THE QUALITY AND EFFICIENCY OF PUBLIC AND PRIVATE FIRMS: EVIDENCE FROM AMBULANCE SERVICES *

Daniel Knutsson & Björn Tyrefors†

July 1, 2021

Abstract

Economic theory predicts that outsourcing public services to private firms will reduce costs, but the effect on quality is ambiguous. We explore quality differences between publicly and privately owned ambulances in a setting where patients are as good as randomly assigned to ambulances with different ownership statuses. We find that privately owned ambulances perform better in response to contracted quality measures but perform worse in response to noncontracted measures such as mortality. In fact, a randomly allocated patient has a 1.4% higher risk of death within 3 years if a private ambulance is dispatched (in aggregate, 420 more deaths each year). We also present evidence of the mechanism at work, suggesting that private firms cut costs at the expense of ambulance staff quality.

JEL Codes: P48, H44, I11, D22, D44, L33.

*We thank the Swedish Competition Authority and Handelsbanken's Research Foundations for financial support and Andreas Madestam, Andrei Shleifer, Ekaterina Zhuravskaya, Erik Grönqvist, Erik Lindqvist, Johannes Haushofer, Jonas Vlachos, Magnus Johannesson, Mikael Priks, Per Pettersson-Lidbom, and the seminar participants at IFN and Stockholm University for their helpful comments and suggestions. We are further indebted to Ola Kostenniemi (at SOS Alarm AB) and Gunnar Ljunggren (at SLL) for helping us access the data. Moreover, Ann-Catrin Eriksson (SLL) provided us with important information on firms and contracts. We wish to thank all of you for your valuable assistance.

†Corresponding author: Tyrefors: Research Institute of Industrial Economics (IFN), Box 55665, 102 15 Stockholm (e-mail, bjorn.tyrefors@ifn.se; telephone: +46(0)8-665 4500. and Dep. of Economics, Stockholm University); Knutsson: IFN, P.O. Box 55665, SE-10215 Stockholm, Sweden (e-mail, daniel.knutsson@ifn.se; telephone: +46(0)8-665 4500).

1 Introduction

Whether the private sector can provide public services more efficiently than the public sector is a central question in economics. Outsourcing to private firms is a common way for governments to supply goods and services to their citizens, and in many OECD countries, welfare services are increasingly being contracted out to private providers. Among OECD countries, on average, 3% of the GDP is used for public procurement that is not consumed by the government itself, e.g., procurement conducted through outsourcing (OECD, 2019). Despite this strong reliance on outsourcing, theoretically, it is not a panacea for public sector inefficiencies.

When private firms have "residual control rights" over the assets used to produce goods and services, they have incentives for cost-saving innovation that the public sector does not (Hart and Moore, 1990). However, the canonical literature also highlights circumstances where private firms have incentives to reduce costs by shirking on quality (Hart et al., 1997). Cost innovation at the expense of quality is more likely to appear when quality cannot be adequately contracted on, innovation pressure is low, competition is weak, consumer choice is inefficient, and reputational mechanisms are weak (Shleifer, 1998).

Thus, the advantage of outsourcing public services is ultimately an empirical question, but credible evidence is scant.[1] Regarding experimental evidence, the few studies that examine the effects of privatization or outsourcing in areas, such as job placement or vocational rehabilitation, point to very small, if any, effects (Bennmarker et al., 2013; Behaghel et al., 2014; Laun and Thoursie, 2014). The evidence on outsourcing in healthcare is concentrated in nursing home markets, and the evidence is mixed (Bergman et al., 2016; Barron and West, 2017; Wübker and Wuckel, 2019; Duggan et al., 2018). The limited number of papers available reflects the inherent problems of self-selection of patients into providers and the lack of good measures of quality in these markets. With self-selection of patients to providers, any given difference in measures of quality could reflect patient choices (and therefore, patient characteristics) rather than the quality of the service delivery itself.

This paper furthers the literature by providing a credible empirical evaluation and comparison of quality between public and private firms in an acute healthcare environment.[2] Analyzing the efficiency of emergency care is of economic importance, as aggregate spending on emergency care could be as high as 10% of national health expenditures (Lee et al., 2013).

[1]For primarily descriptive evidence on outsourcing, see Szymanski and Wilkins (1993); Szymanski (1996) for garbage collection, Bales et al. (2005); Duwe and Clark (2013); Powers et al. (2017); Spivak and Sharp (2008) for prisons, and Bayer and Pozen (2005); Armstrong and MacKenzie (2003) for residential youth care.

[2]Jansson and Castwall (2017) use part of one year of our data as a pilot for a bachelor's thesis developed under the supervision of Tyrefors. They only investigate efficiency measures and not health outcomes.

To overcome the problem of self-selection, we use the ambulance services in Stockholm as a laboratory to answer questions regarding outsourcing in healthcare. From an evaluation point of view, the setting has important features. After auctions based on competitive tendering in 2011, two private ambulance firms were granted a 5-year contract at a fixed annual reimbursement to operate the ambulance services side by side with one public firm.[3] After the auction, there were no further entry or exit within 5 years, and most importantly, there was no ex post competition by patient choice. Dispatch operators assign an ambulance to each patient by considering the distance to the patient and the alternative use of the unit but without regard for ownership status. According to their guidelines, the nearest ambulance should be assigned, but to maintain readiness for emergencies, exceptions can be made for less severely injured patients. As the ambulances that we study have the same capacity to handle different health emergencies, dispatch operators have no incentives to systematically discriminate between providers. In contrast to the scenario found in much of the literature, our setting offers an advantage in that the provider type is chosen independently of the patients' characteristics and without concern for the patients' provider preferences.[4]

Our first set of results relates to firm-specific ambulance outcomes of each patient serviced. These outcomes are all explicitly contracted on and well documented, with fines ultimately assessed if inferior performance is observed. In terms of contracted, salient outcomes, we find that private firms are more efficient. Private firms were 8% faster in responding to a dispatch and 7% faster in reaching the patient. Since firms were compensated at a fixed rate, these results are not generated by incentives to take up more assignments.[5] We interpret these findings as private firms being more innovative than public firms when outcomes are observable and contracted on.

In our second set of results, we find that private ambulances produced lower quality with respect to noncontracted outcomes (i.e., mortality). This mortality penalty for private firms was already consistently higher from the first day of service and up to three years later. In our preferred model, we estimate an effect of a 0.42 percentage point (pp) increase in mortality for privately serviced patients. This amounts to a 1.4% higher risk of dying within three years of service from a baseline mortality of 31%. A back-of-the-envelope calculation for a given year shows that in Stockholm alone, approximately 420 additional patients each year died within 3 years due to the ownership status of the ambulance dispatched to them.

[3] Andersson et al. (2019), upon reviewing the outsourcing literature, claim that retaining some in-house provision is a common arrangement in practice.

[4] Although our setting is beneficial from an evaluation point of view, without further assumptions, it may not be informative about private provision with ex ante competition, where consumers can assess the quality or if markets are competitive, as discussed by Hart et al. (1997).

[5] During the first three years of the sample, the contracts included a very small variable component based on historical patient flows. The results are, however, not sensitive to excluding these years.

This figure is higher than the number of all traffic deaths in all of Sweden in a single year by more than 50%. Our calculations suggest that the cost of private firms in a given year in terms of loss of quality adjusted life years (QALYs) could be substantial. In relation to the total cost of the ground-based ambulance services in Stockholm, a conservative estimate puts the annual loss of QALYs at 50% of the total annual cost and exceeds considerably the cost savings from outsourcing ambulances to private firms in Stockholm. Considering the two sets of results together, we argue that our findings provide strong support for the theoretical predictions outlined in Hart et al. (1997).

To better understand the mechanisms behind these quality differences, we analyze outcomes that indicate differences in how ambulance crews serve their patients at site. When arriving at the patient site, the staff must first assess the medical condition of each patient. This assessment starts with a thorough interview with the patient (and/or bystander) to assess his or her medical history and current symptoms. Then the ambulance crew members diagnose the patient, asses the severity, and decide on conveyance. We find that private firm ambulances assessed their patients as having less severe conditions, diagnosed more generic conditions, and assessed fewer patients as having cardiovascular diseases (CVDs) than their public counterparts, although the baseline characteristics are ex ante balanced across providers. These findings are intriguing, given that private firm patients have higher mortality in both the short and long term. However, for conditions that are difficult to misdiagnose, such as cardiac arrest, there are no differences across provider types. Consistent with these assessments, private firm ambulances left 32% more patients at home (3.2 pp out of an average of 10%). Moreover, if a patient is conveyed, the admission to a hospital is based on a new hospital assessment. As a result, the patients conveyed by private ambulances had 1.2 pp *fewer* hospital admissions and consistent with the under-diagnosing at site by the crew, many fewer patients were admitted with a CVD as the primary intake diagnoses.

That private firms are leaving more patients at home cannot be explained by financial incentives, as ambulance firms are not paid per assignment. Instead, we argue that our results can be explained by private firms cost innovating at the expense of their staff, thereby attracting and retaining on average lower quality staff than the public firm. This hypothesis is corroborated by descriptive and anecdotal evidence. In follow-up assessments of each firm from 2017 there are indications that the private firms have, on average, both a higher turnover rate and a higher fraction of their staff on temporary employment contracts (SLL, 2017). Additional information from matched employer-employee data at the regional level suggests that private firms have employees who have less experience, change workplaces more often, and have lower grades but higher annual income, although some of the differences lack statistical power. Further information from collective agreements, reports, and newspaper articles suggest that the annual wage difference is not explained by a higher hourly wage but instead by longer workdays, extensive use of overtime, and less on-the-job training.

Our confidence that we have estimated the causal effects on efficiency and quality is supported by a credible design. Although treatment is assigned without concern for ownership status, the identification of causal effects in this setting relies on the fact that the probability of being serviced by a private ambulance is unrelated to a patient's health status. However, this assumption can still be violated in two dimensions that can also be related to patient health. First, and most importantly, the location a patient calls from can affect his or her probability of being serviced by a private ambulance. Patients living close to a private ambulance station and hence having a different probability of private firm service could differ from those residing close to a public station. Second, even conditional on patient location, the timing of the emergency call can be related to the probability of private ambulance service. Differences in both the demand for and supply of ambulances over time and thereby the probability of private ambulance service could be related to patient health characteristics or the general health environment. To make causal inferences about outsourcing in this setting, a credible empirical strategy must account for both geography and time.

To estimate the causal effects of ambulance providers, we employ a fixed effects model with both narrowly defined geographical fixed effects and flexible time fixed effects. Using the GPS coordinates of each patient's location at the time of the call, we create geographical fixed effects of block level size, which measures approximately 190-by-140 meters, to ensure that patient treatment probability is unrelated to patient location. Thus, we compare patients who are experiencing medical emergencies and are living in the same residential block. Ambulance assignment between private and public firms is thereby determined in a stochastic way by ambulance availability and ambulance distance to the patient at the time of the call.

