: ambulance dispatch characteristics (contact center that managed the call, priority, dummy for Advanced Life Support (ALS) ambulance, number of kilometers driven by the ambulance and its square), time fixed effects (weekday, holiday, year), individual demographics (gender and age group) and location fixed effects (municipality and population density). The reported results are, respectively, ordinary least square (OLS), first stage (FS), the intention to treat (ITT) and the instrumented (IV) estimations. The first-stage F-statistic is reported at the bottom of the table. Clustered standard errors at the time, day and municipality level appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Effect of ambulance response time on patient severity (M1) by set of included covariates. Response time is instrumented by amount of rainfall at the hourly and municipality level.

	(1)	(2)	(3)	(4)	(5)
Response Time	0.015**	0.017**	0.017**	0.018**	0.018**
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Population density: Medium	-0.402***	0.016	0.020	0.026	
	(0.085)	(0.057)	(0.056)	(0.057)	
Population density: Low	-0.289***	0.337***	0.330***	0.244**	
	(0.028)	(0.084)	(0.084)	(0.099)	
Gender: Male	0.075***	0.086***	0.085***		
	(0.007)	(0.010)	(0.010)		
Age: 50-79	0.031	0.043	0.041		
	(0.031)	(0.036)	(0.035)		
Age: 80+	-0.020	-0.009	-0.012		
	(0.042)	(0.046)	(0.045)		
Day of the week: Tuesday	0.005	0.004			
	(0.009)	(0.010)			
Day of the week: Wednesday	0.016*	0.019*			
	(0.010)	(0.010)			
Day of the week: Thursday	0.020**	0.023**			
	(0.010)	(0.010)			
Day of the week: Friday	-0.006	-0.002			
	(0.009)	(0.010)			
Day of the week: Saturday	0.003	0.008			
	(0.010)	(0.011)			
Day of the week: Sunday	-0.003	0.005			
	(0.019)	(0.020)			
Public Holiday	0.011	0.008			
	(0.018)	(0.019)			
Year	0.000	-0.003			
	(0.007)	(0.006)			
High Priority Dispatch	0.368***				
	(0.019)				
Type of Ambulance: ALS	0.190***				
	(0.058)				
Distance (km)	-0.002				
	(0.001)				
Distance (km)^2	0.000**				
	(0.000)				
Contact Center 2	-0.105				
	(0.124)				
Contact Center 3	0.207				
	(0.151)				
Contact Center 4	-0.723***				
	(0.202)				
Contact Center 5	0.498***				
	(0.046)				
Municipality FE	✓	✓	✓	✓	
Observations	30,149	30,149	30,149	30,149	30,149
F-statistic	15.92	14.24	14.91	13.34	13.27

NOTES: Column (1) reports the estimations of the baseline specification, which include the full set of controls: ambulance dispatch characteristics (contact center that managed the call, priority, dummy for Advanced Life Support (ALS) ambulance, number of kilometers driven by the ambulance and its square), time fixed effects (weekday, holiday, year), individual demographics (gender and age group) and location fixed effects (municipality and population density). Columns (2)-(5) report the results by excluding, respectively, controls for ambulance dispatch characteristics, time fixed effects, individual demographics and location fixed effects. The first-stage F-statistic is reported at the bottom of the table. Clustered standard errors at the time, day and municipality level appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Effect of ambulance response time on patient mortality (M2) by set of included covariates. Response time is instrumented by amount of rainfall at the hourly and municipality level.

	(1)	(2)	(3)	(4)	(5)
Response Time	0.007**	0.008**	0.007*	0.007*	0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Population density: Medium	0.230***	0.028	0.024	0.024	
	(0.042)	(0.028)	(0.027)	(0.027)	
Population density: Low	0.051***	-0.150***	-0.153***	-0.144***	
	(0.013)	(0.044)	(0.043)	(0.051)	
Gender: Male	0.012***	0.007	0.008		
	(0.003)	(0.005)	(0.005)		
Age: 50-79	-0.016	-0.021	-0.019		
	(0.015)	(0.018)	(0.018)		
Age: 80+	0.003	0.000	0.003		
	(0.021)	(0.024)	(0.023)		
Day of the week: Tuesday	-0.003	-0.004			
	(0.004)	(0.004)			
Day of the week: Wednesday	-0.002	-0.003			
	(0.004)	(0.004)			
Day of the week: Thursday	-0.001	-0.001			
	(0.004)	(0.004)			
Day of the week: Friday	-0.000	-0.000			
	(0.004)	(0.004)			
Day of the week: Saturday	-0.004	-0.004			
	(0.005)	(0.005)			
Day of the week: Sunday	-0.001	-0.002			
	(0.009)	(0.010)			
Public Holiday	-0.003	-0.003			
	(0.009)	(0.010)			
Year	-0.004	-0.013***			
	(0.003)	(0.003)			
High Priority Dispatch	0.016*				
	(0.009)				
Type of Ambulance: ALS	0.010				
	(0.029)				
Distance (km)	-0.004***				
	(0.001)				
Distance $(\text{km})^2$	0.000***				
	(0.000)				
Contact Center 2	-0.098				
	(0.488)				
Contact Center 3	-0.123				
	(0.106)				
Contact Center 4	-0.272**				
	(0.114)				
Contact Center 5	-0.175***				
	(0.023)				
Municipality FE	✓	✓	✓	✓	
Observations	30,149	30,149	30,149	30,149	30,149
F-statistic	15.92	14.24	14.91	13.34	13.27

NOTES: Column (1) reports the estimations of the baseline specification, which include the full set of controls: ambulance dispatch characteristics (contact center that managed the call, priority, dummy for Advanced Life Support (ALS) ambulance, number of kilometers driven by the ambulance and its square), time fixed effects (weekday, holiday, year), individual demographics (gender and age group) and location fixed effects (municipality and population density). Columns (2)-(5) present the results by excluding, respectively, controls for ambulance dispatch characteristics, time fixed effects, individual demographics and location fixed effects. The first-stage F-statistic is reported at the bottom of the table. Clustered standard errors at the time, day and municipality level appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Balancing test: mean and standard deviation of covariates with respect to rainfall (with vs without).

Variables	Wo/Rainfall (1)		W/Rainfall (2)		Difference (3)	p-value (4)	
	Mean	SD	Mean	SD			
Response Time	28.08	14.12	28.92	14.58	-0.83	0.000	***
High priority dispatch (%)	91.84	27.38	92.66	26.08	-0.82	0.067	*
Gender: Male (%)	49.75	50.00	49.99	50.01	-0.24	0.769	
Age: < 50 (%)	13.16	33.81	12.89	33.51	0.28	0.622	
Age: 50-79 (%)	49.76	50.00	49.64	50.00	0.13	0.880	
Age: 80+ (%)	37.08	48.30	37.48	48.41	-0.40	0.617	
Distance (km)	15.80	19.75	15.99	20.57	-0.19	0.571	
Type of ambulance: ALS (%)	18.74	39.02	18.58	38.90	0.16	0.803	
Population density: High (%)	48.26	49.97	50.65	50.00	-2.38	0.004	***
Population density: Medium (%)	44.43	49.69	40.94	49.18	3.49	0.000	***
Population density: Low (%)	7.31	26.03	8.42	27.77	-1.11	0.011	**
Observations	25,896		4,253				

NOTES: Column (1) reports mean and standard deviation of covariates in the absence of rainfall; column (2) reports mean and standard deviation of covariates in the presence of rainfall; column (3) and (4) report, respectively, the difference between the means and the p-value of the difference. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Alternative outcome specifications and ordered probit estimation results for the effect of response time (RT) on patient's condition on arrival at the scene (H1).

	(1)	(2)	(3)	(4)	(5)
Response Time (RT)	0.015**	0.002	0.017**	-0.015	0.016
	(0.007)	(0.002)	(0.008)		
Observations	30,149	30,149	30,149	30,149	30,149
F-statistic	15.92	15.92	15.92		
Average RT				28	28
Average age				70	70
Average distance				16	16

NOTES: The estimations reported in columns (1)-(3) include the full set of controls: ambulance dispatch characteristics (contact center that managed the call, priority, dummy for Advanced Life Support (ALS) ambulance, number of kilometers driven by the ambulance and its square), time fixed effects (weekday, holiday, year), individual demographics (gender and age group) and location fixed effects (municipality and population density). Column (1) reports the baseline results where the outcome is a dummy equal to 1 for highest level of severity (level 4) and zero for levels 1, 2 or 3. Column (2) reports the results when the outcome is a dummy equal to 1 for severity level 3 or 4 and zero otherwise. Column (3) reports the results for the linear model where the outcome range is the level of severity (from 1 to 4). Columns (4) and (5) present the marginal effects obtained by estimating the ordered-probit model (the set of covariates does not include fixed effects at the municipality level in order to allow convergence, and lower degrees of severity (1 and 2) are grouped together due to the low number of observations for degree 1. The three parameters in the ordered probit estimation sum up to 1 and therefore the results of only two are presented: the effect of RT on severity of degree 3 (in column 4) and on severity of degree 4 (in column 5). The table also reports the first-stage F-statistic and the average values of RT, age and distance driven by the ambulance at which the maximum likelihood estimates are calculated. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: Alternative outcome specifications and ordered probit estimation results for the effect of response time (RT) on a patient's condition on arrival at the hospital (H2).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
RT	0.007**	0.018**	0.011**	0.052***	-0.002	0.010	0.006
	(0.004)	(0.007)	(0.005)	(0.011)			
Obs	30,149	30,149	30,149	30,149	30,149	30,149	30,149
F-stat	15.92	15.92	15.92				
Average RT					28	28	28
Average age					70	70	70
Average distance					16	16	16