The notion that we can credibly identify the causal effects of provider type on quality is verified by a test of balance and correspondence of criteria with the assignment center. There is no evidence of a systematic relation between the assignment of the type of provider and patient characteristics.[6] We also show that under a series of regression model specifications and context-relevant exclusions of data, the estimated effects are very close to those of our main specification. For example, we find similar results when considering districts mainly operated by public and private ambulance firms. This suggests that the mechanism producing our findings is not specific to private (public) ambulances operating close to public (private) ambulance stations.

[6]Importantly, our results do not hinge on a specific set of fixed effects but are consistent across a wide variety of different specifications.

Our findings contribute to the literature on outsourcing and quality, surveyed recently in Andersson et al. (2019), and support the conclusion that the ability to contract quality is paramount to the successful outsourcing of public services. Our paper adds new evidence to the small literature on quality outcomes from private firms providing healthcare for the public sector. Using mortality as the outcome, Bergman et al. (2016) show that private nursing homes in Sweden produced better quality than public nursing homes. In Germany, Wübker and Wuckel (2019) show that for common diagnoses, for-profit hospitals performed no worse—or even better—than public hospitals. One clear difference from these studies is that our results emerge in a context where consumers cannot choose their provider, which could explain the difference.

Of further relevance to the present study is the fact that few papers have studied staff productivity, which can be crucially important in healthcare. Bartel et al. (2014) find that the nurses' education and on-the-unit experience increased productivity in hospital environments. Our findings suggest that even important health outcomes can be susceptible to healthcare staff quality. Moreover, Gruber and Kleiner (2012) find that during the nurses' strikes in New York, in-hospital mortality increased by 18%, even when hospitals engaged temporary staff. Similarly, Gruber et al. (2018) studied physician practices in emergency departments (EDs) when a new regulation imposed a strict ceiling on waiting time before seeing a physician. As patients were cared for earlier, more patients were admitted to a hospital ward, and mortality decreased by 14%. Our findings confirm the susceptibility of the group of patients seeking emergency care to small changes in staff practices on mortality.[7]

In markets where for-profit firms are at risk of shirking their responsibility for service quality, a possible solution is to incentivize not-for-profit firms to produce services (Bennett and Iossa, 2009). Not-for-profit private firms have the potential to credibly commit to non-contracted soft incentives while maintaining intrinsic motivation and keeping costs down (Glaeser and Shleifer, 2001; Francois, 2003). Several studies on nursing homes have documented that private for-profit firms perform worse than private not for-profit firms (Tanuseputro et al., 2015; Grabowski et al., 2013). The ambulance market could be a suitable context for not-for-profit firms if public in-house production is inadequate.[8]

[7] Other evidence on the importance of healthcare resources is provided in Doyle (2011); Doyle et al. (2015). The consequences of EMS staff fatigue is explored in Patterson et al. (2012), Rivard et al. (2020) consider how multiple jobs and workload affect the emergency medical services (EMS) employees job satisfaction, and Rogers et al. (2004) analyze the relationship between long working hours and medical errors. For descriptive evidence on mortality among non-conveyed ambulance patients, see Lederman et al. (2020); Ebben et al. (2017). Regarding mortality, several studies document the importance of early detection, treatment, and adherence to treatment for CVDs (Korhonen et al., 2017; Orchard et al., 2018; Rothwell et al., 2007; Lindahl et al., 2017).

[8] However, our results do not relate to the performance of *not*-for-profit firms, as there are no such firms competing for ambulance contracts in Stockholm.

The remainder of the paper is organized as follows. The next sections address the background and institutional details, followed by section three, where we describe the methodology and empirical design. In section four, the results are presented, and in section five, we provide further analyses on mechanisms, heterogeneity, and robustness. In section six, we conclude the paper. Throughout the paper, the appendix figures and tables are available in the Online Appendix.

2 Background

2.1 The ambulance market in Stockholm

The Stockholm region is the largest healthcare region in Sweden, with 2.3 million inhabitants in 2017. The ambulance service is completely financed by the public sector (through regional income taxes) but with firms competing for contracts in lowest bid auctions spanning approximately six years.[9] The total costs of ambulance ground transports were approximately 36 million EUR in 2010, when 55 ambulances took on 145,000 assignments (Hälso och sjukvårdsförvaltningen, 2011; HSF, 2018). From 2010, reflecting the increasing population of Stockholm, the number of ambulances increased over time through supplementary contracts. In 2016, the inhabitants of the region of Stockholm were serviced by 73 ambulances that responded to 220,000 dispatches at a total cost of 52 million EUR.

From being completely run by a public firm, in 1993, competition was introduced into ambulance services. Stockholm County was divided into 7 commercial sectors for which the firms competed. Several firms have entered and disappeared over time, and in 2006, it was decided that the public firm was to be freed from competition and provided with two sectors, located in the central and western parts of Stockholm, that were not up for auction. Private firms have since competed for five sectors, with each firm allowed to operate at most in three out of the five sectors (AISAB, 2020). In 2009, in addition to the public firm, three private firms operated in five sectors. Only one auction was held during our period of study (in 2011). During this auction, one of the firms lost its contract, meaning that only two private firms remained. Currently, the two remaining private firms still comprise the market structure, although the auctions are open for entry and other firms have attempted to compete (Online Appendix Figure 1 shows the locations of the stations during this period).

In the auctions before 2011, winning the right to operate a sector was a function of being able to comply at the lowest cost with the rather precisely specified contract. The contract specified a fine if the response times after dispatch calls were too slow. The ambulance crews

[9] The contracts last for five years, with the additional possibility of an extension of up to two years. Auctions for ground-based ambulances were held in 2005, 2011, and 2017.

were to be ready to depart within 90 seconds after a dispatch for an emergency (priority 1) and within 180 seconds for other assignments (priority 2 or 3). Fines could also be generally applied if the contracted ambulances were not operational as required (on repair without a replacement car or without staff) or if staffing requirements were not fulfilled (at least one masters degree (specialist) nurse in the ambulance crew). The transport time to each patient was carefully monitored but not tied to a monetary fine. The ambulance firms were compensated with a fixed and a variable component. The fixed amount, allowing firms to maintain ambulances, staff, and garages, was slightly adjusted on a yearly basis for changes in the number of dispatches compared to those of the previous year. More dispatches thereby led to marginally higher compensation (SLL, 2005).

From 2011, the county of Stockholm started to weight quality as 25% of the bid, but the same structure was maintained with respect to fines. The variable compensation was also removed, and the requirements to gradually increase the use of renewable vehicle fuels were tied to fines. The maximum fine that firms could be required to pay was 3% of the contract value (Landstingsrevisorerna, 2013). Fines were not conditioned on health outcomes, mortality, or any other direct quality measure. The contracts also specified the number of ambulances in each sector, their equipment, and their service time availability. Additionally, since 2008, together with at least one trained paramedic, a registered nurse with a master's degree in nursing was required to be present on an emergency ambulance.[10] When studying the continuous evaluations between the region of Stockholm and ambulance providers, these factors are mentioned, measured, and discussed.

Table 1 describes the annual contracted service hours and the outcome of the auction in 2011, i.e., the annual compensation. The private firms were contracted on approximately 5,300 hours per week, which comprises the bulk of the hours serviced (67%). They received an annual compensation of about 25.7 million EUR in 2012 vs. 15.9 million EUR for the public provider. The average cost per hours serviced was 116.7 EUR for the public provider

[10] The contracts included other types of units. From 2009 to 2012, these other units consisted in part of several transport ambulances that were staffed only by paramedics and mainly responded to priority 3 dispatches where medical treatments were not required. There were also two emergency cars in Stockholm. These vehicles included a nurse with a specialist degree in anesthesiology and were not able to transport patients; they were only allowed to assist an ambulance when the patient was severely ill or if there was a need to intubate. Furthermore, two helicopters were contracted, of which one was available during the summer months (Hälso och sjukvårdsförvaltningen, 2011). Under the new contract from 2012, the ambulance system changed somewhat. An emergency car with a specialist physician onboard was instituted. The two emergency cars that were already operational were upgraded to have patient transportation capacity. This meant that these units could take on emergencies alone if required and transport patients to hospitals. The staffing of these vehicles was unchanged, and they were still intended to be used in medical emergencies and in support of other ambulance units. From 2012, helicopters were supplied by a separate private firm that was not operating ground transports, while emergency ambulance cars were supplied by private firms (HSF, 2018).

but about 24% lower, namely, 93.9 EUR, for the private providers. A normal ambulance transport is staffed with one nurse and one paramedic, and the labor cost for an hour is approximately 75 EUR for both, about 75% of the compensation. Thus, taking the public provider cost per hour as baseline, the cost saving from using private providers is about 6.2 million EUR annually.

Table 1: The cost of public and private ambulances

Provider	**Hours/week**	**Share**	**Annual compensation (EUR)**	**Cost/hour (EUR)**
Public	2,621	33%	15,900,000	116.7
Private	5,274	67%	25,741,026	93.9

Note: The authors' calculations from the contracts and corresponding documents describing the auction in 2011. We convert exchange rates by, 1EUR=10SEK.

2.2 The geography of ambulance dispatch and transport in Stockholm

The ambulance dispatch firm in Stockholm, SOS Alarm AB, has two main objectives. First, patients should be prioritized such that they can be served based on the severity of their condition. Second, patients are matched to ambulances with as little waiting time as possible before the ambulance arrives while still maintaining sufficient readiness for other medical emergencies. These objectives are crucial for critically ill patients with a serious condition, e.g., an ongoing heart attack, and for whom time to treatment is of the essence (SOSAAB, 2020).[11]

When calling the emergency number, 112, the caller is connected to an emergency operator. Based on the need for assistance from the police force, fire brigades, or ambulance services, the caller is connected to a specialist (normally not a health professional) to further assess the emergency. Once the specialist has made an assessment of the patient needs (priority), an ambulance work description is sent out, including the priority of the emergency, patient characteristics, characteristics of the injury, and patient location. A dispatch operator then takes the job and assigns the patient to an ambulance according to the main objectives. If the patient's condition is sufficiently severe, several units can be assigned the same patient.[12] The dispatch operator observes the patient's location and priority when assigning an ambulance to the patient. Sometimes the dispatch operator has more information, especially when there are critically ill patients or many injured. According to SOS

[11] Information regarding SOS Alarm AB and how it operates has primarily been collected through personal contacts with employees in the firm.

[12] Multiple unit dispatches are rare (4% in the data). We exclude these in the robustness analysis (Figure 4 row 5) and use multiple ambulance dispatches as a pre-determined characteristic in our balance test.

Alarm AB, the assignment should immediately be directed to the closest ambulance if the patient's condition is life threatening (priority 1) or potentially life threatening (priority 2). The assignment can be placed on hold if no suitable units are available and the patient's condition is not affected by a time delay (priority 3). In reality, cases with priority 2 or 3 can be placed on hold to facilitate readiness for emergencies (RiR, 2012).

Within each firm-specific sector, there are several ambulance stations. These stations, containing one or more ambulances, are connected to a geographical area in which they are primarily responsible for transporting patients, i.e., an ambulance district. If the local ambulance is unavailable in its district, e.g., if it is occupied by another patient or relatively far away from the patient, another ambulance is called in to service the patient based on proximity. Dispatch operators are explicitly not limited by the boundaries formed around each ambulance station, and the operators can freely engage ambulances as they find appropriate. An ambulance can be dispatched to any location in the county, although very long transport times to patients are rare. In essence, whether a patient is picked up by the local ambulance or by one from an other station depends on the relative distance and availability of that ambulance compared to others nearby.

Importantly, two patients residing in the same neighborhood have the same assigned hospital. Each ambulance takes them to their designated hospital, determined by their residence. Exceptions can be made if the patient is currently under treatment at a different hospital or if specialist care elsewhere is required. However, this is not at the discretion of the ambulance crew. Therefore, the distance to hospital or hospital resources are not mechanisms that can explain any results that we find, in contrast to other studies, e.g., Doyle et al. (2017).

3 Empirical design and data

3.1 Empirical design

Estimating the causal effects of outsourcing public health services to private firms on quality outcomes is inherently challenging. The main challenge comes from patients being able to sort themselves into different health providers based on social and material resources, such as income, education, personal contacts, and residence. Furthermore, the patients' health may determine by whom and at what location they receive care.

The ambulance services that we study provide some clear benefits with respect to this type of sorting. Patients cannot sort themselves into either a private or a public ambulance. The dispatch operator assigns with certainty the patient to an ambulance based on criteria that the patients themselves cannot affect. The dispatch operators are responsible for minimizing the waiting time for patients while ensuring that the time for a future patient in distress is minimized (readiness for emergencies). This is done manually by observing a digital map on which ambulance resources and patients are marked. For severe cases (priorities 1 and 2), distance (or time to patient) is the sole ambulance allocation mechanism. Furthermore, ambulances, in general, do not differ in terms of resources. All ambulances have a fixed set of equipment and requirements of formal staff quality.[13]

However, even if patients cannot choose ambulance providers, they choose where to live, and providers operate more frequently in certain areas. The public firm mainly services the inner city and a few nearby districts. The private firms mainly service the suburbs and distant cities, villages, and rural communities in the larger Stockholm region. These populations are likely very different in terms of many characteristics. A naive comparison of outcomes by public-private status is hence likely to yield a biased estimate due to noncomparable treatment and control groups. Fortunately, ambulance assignment is not bound by the district borders. This means that almost all patients have a nonzero probability of being serviced by a private (public) ambulance regardless of residency.

To identify the causal effects of the ownership status of ambulances, we rely on the cross-district movements of ambulances. There are several different situations that give rise to cross-border movements. First, the district-specific ambulance can already be servicing a patient and thus be unavailable for a new assignment. Second, an ambulance from a different district might be closest to the patient and thereby the best choice given that the patient is in need of urgent assistance. Third, maintaining readiness for future emergencies might require assigning a different ambulance than the district-specific ambulance. These situations are unpredictable and unrelated to patient health given the priority set by the emergency operator and the location of the patient.

To be able to use movements of ambulances across districts, we adopt a nonparametric approach by controlling for neighborhood fixed effects. By using detailed information on each patient's location with GPS coordinates, we construct a grid of squares covering the full county of Stockholm. The patients requesting an ambulance from a given grid cell are assigned to a group variable that we use as fixed effects. With these fixed effects, we essentially compare patients who live very close to one another. Our baseline grid cells are approximately 190 x 130 meters, resulting in approximately 16,000 effective geographical fixed effects with variation in the ambulance provider.[14] As shown in Miller et al. (2019), the

[13]Before 2012, only ambulances operating around the clock were equipped with ECG monitoring capabilities, not all ambulances, but the ambulances so equipped were not exclusive to any particular provider.

[14]Theoretically, we can estimate up to 720,000 fixed effects. However, many of these grids will not be covered within the county, will not be populated, or will only contain a single observation. Our baseline model estimates around 32,000 fixed effects, while the number of grids with private-public common support

chosen level of fixed effects may not only reflect a trade-off between bias and variance if the treatment status only varies within some groups and the treatment effects are heterogeneous. In the Online Appendix, Figure A6, we show that our results are not sensitive to varying the size of the grid cells.

These patients, by virtue of their residence, have a very similar probability of being serviced by a private (public) ambulance. By holding the probability of a private ambulance constant, we use the quasi-random variation in ambulance assignment generated by idiosyncratic shocks to local ambulance demand and ambulance locations at the time of the call. In other words, the theoretical experiment that we propose is as follows. Two patients living in the same neighborhood call for an ambulance for a similar condition at approximately the same time. Due to which ambulance that happens to be nearest and the restriction on readiness for emergencies, either a private or public ambulance is dispatched. When we reduce the size of the geographical unit in which we compare patients, the probability of a patient being assigned a private ambulance essentially converges to a constant, and the assignment of ambulances in terms of ownership status should not be related to patient health.

This design provides further benefits for identification. Comparing patients living near one another means that the patients not only face a similar probability of treatment but also should be transported to the same hospital from approximately the same transport distance. Therefore, the distance to the hospital or the hospital resources are not mechanisms that can explain any results that we find. Moreover, patients residing near one another are similar on a range of dimensions, allowing us to estimate precise coefficients, as our fixed effects can explain much of the residual variation.

In addition to geographical fixed effects, we include time dummies (year and month dummies) in our main specifications, and we include dummies for how urgent the dispatch operator determines the assignment to be. Transport urgency is part of the observable treatment allocation mechanism, and it may be important to hold it constant. Similarly, with time effects, we can hold constant changes in the treatment probabilities over time that result from changes in supply and demand (e.g., introducing new ambulances to ambulance stations). Although we consider more flexible fixed effects in the robustness section, we refrain from doing so in our main specification, as overspecifying a model with fixed effects can have other adverse effects (Miller et al., 2019).[15] Including these control variables, our

is smaller at approximately 16,000 grids.

[15]To be able to estimate a large set of fixed effects efficiently, we use the STATA command *reghdfe* (Correia, 2014).

main model to be estimated by OLS is:

$$\text{Outcome}_{igy} = \alpha_g + \beta * \text{Private}_{igy} + \text{Year}_y + \text{Month}_y + \text{Priority}_{igy} + \epsilon_{igy} \tag{1}$$

In the above equation, the outcome of individual i in grid cell g at time y is causally related to the ownership status of the ambulance servicing the patient. Clearly, the aim of this analysis is to mimic a random allocation of ambulances to patients. Assuming that the treatment is as good as randomly allocated within the grid cells, with a different but within-group constant probability of treatment, we have to control for fixed effects at the group level (see, e.g., Krueger, 1999). A testable implication of our design is that we expect to see similar characteristics between patients in the same grid cell serviced by either private or public firms.

To test whether patients serviced by private or public firms are similar before dispatch, we use predetermined variables on the left-hand side in equation (1) and estimate the coefficients on the *Private* dummy variable (the treatment). The estimated coefficients are now the differences in these characteristics between patients serviced by a public and a private ambulance conditional on the control variables included. The predetermined variables that we consider are either measured before the study period begins in 2007-2008 (hospitalizations and outpatient visits) or at the time of ambulance dispatch (age, sex, multiple units, previous ambulance transport, and residence in a nursing home). To further show that our design solves inherent imbalances between patients serviced by private and public ambulances, we also display correlations between the predetermined variables and the *Private* dummy, i.e., the results from bivariate regressions.

With respect to the standard errors, we follow Abadie et al. (2017). On the one hand, our patients are randomized to treatment (resembling a randomized control trial); therefore, there is no general need for clustering. On the other hand, our design is based on comparing patients within small but fixed geographical grids. Since ambulance stations are fixed in space, it is not unreasonable that there is some positive, however not perfect, correlation of treatment assignment within grids. Abadie et al. (2017), shows that if there are also heterogeneous treatment effects in this type of setting, a clustering adjustment can be motivated. Since the latter cannot be ruled out, we choose to be conservative and cluster our standard errors at the fixed effects level, although we acknowledge that our standard error will be too large. However, our choice of the level of clustering turns out not to be important for our results. In the Online Appendix, we show the sensitivity when clustering on different grid sizes (Figure A6), the non-clustered standard errors, clustering at the patient levels (patients are sometime dispatched multiple times in our sample), and randomization inference (Figure A7).

3.2 Data and variables

Important for this study is that we observe and evaluate contracted and noncontracted quality indicators. Contracted indicators are time stamps in the data describing the time interval of the ambulance dispatch during all stages. Slow firm average response times or slow times to patient are parameters under contract and are associated with fines or feedback in follow-up meetings. Health outcomes are noncontracted and possibly noncontractible since contracting requires large amounts of data and measures that are robust to provider moral hazards (Gupta, 2017). In our data, we observe more than 1.1 million ambulance assignments for close to 500,000 unique patients, where almost 60% of the assignments are provided by private firms. These data include all inpatient and outpatient visits, mortality, and importantly, information determined prior to the ambulance dispatch.

The data come from the VAL database, which originates from Region Stockholm. Region Stockholm is responsible for all publicly financed healthcare in greater Stockholm and maintains health registries for the research, evaluation, and monitoring of the care that it finances. We make use of the ambulance registry and pool all ambulance dispatches between 2009 and 2016. An ambulance assignment, i.e., an observation, is only registered in the database if a patient was present on the scene. With the ambulance data, we merge several other registries from the same database, including the registries for inpatient care, outpatient visits, and mortality. Finally, we are able to merge the patient GPS position data from the dispatch services, SOS Alarm AB, and we can uniquely merge this data to each ambulance assignment ID number.

We have information on time stamps for each dispatch, i.e., the response time to a dispatch, time to patient, time at patient, time to hospital, and time at hospital.[16] Since some of these are important quality indicators and are contracted on, they are used as outcomes to investigate how public and private firms respond differently to contracted incentives.

We make some data restrictions in the analysis by focusing attention only on emergency ambulances that are generally comparable in terms of staff and equipment requirements. This means that we are excluding transport ambulances (which were phased out from 2009), intensive care ambulances (only one unit that was publicly owned), and a few other rare ambulance types that are directed to particular patients (including helicopters, which are strictly privately owned). Since the vast majority of units during this time are regular emergency ambulances, this exclusion amounts to a loss of 38,500 observations. Our final sample that we use consists of at most 1,141,939 observations. For further information on the data, see the Online Appendix, Section 6.

[16] In the original VAL database, these variables had many missing observations on some of the time stamps. See the Online Appendix, Section 6, for more information on the time stamps.

The geographical position system information we have for each patient is used to generate the grid cell fixed effects. To do this, we first transform the GPS data (latitude/longitude) to Cartesian coordinates by using the STATA command *geo2xy* (Picard and Stepner, 2015). We then use the two positions that describe and include the entire region of Stockholm (northwest and southeast) and generate the grid as X times X equally sized rectangles between these points by grouping observations in each cell (the region of Stockholm is around 160 by 110 kilometers in size). We do this for several different values of X, but our baseline model includes 850 times 850 grid cells, with each cell measuring 190 by 130 meters, which is about the size of a residential block.