NOTES: The estimations reported in columns (1)-(3) include the full set of controls: ambulance dispatch characteristics (contact center that managed the call, priority, dummy for Advanced Life Support (ALS) ambulance, number of kilometers driven by the ambulance and its square), time fixed effects (weekday, holiday, year), individual demographics (gender and age group) and location fixed effects (municipality and population density). Column (1) reports the baseline results where the outcome is a dummy equal to 1 for highest level of severity (level 5) and zero for levels 1, 2, 3 or 4. Column (2) reports the results when the outcome is a dummy equal to 1 for severity level 4 or 5 and zero otherwise. Column (3) reports the results when the outcome is a dummy equal to 1 for severity level 3, 4 or 5 and zero otherwise. Column (4) reports the results for the linear model where the outcome range is the level of severity (from 1 to 5). Columns (5)-(7) present the marginal effects obtained by estimating the ordered-probit model (the set of covariates does not include fixed effects at the municipality level in order to allow convergence and lower degrees of severity (1 and 2) are grouped together due to the low number of observations for degree 1. The four parameters in the ordered probit estimation sum up to 1 and therefore the results of only three are presented: the effect of RT on severity of degree 3 (in column 5); the effect of RT on the severity of degree 4 (in column 6); and on severity of degree 5, i.e. the mortality rate (in column 7). The table also reports the first-stage F-statistic and the average values of RT, age and distance driven by the ambulance at which the maximum likelihood estimates are calculated. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 10: Descriptive Statistics: ambulance driving times (ADT) on the way to the scene (from the time of dispatch until arrival on the scene) and on the way back to the hospital (from the time of departure from the scene to arrival at the hospital).

Variables	Obs	Mean	Median	Std. Dev.
ADT on the way to the scene	30,149	15.86	13.22	12.20
ADT on the way to the scene	27,729	15.22	12.80	11.28
ADT on the way back to the hospital	27,729	11.51	8.52	15.61

NOTES: The first row reports for the full sample. The second and the third rows report only for the observations with complete records for both the way there and the way back. Driving times are expressed in minutes.

Table 11: Average difference in driving time between the trip to the scene and the trip back.

Dependent variable: Driving time on the return trip	
Go	3.70***
	(0.12)
Constant	11.51***
	(0.09)
Observations	27,729

NOTES: The parameter *Go* is the estimated difference in driving times between the trip to the scene and the trip back. Robust standard errors appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1

7 Appendix

7.1 Sample construction

The dataset includes 43,713 observations. 4,171 observations were omitted because one of the following was missing: dispatch priority, response time (RT), health outcomes (H1 or H2), gender or age of the patient. Outlier observations were omitted: age over 100 years (125 observations) and the 99th percentile of RT, i.e. 90 minutes or greater (3,836 observations). Finally, in case where multiple EMS vehicles reached the scene only the response time of the first vehicle is included and as a result 5,432 observations were omitted.[45] The final sample consists of 30,149 observations.

Table 12: Sample construction

Omitted observations	Observations
Raw data	43,713
Missing values[1]	39,542
Patients over 99 years-old	39,417
Truncation at 99th percentile of response time	35,581
Ambulance not first on the scene	30,149

NOTES: [1] Missing values include: dispatch priority, response time (RT), health outcome (H1 or H2), gender or age of the patient.