As our main dependent variable on health, we use mortality. Since we have the date of death for all ambulance patients, we create several variables indicating whether the patient died within a certain length of time after each ambulance dispatch. The time periods that we consider are one day, one week, one month, three months, six months, one year, two years, and finally three years after each dispatch.

Although we present our main results on mortality for all time frames that we consider, we will more generally use a 3-year mortality time frame throughout the paper. This choice is not motivated by a particular expected health mechanism but more reflects the variety of conditions ambulance crews face on a daily basis. Trauma and medical emergencies (such as acute cardiovascular conditions, car accidents, and drug overdoses) have a higher likelihood of mortality close in time to the emergency call. Other ambulance dispatches due to geriatric conditions or milder symptoms, which are common in the data, will have a higher likelihood of mortality with a longer delay, as many of the patients are of old age. Three-year mortality provides a complete measure of quality in outsourcing and enables the consideration of all kinds of conditions and types of patients. Not evaluating longer mortality measures is motivated by data availability and by the fact that in the long run, treatment effects on mortality between any two groups must converge.

In the Online Appendix, Table A1 shows the descriptive statistics for the predetermined variables of the data used in the analysis. Forty-five percent of the observations in the data include patients with a previous ambulance contact within one year, suggesting that the average serviced patient is prone to acute illness or has a severe chronic health condition. This can be explained by the ambulance serviced population being old, with the average patient being 62 years old and 7% of patients residing in a nursing home. The patients in this population were also sick; 10% had a diabetes diagnosis, and 7% had a chronic obstructive pulmonary disease (COPD) diagnosis in 2007-2008. This picture is further reinforced in the Online Appendix, Table A2, where the outcomes are described. Ambulance-serviced patients have close to a 31% three-year mortality rate and a 6% one-month mortality rate. Thus, ambulance-serviced patients are often old, sick, and frail. Minor changes in their care or treatment could thereby have vast implications for their health and survival.

4 Results and discussion

4.1 Balance

As noted in the Online Appendix, Figure 1, the ambulance stations are not randomly allocated, and the average person serviced by a private ambulance could be different from the average person serviced by a public ambulance. However, we argue that any imbalance across groups will reflect geographical differences in the probability of being serviced by a private ambulance, which are correlated with differences in patient characteristics. If this is the case, any observed differences should vanish if conditioned on fixed effects, based on sufficiently small geographical units (grids).

To emphasize this argument, Figure 1 and Figure 2 show the results from balance tests. In these figures, we use 36 predetermined characteristics as outcomes, including the most common diagnoses we observe in the patient data in 2007-2008.[17] In the left panel of both figures, we show the results when no grid fixed effects or other controls are included, i.e., the results from a simple bivariate regression. In the right panels of the same figures, we use our design-based approach, namely, the full specification of Equation (1).

Starting with the left panel of Figure 1, we conclude that the patients for whom a private ambulance was dispatched are indeed unconditionally different. For example, private ambulance-serviced patients are 0.05 standard deviations younger than public ambulance-serviced patients. This amounts to a raw average age difference of almost 1.25 years. Reflecting this age difference, private ambulance-serviced patients seem, on average, to be healthier than public ambulance-serviced patients. They were less likely to live at a nursing home and had fewer healthcare contacts in 2007-2008. Although private ambulance-serviced patients appear unconditionally healthier, they are more often allocated more than one ambulance. Being assigned more than one ambulance or assigned an additional unit with a specialist competence is a strong indication that a life-threatening event has taken place.[18] The left panel of Figure 2 shows a more complicated picture, with private ambulance-serviced patients having more documented lifestyle-related diagnoses (such as diabetes, hypertension, and hyperlipidemia) but fewer diagnoses related to substance abuse.

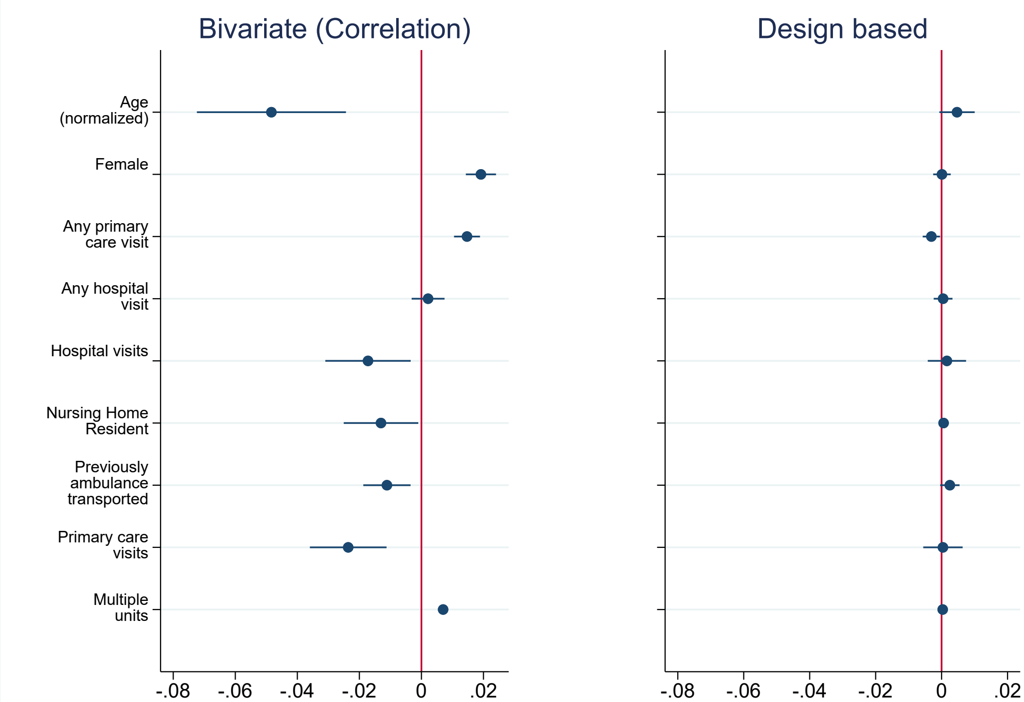
[17] We create 5 groups by using information from the diagnoses, namely, CVDs, respiratory diseases, psychiatric diseases (including drugs and alcohol), surgery related diseases, and an indicator variable for having more than one diagnosis in 2007-2008. Thus, these variablesare functions of the other predetermined characteristics.

[18] Patients serviced by multiple units have a 15% risk of dying within one day of ambulance service.

In the right panels of Figure 1 and 2, we re-estimate the same models using our preferred specification (adding grid cell fixed effects, time dummies, and dispatch priority). Out of 36 tested hypotheses, only two, the hypotheses regarding whether the patient had any primary care contacts in 2007-2008 and whether an anemia diagnosis had been made, are significant at the 5% level. The first of these estimates is negative at 0.3 pp and suggests that, if anything, patients serviced by private firms were healthier prior to the dispatch. The second significant estimate is positive and suggests that private ambulance patients are 0.08 pp more likely to have had an anemia diagnosis in 2007-2008. However, since these estimates are very close to zero, marginally significant, and 36 hypotheses are tested, we do not interpret this as a sign of imbalance but close to what we would expect from random sampling. The other variables tested are close to zero, insignificant and, most importantly, provide support for the assumptions underlying our design that conditional on the fixed effects, ambulance firm assignment is unrelated to patient health. Furthermore, our preferred model explains much of the residual variation in the outcomes, as reflected by meaningful reductions in the standard errors on all estimates. We interpret these findings as strong support for our design's ability to address geographical selection as present in the data.[19]

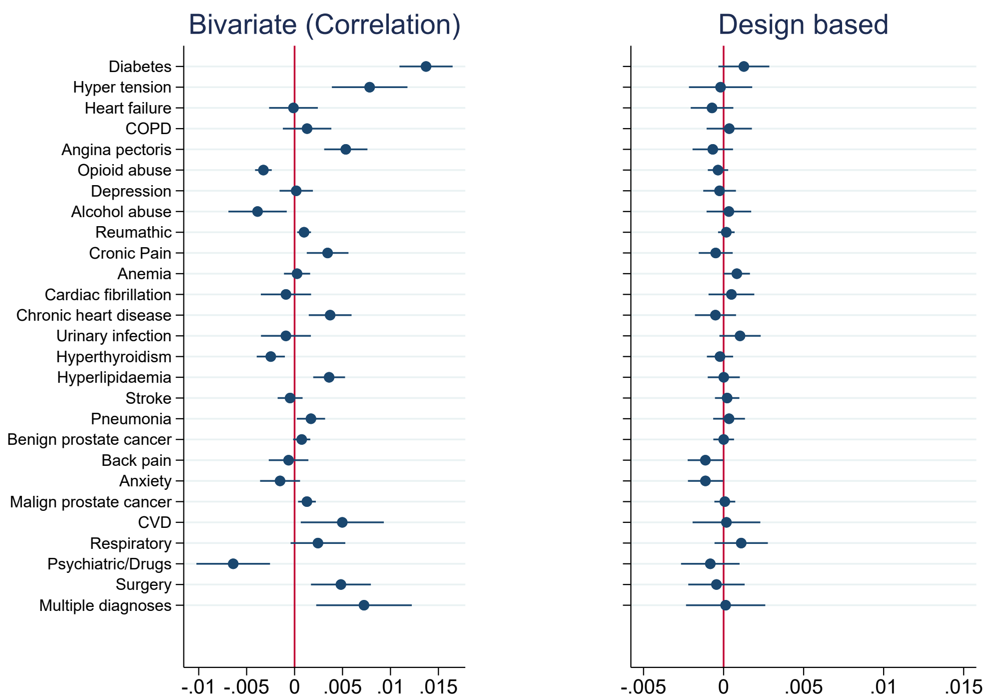
[19] As age could be the potential driver of many of the unconditional health differences, we provide kernel density plots of patient age distributions between public and private ambulance-serviced patients (see the Online Appendix, Figure A2). We show both the unconditional and conditional difference in age distributions between public and private firms. We find that the two estimated distributions are extremely close to one another once we residualize by using our main specification.

Figure 1: Balance of predetermined variables: Bivariate and design based



Note: Both panels show estimates of *Private*, with the predetermined variable indicated to the left of the figures as the outcome. The left panel shows the bivariate linear regression estimate using OLS (only the outcome and private ambulance indicator), while the right panel displays the same results using our preferred model. Several variables here are measured at the time of each ambulance transport. Age in years (normalized by demeaning and dividing by the standard deviation), sex, residing in a nursing home, dead at arrival, multiple units dispatched (to the same patient), and previously transported by an ambulance (after 2009) have this property. Primary care and hospital care visits are based on data from 2007-2008. All regression models include year, month, and dispatch priority dummies. We include 850 by 850 grid fixed effects and cluster the standard errors at the grid cell level.

Figure 2: Balance of diagnoses in 2007-2008: Bivariate and design based



Note: Both panels show estimates of *Private*, with the predetermined variable indicated to the left of the figures as the outcome. The left panel shows the bivariate linear regression estimate using OLS (only the outcome and the private ambulance indicator), while the right panel displays the same results using our preferred model. Diseases are based on ICD-10 codes registered at the hospital and primary-care visits between 2007 and 2008. The diagnoses that we choose to include are the most frequent that we could find in the data. If a diagnosis is documented, the patient indicator variable for that disease will be 1. If no indication, it is 0. All regression models include year, month, and dispatch priority dummies. We include 850 by 850 grid fixed effects and cluster the standard errors at the grid cell level (also in the bivariate regressions).

4.2 Effects on contracted outcomes

We obtain information on two outcomes that are perfectly observable and contracted or reviewed: the response time after the dispatch call and the ambulance time to the patient. Again, we present the full specification of Equation (1). In Table 2, column 1, we show that private ambulances respond 8 seconds faster – a decrease in the response time of nearly 8%. The standard errors in brackets show that the estimate is highly statistically significant, with a t-statistic of more than 10. This outcome is contracted and entails fines if performance is sufficiently poor. We further find, as reported in column 2, that private firms reach their patients faster. Private firms are 61 seconds (or approximately 8%) faster, which is a considerable amount of time given that a time delay is a key detrimental factor in the health emergency literature (Jena et al., 2017; Lucchese, 2020). Again, the estimate is highly statistically significant. This outcome is not contracted with fines as the ultimate consequence but is an important indicator that is followed over time and discussed during follow-up meetings. Thus, the results confirm the predictions in Hart et al. (1997).

Table 2: Contracted outcomes

	Response time to a dispatch	Travel time to patient
	(1)	(2)
Private ambulance	-8.3122	-61.1357
	(0.8752)	(5.2931)
Public outcome mean	107.6759	716.5211
Observations	1000902	1017910

Note: Each column display results from a separate regression following Equation 1. The two outcomes are measured in seconds between two time stamps. Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters in size, as well as year and month dummies and priority of patient dummies. Standard errors are clustered at each grid cell.

4.3 Effects on noncontracted outcomes: mortality

The ultimate quality measure for health services is often mortality. Again, we present the specification of Equation (1), now using mortality as the outcome. In Table 3, column 1 displays the effect of a private ambulance being dispatched instead of a public ambulance on mortality within 1-day of service. Among private ambulance-serviced patients, the mortality within one day is significantly higher (0.1 pp). The effect increases rather monotonically over time.[20] That the effect is increasing over time suggests worse health outcomes not only

[20]See Online Appendix, Figure A3, for a graphical illustration.

Table 3: Noncontracted outcomes: Mortality

	Mortality after ambulance transport occurred within:							
	1 day	1 week	1 month	3 months	6 months	1 year	2 years	3 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Private ambulance	0.0010	0.0009	0.0014	0.0023	0.0011	0.0026	0.0041	0.0042
	(0.0003)	(0.0005)	(0.0008)	(0.0010)	(0.0011)	(0.0013)	(0.0014)	(0.0015)
Public outcome mean	0.0148	0.0294	0.0580	0.0954	0.1290	0.1758	0.2477	0.3066
Observations	1075958	1075958	1075958	1075958	1075958	1075958	1075958	1075958

Note: Each column display results from a separate regression. The outcomes are dummy variables indicating if the patient had died within each time window. Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters in size, as well as year and month dummies and priority to patient dummies. Standard errors are clustered at each grid cell.

for critically ill patients, who have a high probability of dying soon after the emergency call is made, but also for patients with milder symptoms that are more likely to die later on. There is also no significant increase when evaluating the difference between 2- and 3-years mortality, consistent with the fact that the treatment effect must decrease at some point when evaluating mortality over time.

We find the largest absolute effect when evaluating 3-year mortality. The effect size is 0.42 pp, which (at a mean of 31%) yields an increase of 1.4%. To understand the magnitude of this effect, we use the following example. Considering that private firms service approximately 100,000 patients each year, the estimated effect suggests that private ambulance service led to an additional 420 deaths (within three years) each year. This is a considerable amount; for example, traffic accidents in Sweden cause a total of 200-300 deaths each year. In the Online Appendix, Section 4, we calculate standard Quality Adjusted Life Years (QALYs) based on our estimates. Even under conservative assumptions on lives lost due to private service, the total cost is still substantial. Considering our most conservative QALY weight of 0.1, the yearly cost is still 10 million EUR or 100 million SEK, a quarter of the total cost of ground-based ambulance services in Stockholm in 2012. As discussed previously, outsourcing ambulances in Stockholm could generate cost savings of up to 6.2 million EUR. Even if we use a conservative evaluation, our calculations suggest that the cost of reduced quality vastly outweighs the cost savings that private firms generate. The aim of the next section is to understand the underlying mechanism of why private ambulances underperform in terms of quality in both the short and long run.

5 Mechanisms and heterogeneity

The superior performance of private firms on contracted and their inferior performance on the non-contracted quality measures line up well with the predictions of Hart et al. (1997). In this section, we further the analysis and argue that the results are driven by the firms' personnel policy affecting staff quality and thereby patient health. Although staff quality is regulated with respect to education and basic training, firms can still hire staff with lower experience or sustain an extensive use of overtime, potentially affecting the quality of service.

We start by showing that diagnosing at site differs substantially across providers and that this affects the probability of receiving proper care. To further understand the mechanism, we next present a heterogeneity analysis that sheds light on which types of patients drive our main results. Finally, we present anecdotal evidence from credible sources documenting fundamental differences in personnel policy between firms. From matched employer-employee data, we document staff differences related to staff quality. Moreover, we discuss competing mechanisms based on sorting unrelated to the private firms' production choices.

5.1 Immediate responses

In Table 4, we study the immediate actions taken by the staff when arriving at a patient's location. We know from Figures 1 and 2 that patients within a cell are comparable. Thus, any difference in assessment by the crew when arriving is due to crew behavior.

When arriving at the patient, the staff must first assess the medical condition of each patient. A thorough interview with the patient (and/or bystander) to assess his or her medical history and current symptoms can be combined with several clinical examinations, such as blood pressure, respiratory rate, blood sugar, ECG monitoring, pulse oximetry, and pulmonary auscultation assessments. We do not have access to information on these examinations; however, the diagnosis, as assessed by the ambulance crew, is noted in the patient's journal. These assessments are fixed in the journal system (consisting of around 160 categories) and are important, as the ambulance crew assessment motivates the actions taken (and not taken) during the assignment. For example, if the ambulance crew assesses a patient to have an acute myocardial infarction (AMI), the ambulance guidelines dictates which immediate treatments should be provided. Moreover, to prepare the ER for ambulance arrival, the information should be provided to the receiving hospital together with an ECG for further assessment by a specialist. Guidelines were available for several acute conditions, and ambulance clinicians (specialist nurses) have generous delegations of medicines to administer. However, the most important guidelines were for conditions where time is a crucial factor, such as a stroke or a cardiac arrest. Thus, making correct patient assessments is important.

Table 4: Does firm affiliation affect ambulance crew behavior?

	Diagnoses set by ambulance crew							
	Unspecified/ General	CVD	Psyciatric/ drugs	Respiratory diseases	Surgery	Cardiac arrest	Severity (higher worse)	Patient stays home
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Private ambulance	0.0139	-0.0059	-0.0006	-0.0015	-0.0011	0.0003	-0.0523	0.0322
	(0.0010)	(0.0010)	(0.0007)	(0.0009)	(0.0007)	(0.0002)	(0.0029)	(0.0011)
Public outcome mean	0.0399	0.1364	0.0594	0.0910	0.0545	0.0046	2.8507	0.1010
Observations	1075984	1075984	1075984	1075984	1075984	1075984	1061675	1075984

Note: Each column reports results from a different regression where each model includes 850x850 grid fixed effects covering the entire county of Stockholm, for which each grid is approximately 190x130 meters in size, as well as year, month, and priority to patient indicator variables. The diagnoses set by the ambulance crew are obtained from the ambulance data. The ambulance report requires a listing of the diagnosis, which is chosen from approximately 160 common conditions. In column 1, the outcome is an indicator variable denoting if the ambulance crew chose an unspecific category, e.g., "General, unspecific." In column 2, the outcome variable indicates if the ambulance crew chose a category consistent with a CVD disease, e.g., cardiac arrest, chest pain, AMI, or stroke. In column 3, we use an indicator variable for psychiatric diagnoses as the outcome. This includes all intoxications due to drugs, medicines, alcohol, and toxic chemicals. The outcome in column 4 indicates if a respiratory disease, such as COPD, the common flu, and respiratory distress, was listed as the main diagnosis. The outcome in column 5 indicates if surgery was associated with the complaints diagnosed, such as abdominal pain or ulcer. In column 6, we single out cardiac arrest from the CVD category as a standalone outcome using three listed conditions including the term "cardiac arrest." Further information on the ambulance crew diagnoses can be found in Section 6 in the Online Appendix. The outcome in column 7 is a severity measure reported by the ambulance crew that ranges from 0-7. The outcome in column 8 is a variable recorded by the crew that indicates if the patient was transported by the ambulance or left at home. Standard errors are clustered at the grid cell level.

Our first result in Table 4, column 1, suggests that private ambulance staff are making less specific assessments. Compared to the public ambulance staff, the private ambulance staff have a 35% higher likelihood of using a generic category (such as "general other") to define the patient's conditions. The natural follow up question then is what the under-assessed diagnoses are. We present four broad categories and one distinct diagnosis that are common and jointly make up approximately 30% of all patient diagnoses. These are CVDs, psychiatric/drugs, respiratory diseases, and surgery.[21] From the CVD category, we take out cardiac arrests as a stand-alone diagnosis that is difficult to mis-diagnose for the ambulance crew. Thus, we do not expect any large differences in the assessments for cardiac arrest. The results are presented in Table 4, columns 2-6. We find that private firms diagnose fewer patients as suffering from CVDs, including AMI and stroke, where timely treatment is of the essence. However, there are no significant differences for the other categories. Reassuringly, in column 6, we do not find any statistically significant difference in diagnosed cardiac arrests.

Although we have shown that private ambulance staff make different assessments regarding their patients' conditions, we do not know if they make different assessments with respect to the severity of the condition. To understand the differences in severity assessments, we use that the staff must classify each patient by using a number from 0-7 to denote the severity of the patient's condition (7 is the most severe). In column 7, using the ambulance crew assessment of the patient's severity as the outcome, we show the effect of having a private ambulance dispatched. We find that the patients of private firms are reported as being less severely injured than the patients of public firms. Although this difference is not very large (approximately 5% of a standard deviation), this finding is intriguing, given the increased mortality risk from a private ambulance dispatched in the short run. The possibility that private firm ambulances are somehow matched with more severely injured patients, even if this is at odds with our balance tests results, seems further implausible.