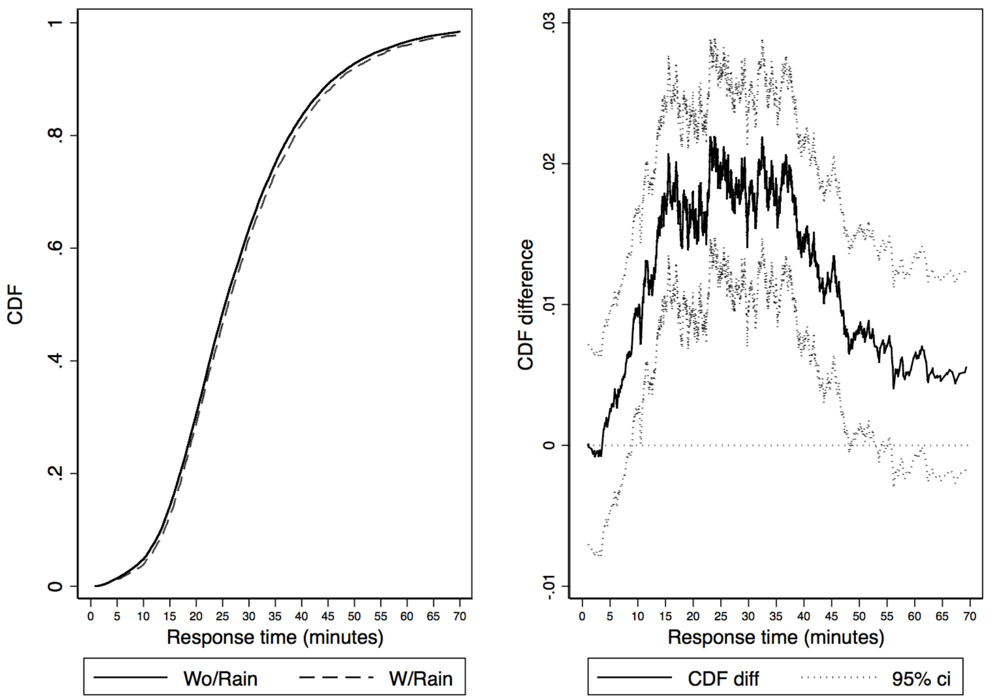
7.2 Internal Validity: Monotonicity

The interpretation of the estimation results as local average treatment effects (LATEs) requires monotonicity, that is, stochastic dominance of response time distributions with respect to rain. This condition is tested using the approach of Angrist and Imbens (1995). The left panel of Figure 5 reports the conditional

[45] Data are collected for each ambulance run. Information on the total number of vehicles and their characteristics is reported at the patient level and multiple records are omitted.

distribution functions (CDF) of response time with and without rainfall (dashed line and solid line, respectively) while the right panel shows the difference between the two CDFs and its confidence intervals. As can be seen, stochastic dominance is fulfilled since the two CDFs never cross. The difference between the CDFs reveals the compliers, which are concentrated in the densest part of the response time distribution (see Figure 4 for response time distribution).

Figure 5: Cumulative distribution function (CDF) of response time with and without rainfall and the difference between them.



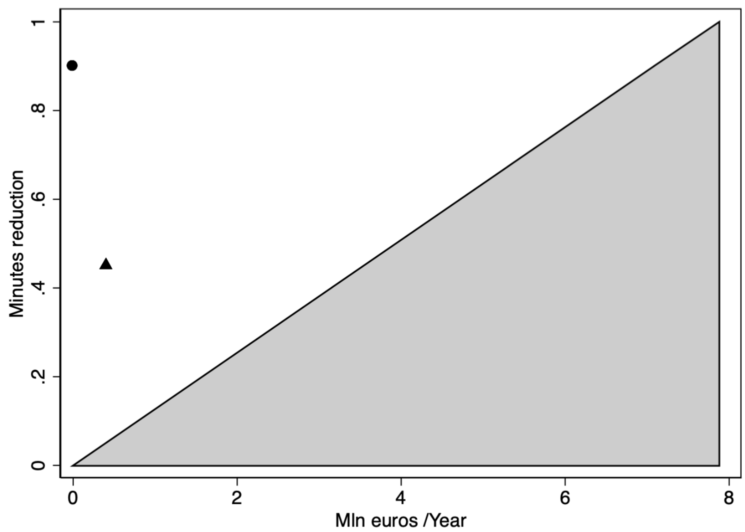
NOTES: The left panel shows the CDF of response time with rainfall (dashed line) and without (solid line). The right panel shows the difference between the CDFs and the 95% confidence intervals.

7.3 Cost effectiveness of increasing the number of available ambulances

Following Wilde (2009), Pons and Markovchick (2002), and Pons et al. (2005), the cost of one additional staffed ambulance is assumed to be 450,000 euros per year. The benefit from one additional ambulance in terms of reducing response time in Liguria is calculated based on the following: (i) the average distance driven by an ambulance in Liguria is 10.4 kilometers per incident; (ii) the average driving time from the ambulance's dispatch to arrival on the scene is 15.86 minutes; (iii) each day there are an average of about 41 ambulance runs that involve a cardiac event. If we assume that the additional ambulance reduces the average distance driven by all ambulances, average distance would be reduced from 10.4 to 10.15 kilometers (i.e. 41*10.4/42) in the best case scenario. On average, an ambulance travels one kilometer in 1.5 minutes. As a result, reducing the average distance by 0.25 kilometers would reduce average response time by 0.4 minutes.

Figure 6 compares the cost effectiveness of adding an ambulance to that of installing technology to improve communication during an emergency call. As can be seen the latter option is far more worthwhile.

Figure 6: Policymaker's indifference curve and two alternative policies.



NOTES: The vertical axis measures the reduction in response time while the horizontal axis measures annual cost. The solid diagonal line represents the policymaker's willingness-to-pay to obtain a given response time reduction in the case of patients experiencing a cardiac event. The dot represents the cost-benefit ratio associated with introducing a technology that improves communication during the emergency call while the triangle represents the cost-benefit ratio associated with the acquisition of one additional ambulance.