Finally, the assessment is also linked to the choice of conveyance (or not). As shown in Table 4, column 8, private firms more often elect to leave patients at home, consistent with their severity assessment. In total 12% of all patients are not conveyed, and our estimated effect suggests that private firms leave 3.2 pp more patients at home than public firms (a 32% increase from the public mean). We argue that this is strong support for immediate differences in actions taken upon arrival at the patient's location. Leaving more patients at home could provide a possible explanation for why we find effects on mortality immediately

[21] The psychiatric/drugs category includes all intoxications, including alcohol and other psychotic and psychiatric diagnoses. CVDs are a combination of cardiac arrests, AMI, fibrillation, heart failure, stroke, and general circulatory complaints. In the respiratory category, we include asthma, COPD, the common flu, respiratory obstructions, and other respiratory diagnoses. The surgery category includes ulcer, pain, ileus, and bleeding from the intestinal tract. See further Online Appendix Table A14.

after each dispatch. Lederman et al. (2020) studied non-conveyed patients in Stockholm in 2015 and found that among the patients who were 65 years or older, representing nearly half of their adult sample, 28% visited an ED, and 21% were hospitalized within 7 days of being left at home. Moreover, 0.4% of this older group died within 1 day of non-conveyance, and 1.1% died within 7 days. A recent literature review found that 1 day mortality ranged between 0.2% and 3.5% among non-conveyed EMS serviced patients (Ebben et al., 2017). Given the high mortality among non-conveyed elderly patients, private ambulances not conveying older patients could explain at least part of the direct effects on mortality that we estimate. To provide more direct evidence on the impact of non-conveyance, in the next section, we make use of hospital admissions data.

5.2 Hospital admissions

The difference in assessments and conveyance rates by private ambulances could be innocuous if the patients were assessed correctly or if the patients received proper treatment through later hospital visits. Even though the difference in mortality rates speaks against this, we can further investigate the importance of conveyance by evaluating hospital admissions data. Hospital admissions are based on a new assessment of severity made at the ED by a physician. We interpret differences in hospital admissions between patients serviced by public or private ambulances to be due to the difference in ambulance crew diagnosing and conveyance rates that we have found.

In Table 5, column 1, we show that patients serviced by private ambulances have a 1.2 pp lower chance of being admitted to a hospital within one day of service. This cannot be explained by mortality attrition, as the 1-day mortality is only 0.1 pp. The likelihood of being admitted to a hospital within 30 days is somewhat lower (column 2), indicating that only a small fraction of the patients are receiving additional care after not being conveyed. This means that the missed opportunity of an hospital admission and the health impacts thereof persist over time and can lead to lasting health differences. In columns 3-6, information on the primary diagnosis for the patient if admitted is used as the outcomes. For patients admitted with a CVD as the primary diagnosis, which is the most common cause of admissions, in column 4, we see a clear decreased chance of being admitted if a private ambulance is dispatched. We also find smaller decreases in admissions for diagnoses related to psychiatry and surgery, likely reflecting the large difference in conveyance rates between providers. Thus, the main message is that patients with ongoing cardiovascular events are the ones missing out on proper treatment due to a private ambulance being dispatched. CVDs comprise 36% of the decrease in hospital admissions caused by private ambulance service. This is consistent with our previous findings showing that private ambulances under-diagnose CVDs at

Table 5: Does ambulance firm affiliation affect hospital admissions?

			Primary intake diagnosis			
	Patient admitted $\leq$ 1 day	Patient admitted $\leq$ 30 days	CVD	Psychiatry/ drugs	Respiratory diseases	Surgery
	(1)	(2)	(3)	(4)	(5)	(6)
Private ambulance	-0.0123	-0.0086	-0.0043	-0.0017	-0.0008	-0.0023
	(0.0015)	(0.0014)	(0.0012)	(0.0006)	(0.0007)	(0.0007)
Public outcome mean	0.5581	0.6232	0.1065	0.0384	0.0615	0.0599
Observations	1075984	1075984	1075984	1075984	1075984	1075984

Note: Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters in size, as well as year, month, and priority to patient indicator variables. Each column describes a different regression. In column 1, the outcome is an indicator variable denoting if the patient was admitted to a hospital the same day or the day after the ambulance dispatch. We use both these days because waiting times at the ER can result in admissions taking place in the following day and we cannot determine if a transport resulted directly in an admission. In column 2, the outcome variable indicates if the patient was admitted at any time within 30 days of the ambulance dispatch. In columns 3-6, the outcomes are the main intake diagnosis for patients that were admitted to a hospital within 1 day of ambulance dispatch (as in column 1). Intake diagnoses have been coded based on the first letter in the ICD10 coding system for the primary diagnosis in the data. See the Online Appendix, Figure A10, for the results for all available categories. Standard errors are clustered at the grid cell level.

site (see Table 4).[22]

We have shown that patients serviced by private ambulances are more often left at home and thereby less often admitted to a hospital. This missed healthcare contact either with an ED physician or through an admission could be immensely important for health in the short run. For example, early treatment of a minor stroke or TIA can reduce the risk of reoccurring stroke by 80% (Rothwell et al., 2007). Moreover, not being properly examined and diagnosed can reduce access to secondary prevention options, for example, medications, potentially explaining why we find larger effects on mortality later on (Korhonen et al., 2017; Orchard et al., 2018; Lindahl et al., 2017).

5.3 Heterogeneity in the effect of private ambulance service on mortality

A heterogeneity analysis can shed light on which patient groups are affected the most from having a private ambulance dispatched. In Figure 3, we show the results from regression models where we interact the private ambulance dummy with a set of pre-determined char-

[22]In-hospital care can be important for patient health through access to advanced diagnostic tools available only while in care. An admitted patient's diagnoses, treatments, and medications will be provided by physicians specialized in the relevant area of medicine. Moreover, once dismissed from hospital care, and if needed, outpatient follow-up at a hospital based clinic are often provided. Moreover, after a hospital admission, home assistance with daily tasks will be organized before discharge if necessary.

acteristics. We run a interaction regression model separately for each characteristic. In the Online Appendix, Table A10, the complete regression results are shown. Figure 3 shows the estimated coefficient for the interaction between the private ambulance indicator and the characteristic. The figure depicts an interesting pattern; the mortality for priority 1, male, and old patients is more affected by having a private ambulance dispatched than other patients. Interestingly, patients that had *no* pre-treatment hospital visits in 2007-2008 are affected more. When interacting the private indicator with different diagnoses from 2007-2008, there is no incremental effect.

Thus, inferior diagnosis and treatment by the ambulance crew seems therefore to be linked to historically healthier patients that do not frequently visit the hospital. This makes sense, as a part of the diagnosis and treatment at site is a function of interviewing the patient about their historical health status; i.e., historically sick people are easier to properly diagnose. A recent observational study using Swedish data corroborates the validity of this interpretation (Figtree et al., 2021). The authors find that among severe AMI patients, those without risk factors had within 30 days a 50% higher mortality, which did not converge until after 8 years. One of the authors argue that a reason for this paradoxical finding could be that patients without risk factors are assumed to have a better prognosis by healthcare professionals and are thereby provided with less adequate treatments after the event (Hake, 2021). Otherwise healthy individuals experiencing an acute medical event could also be undertreated in the ambulance setting.

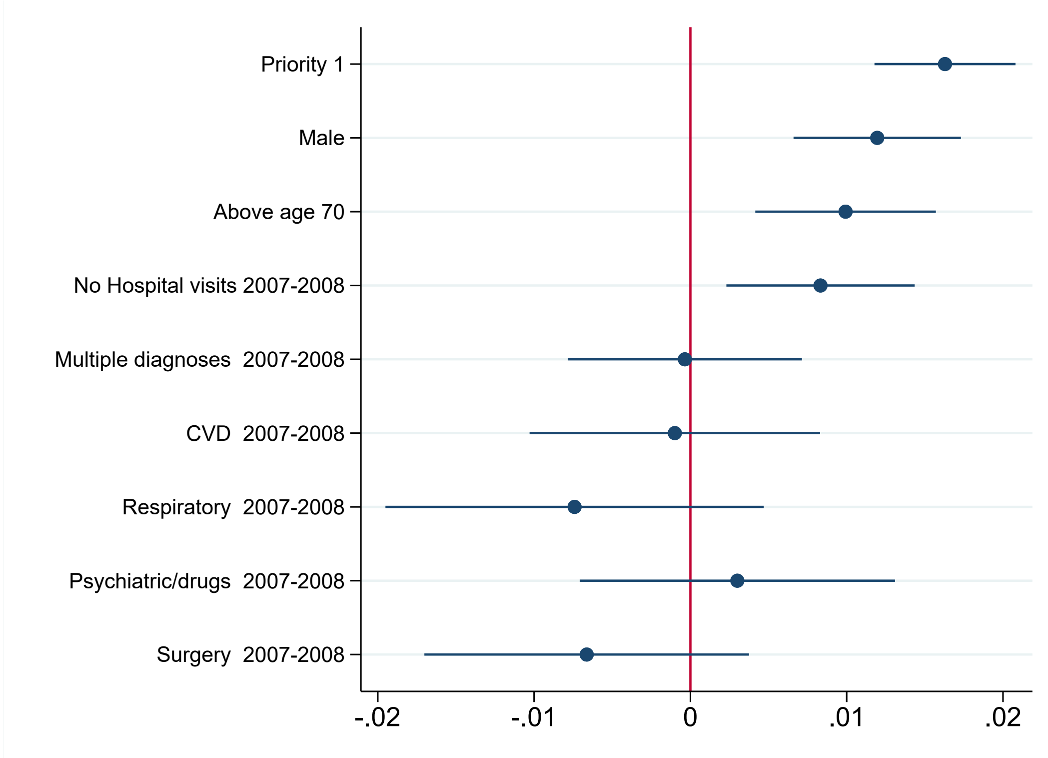
5.4 Personnel policy differences

When examining credible sources, such as newspapers, public service mass media, and information from the unions, there are many indications of differences in personnel policies. First, there are some differences with respect to the collective agreement and circumventions of labor laws between the private and public employers.[23] A standard work week in the public firm is 37 hours instead of 38.25 in the private firms, a 3 percent difference. Working more will enable the employee to incur extra compensation. Thus, this potentially results in lower labor cost and less well-rested personnel for the private provider.

In addition, on-call duty time can be scheduled. According to one of the leading Swedish dailies, the independent conservative Svenska Dagbladet, the ambulance staff in one of the private firms was normally scheduled for 42 hours per week (including on-call duty) (Cunvik, 2011). The on-call duty leads to around 200 hours additional working hours a year compared

[23]The collective agreements form the basis of "the Swedish Model", in which the conditions in each sector are determined jointly by the trade unions and the employer organizations. The government does not interfere, though it may set the boundaries through labor laws.

Figure 3: Heterogeneity in the treatment effect: 3-year mortality



Note: Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters in size, as well as year, month, and priority to patient indicator variables. Each row represents a separate regression where the dummy variable for private ambulance is interacted with a predetermined characteristic. The main effects are included, but only the interaction term is presented in the figure. See the Online Appendix, Table A10, for a more thorough description of the models. In row 1, the interaction variable is created by interacting the private dummy with a binary variable indicating if the priority to the patient was a priority 1 dispatch. In row 2, the interaction variable is created by using a "male" dummy. The interaction variable in row 3 is created by using a binary variable to denote if the patient is above 70 years of age. In rows 4-9, we use the predetermined health variables measured in 2007-2008 and add them together into 6 distinct categories: any hospitalization, more than 1 diagnosis, any CVDs, any respiratory diseases, any psychiatric conditions and drug abuse, and any surgical diseases, including cancer and pain conditions. Standard errors are clustered at the grid cell level.

to what the collective agreement stipulates. Together with additional scheduled working time in private firms, private employees could have been working up to 18% more regular hours than public employees on a full time contract.

Moreover, the Swedish labor laws stipulate a maximum of 200 hours of overtime per annum. To circumvent the law restricting the use of overtime, one private firm also employed their staff in a subsidiary staffing company, which essentially allowed for an unrestricted overtime (TT, 2012; Olsson, 2017). Moreover, in 2017, according to a firm evaluation produced by the county, there were indications that private firms had, on average, both a higher turnover rate and a higher fraction of their staff on temporary employment contracts (SLL, 2017).

Dagens Arena, an independent left wing newspaper, reported in 2014 that the private providers were less prone to send their employees to on-the-job training and reported that the reason was that the private providers had fewer regular employees and relied to a higher degree on overtime and temporary staff (Leander, 2014). An indication that this reliance on temporary staff continues in 2017 can be found in firm specific evaluation reports. According to the evaluations, private firms produced 713 ambulance hours per regular employee, while the public firm produced 582 hours, a 23 % difference. This difference suggests that the private firms relied more on temporary staff, overtime, and longer regular work hours to staff their ambulances (SLL, 2017).

A high staff turnover and a high reliance on temporary employed staff can make a firm vulnerable to sudden staff shortages. The Public Service radio reported in 2013 that because of a staff shortage, one of the private providers was levied with fines for failing to include a specialist nurse in each ambulance, as contractually required (SR, 2013). There is ample evidence suggesting that private firms cost innovate at the expense of their staff by providing more regular hours, more overtime, and fewer full-time contracts. This could lead to higher staff turnover, the recruitment of low-quality staff, and a larger share of ambulance workers that are more affected by fatigue. In the following section, we address these questions by using Swedish administrative matched employer-employee data.

5.5 Staff quality differences

We argue that one likely explanation for the private firm penalty in mortality comes from staff quality differences.[24] The literature has shown that staff quality in healthcare is important

[24] To cost innovate, for these firms, there are arguably few other ways that could sufficiently affect quality. Firms can purchase less expensive vehicles and other supplies, although the ambulances must be equipped according to regulations and contracts. One way that better vehicles could affect health is through faster transport times. What we find is that private firms reach patients faster, and public firms reach hospitals faster. Jointly, there is not much difference in total transport times, measured from dispatch to drop off,

for health outcomes. Gruber and Kleiner (2012) found that the nurses' strikes in New York hospitals increased mortality among patients admitted during the strikes by 18%. This effect was similar for hospitals that engaged temporary staff, stressing the importance of location-specific experience in healthcare. Bartel et al. (2014) provide even more direct evidence on the importance of the nurses' job experience and education. The departures of experienced nurses, the use of temporarily contracted nurses, and the employment of new hires were all associated with longer hospital stays in a US hospital environment. Gruber et al. (2018) showed that setting a maximum waiting time to see a physician for patients in EDs in the UK reduced mortality by 14%.

The complexity of ambulance care comes not only from the variety of patients who are serviced. In contrast to hospital care, the ambulance clinicians are medically responsible for each patient and make independent decisions regarding treatments and diagnosis. Access to greater medical knowledge and second opinions is limited in the field.[25] The patients who are serviced have many kinds of ailments, diseases, and comorbidities, requiring the ambulance clinician to be a generalist with a wide range of competencies. Furthermore, compared to other medical staff, e.g., ED nurses, who have access to diagnoses set by a physician that are often based on more advanced examinations, the ambulance staff in Stockholm do not have structured access to feedback on their medical decisions. Ambulance staff can therefore rarely evaluate whether their decisions were accurate. Staff quality could thereby be very important for quality outcomes in this setting.

Unfortunately, we have no data on the ambulance staff for every dispatch. However, we can link matched employer-employee data and calculate average quality measures of ambulance workers across private and public ambulance employers in Stockholm County for the relevant years. We have matched staff working for different ambulance providers and also separated out the specialist nurses. Table 6 shows the differences in means for the full crew in panel 1 and those for the ambulance clinicians in panel 2. Starting with differences in turnover in column 1, we find a distinct higher (66%) turnover for private employees. The annual wage income is about 10% higher in the private sector. However, we do not interpret this as higher hourly wages, as the number of hours worked per regular employee is higher in private firms, as discussed in section 5.4. Experience is somewhat lower and age higher in the private sector, but neither are statistically different. The measure for cognitive skills

between the different firms (see the Online Appendix, Table A4). Other equipment is also tightly regulated, and ambulance station rents are normally fixed, as the ambulances often share facilities with fire brigades or occupy other fixed locations.

[25] An emergency trained physician is available on call but can be occupied with other calls, and the situation does not always allow for lengthy phone conversations. One unit staffed with a physician was available during the day.

is missing for females when using conscription data and is not registered for those in the cohort with a 9th grade education level who graduated before 1988, yielding a lower number of observations. Although not always precisely measured, the overall picture is that the employees in private firms score lower on these cognitive skills measures (column 5-7).

We interpret the evidence jointly to indicate that private firms, on average, are hiring different competences and have more difficulties retaining staff. Cost innovation at the expense of staff could cause these problems, leading these firms to be less attractive employers. We have described how private firms to a higher degree use temporary contracts and demand more hours for a full time contract, which is consistent with this interpretation. Nevertheless, there are other possible explanations for why private firms would hire staff of lower quality. In the following section, we discuss and address several of these competing mechanisms.

5.6 Competing mechanisms

One could consider other sorting mechanisms not related to the private firms' production choices, as in Hart et al. (1997). It may be a job disamenity to have the ambulance station located far from amenities in the central part of Stockholm (public transportations, restaurants, etc.). As a result, better nurses may want to work for the public ambulances, as the stations are often located centrally. An empirical finding that speaks against this can be found in the results in Online Appendix, Table A3, where we restrict the sample to only include stations in the inner city and other proximate parts of Stockholm. In the restricted sample, the amenities are more similar. For example, there is a highly frequent subway-, commuter train-, and bus-system. Within this restricted sample, the transport time to the central station is never more than 30 minutes from the nearest train station. It is therefore reassuring that the results are robust when using this subsample.

A related sorting theory could be that all high skilled personnel want to work for public ambulances and that the skill distribution of ambulance nurses is fixed. Then, average performance is not affected by changes in institutional arrangements. Figure A5 in the Online Appendix plots the number of specialist nurses residing in Stockholm and relates that to the number of specialist nurses in the ambulance sector. After a standard in-house training of 10 weeks, any specialist nurse can start to work in an ambulance. This shows that there are a large number of specialist nurses in the potential pool in the Stockholm Region (a factor 10 larger than the pool of ambulance nurses). However, this may not be a potential pool to recruit from if the highly skilled nurses' preferences are extremely in favor of a public job. In Table A7 in the Online Appendix, we test for this by studying transfers of nurses between providers. We find that there is a substantial movement between sectors. For example, in relation to average annual turnover in the public ambulance firm, 20% of

Table 6: Firms and employees

	Turnover	Annual Wage income (EUR)	Experience (Ambulance firm years)	Age	Cognitive skills (conscription)	9th grade GPA	9th grade math grade
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All employees							
Private ambulance	0.085	3307.519	-0.197	1.035	-0.245	-0.137	-0.098
	(0.011)	(669.727)	(0.417)	(0.622)	(0.139)	(0.075)	(0.077)
Public outcome mean	0.1277	36271.5797	8.4445	41.2344	5.0774	-0.061	-0.044
Observations	5436	5436	5436	5436	1776	2321	2321
Specialist nurses							
Private ambulance	0.069	4429.354	-0.558	0.678	-0.227	-0.228	-0.163
	(0.019)	(1186.151)	(0.654)	(0.966)	(0.232)	(0.096)	(0.111)
Public outcome mean	0.1625	39477.7871	7.6989	39.9034	5.4283	0.248	0.183
Observations	1869	1869	1869	1869	560	1046	1046

Note: Each model includes year dummies and includes the years 2009-2016, unless otherwise noted below. In column 1, the outcome variable *turnover* is an indicator variable that takes the value 1 if an employee is observed in a specific ambulance firm in year t but not in t+1. It is a combination of employees exiting the sector and employees that change employer (if we use only the new contract period, 2012-2016, we find a smaller but significant estimate of 0.04 for all employees). In column 2, the outcome is the yearly wage income (using the exchange rate, 1 EUR=10 SEK) according to each individuals filed tax report. In column 3, experience is a variable that describes how many years an employee has potentially been working in the ambulance sector in Stockholm (in one of the firms operating in 2009-2016), with 1990 as the earliest year of observation. In column 4, age in years is the outcome. In column 5, the outcome is cognitive skills measured in Swedish conscript data and the cognitive tests available therein. We use the average score on 4 tests measuring logical, verbal, technical, and spatial skills and that are standardized to a standard nine scale (1-9) as the outcome. For men, this data is only available with full coverage before the early 1980s; the data provides poorer coverage of younger men born in the mid-1980s or later when mandatory conscription was phased out. Columns 6 and 7 use register data on grades (from the 9th grade) available from 1988 until today. We use the GPA in column 6 and the math grade in column 7 as the outcomes. Both variables are normalized to a mean of zero and a standard deviation of one by each cohort. This data is available for both men and women but only for individuals born after 1971. Panel 1 shows the results for all employees, and Panel 2 for nurses with a specialist degree only. See further the Online Appendix, Section 6, which describes the firm data. All data on firms and employees come from the FAME database. Standard errors are clustered at each individual employee in the data.

the nurses move to a private ambulance firm on average each year. Moreover, we find no statistical difference between the relative flow of nurses from private to public ambulance firms compared to flows from public to private. We argue that this speaks against very strong preferences in favor of working for a public ambulance firm.[26] Moreover, in column 3 of Table A7 in the Online Appendix, we show that exits from the ambulance sector are more common for employees in private firms. Although this result could be generated by the selection of workers to firms, it indicates that the market is not fixed and that poor working conditions in private firms could reduce the supply of scarce ambulance staff.

We acknowledge that it is ultimately difficult to rule out or quantify competing mechanisms and that the tests provided are only indicative. However, we argue that the sorting mechanisms discussed above do not easily provide the predictions that privately employed nurses would dominate public providers on the contracted measures, such as time to patient, but not on the non-contracted quality measures.

5.7 Threats to identification: Robustness

There are multiple factors that could threaten the internal validity of our design. In Figure 4, we address several issues related to ambulance services and market structure by displaying the point estimate with corresponding 95% confidence intervals for a range of specifications. Using three-year mortality as the outcome, in the first row, we display the bivariate, non-causal correlation. This correlation is negative and substantial in size, suggesting that the compositional bias between patients serviced by private firms and those serviced by the public firm is negative. This finding is expected, as private firms service a younger population, on average, as shown in the left panels of Figures 1.

In the second row, we show that when adding the grid fixed effects and the year and month dummies, the estimated effect shifts sign and is now positive and of a similar magnitude as in our preferred model. This specification could, however, be problematic and sensitive, as the treatment allocation mechanism could be related to the priority assigned to patients by the dispatch service. For ambulance dispatchers, in addition to the locations of the patients and ambulances, this is the most vital piece of information they have. Nevertheless, flexibly controlling patient priority has no effect on the point estimate (row 3). This model is also the preferred model that we utilize more broadly in the paper.

Having shown how mortality responds to private providers as we implement our preferred model, we now present reasonable deviations from this model to demonstrate the robustness

[26] A related theory is that if working in a public ambulance provides better opportunities for moving to a career in public hospitals, it will attract the best nurses. Evidence against this type of sorting is provided in the Online Appendix, Table A7, column 5, where we find no predictive power of the ambulance nurse provider type on the probability of changing a job to work in a public hospital.

of our results. We start by excluding the two private ambulance stations where two special ambulance units were stationed (row 4). These units were installed in 2012, had staff with higher competence (nurses with anesthesiology training), and targeted the most severely injured patients as a supporting unit to a regular ambulance. While these units could spuriously relate private providers to mortality, this exclusion has only a minor impact on the estimated effect.

Since public and private ambulances operate in different areas with different characteristics, one could suspect that there are differences in the types of patient injuries between the areas and that this is spuriously reflected in ambulance provider assignment. For example, as private units are located in areas with larger roads and faster traffic, they could respond to more traffic accidents and thereby face a higher probability of patient deaths. Moreover, they could be called to the scene more often as a "leading unit", which would be responsible for organizing the care and transport of multiple injured patients. To this end, we exclude all ambulance dispatches where more than one ambulance is engaged. Removing multiple-unit dispatches excludes not only severe traffic accidents but any other multiple injury events, such as cardiac arrests and other severe injuries where more personnel or a specialist anesthesiologist nurse is required.[27] This exclusion does not change the estimated effect (row 5).

We also exclude all priority 4 dispatches, which involve patients without a need for medical care but unable to take a taxi for different reasons. Ambulances were not supposed to transport these patients but occasionally did so. Private ambulances did so even more often in the data. These patients, although not critically ill, had a high three-year mortality risk, which could feed into the overall effect of private care (row 6). However, we find that neither this exclusion nor the exclusion also of priority 3 transports in row 7 changes the estimated relationship.

More generally, we use several approaches to challenge our estimated effect with respect to geographical or temporal confounding. First, although we have already shown that the control variables are balanced in our basic specification, we include all the control variables we have in row 8. As expected, this inclusion does not have a major impact on the estimated effect.

Another threat to identification comes from changes over time within the geographical units that we study. Several new ambulance units were introduced during the period studied, and the underlying composition of residents in different neighborhoods could have changed

[27] This variable is somewhat more inclusive than the pre-determined variable we use in the balance test. In the balance test, we create a variable indicating that more than one unit was dispatched to the same patient. Here, we also include multiple dispatches to different patients at the same scene, thereby including traffic accidents and other multiple victim events.

due to certain factors, e.g., housing constructions or some other demographic shifts. To account for secular changes over time in a flexible way, we interact the grid cell fixed effects with year dummies. In this specification, we only compare private and public ambulance dispatches within a grid cell and unique year. This specification does not change the results (row 9). In row 10, we use an even more restrictive specification by interacting the fixed effects with both year and month and show that the main result is still robust.

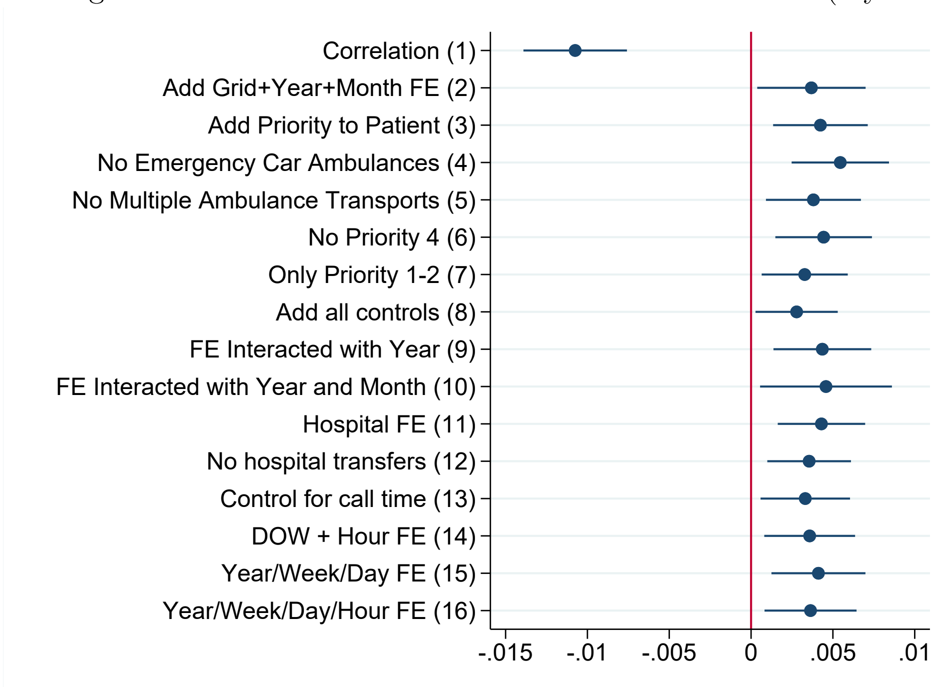
Doyle et al. (2015) leverage that ambulances in the US are essentially randomly assigned to patients but use the fact that different ambulance firms have different hospital preferences. We have argued that this is not a feature of the ambulance services in Stockholm, as patients are normally transported to a predetermined hospital based on their residence. To empirically validate this claim, we augment our preferred model with hospital fixed effects. Any hospital-specific mechanism is thereby subsumed by these indicators. Row 11 of Figure 4 shows that our claim is valid, as there is no change in the effect when introducing hospital dummies.

Moreover, as the geography of private and public ambulance stations is fixed, there could be persistent differences that affect ambulance allocation. We consider hospital transfers as one such persistence. Hospitals are placed in different ambulance areas, and transferring patients between them could induce dispatch operators to selectively choose ambulances based on characteristics related to the provider. For example, transferring a patient from Huddinge Hospital (located in a private district in the southern part of the city) to an inner-city hospital (there are 3 such hospitals in the public districts) could imply that a public ambulance is more likely to be engaged because it will be in its home district when ready for a new assignment. By excluding transports that start at a hospital (i.e., a hospital transfer) in row 12, we find that this potential bias does not affect our estimates.

Furthermore, the fixed geography of ambulance stations could affect the time before being assigned an ambulance. As more urgent patients, even within a priority level, could be serviced faster than others, the time of the interview and dispatch could be informative of each patient's urgency. Although private ambulances leave more patients at home and estimate them to be less severely injured, it cannot be completely excluded that they are assigned to more severely injured patients with a higher mortality expectancy. We control for the time between the emergency call and the ambulance dispatch as a proxy for unmeasured urgency in our preferred specification (row 13) and find that this control variable has a negligible impact on the estimated effect.

Finally, we consider time as a possible confounder for our results. If the patients' health conditions are correlated with private assignment through the time of the day (or day of the week), we could be estimating a spurious relationship. Private ambulances could work more often in public districts at certain times or dates when patient characteristics are different or the health environment is changed (e.g., during nights). If this is true, our estimates should be sensitive to more flexible time controls. However, including day-of-the-week fixed effects, hour fixed effects, or even fixed effects by each unique hour does not change the results (rows 14-16).[28]

Figure 4: Robustness: Deviations from the basic model (3-year mortality)



Note: This figure presents point estimates and 95% confidence intervals for different specifications of Equation 1. The first two rows impose fewer restrictions than Equation 1. The first row is the raw correlation between 3-year mortality and private ambulance providers. Row 2 adds fixed effects, leading up to our preferred specification in row 3, which also includes priority to patient indicators. Rows 4 and below challenge this specification by adding structure or restricting the sample in different ways, as described in the table and text. Standard errors are clustered at the grid cell level.

Although we have considered several potential differences between the districts, one concern could be that between private and public districts, there are more subtle differences that we do not capture. The composition of patients in public areas might differ in a way that makes patient knowledge very important and hence grants public ambulances a quality advantage in their own districts. If this were true, private ambulances would perform poorly in public districts but well in their own districts. To test this hypothesis, we divided Stockholm into private and public districts based on the patient's location at the time of the call. We estimate Equation 1 on each of these samples using 3-year mortality as the

[28]In the Online Appendix, Section 5, we replicate the same results for 2-year mortality in Figure A9.

outcome and report the results in Online Appendix, Table A5. We find that estimating the effect of private care on mortality is similar in size in both samples and that the estimates are statistically indistinguishable from one another.

Another way to make our identification strategy more credible is to focus on grid cells where the probability of being serviced by a private ambulance is close to the probability of being serviced by a public ambulance. If much of our identifying variation came from grid cells with poor overlap between private and public ambulances, i.e., public (private) areas where private (public) ambulances rarely go, our approach would be much more problematic. In grid cells with poor overlap, there is a higher risk of nonrandom assignment of ambulances to patients. Grid cells with an even distribution of private and public ambulances are neighborhoods with a similar distance to a public and a private ambulance station or around the borders of their respective districts. In these grid cells, the probability of being serviced by a private ambulance is close to 50%. To approach this question, we take the average of the dummy variable *Private* within each grid cell and year and condition our regression models on there being at least a certain fraction of private and a certain fraction of public ambulance dispatches in each cell. In the Online Appendix, Figure A4, these results are shown for the fraction of private dispatches of 0.1-0.9, 0.2-0.8, 0.3-0.7, and 0.4-0.6. We find that the results line up well across these different samples, even though in our most restricted sample, we only utilize approximately 10% of the full sample. These findings suggest that "border" areas provide much of the identifying variation that we use for estimation.

6 Conclusions

Outsourcing public services to private firms is a tool to help the public sector contain costs without reducing quality. Theoretically, the economics literature describes a tradeoff related to the contractibility of quality in the public domains considered for outsourcing. This tradeoff comes from private firms' strong incentives to cost innovate but weak incentives to uphold quality if it is not explicitly profitable. When private firms are not sufficiently incentivized in contracts to produce a minimum level of quality, it is possible or even likely that they will produce lower quality. When quality can be contracted on, private firms can theoretically produce adequate quality at a lower cost than public sector in-house production.

We test these predictions in the healthcare sector, which has received little attention in the economics literature on outsourcing. Health outcomes are often difficult to contract on, as they may reflect the selected characteristics of patients rather than provider quality. We solve this challenge in an ambulance services environment where both private and public firms operate side-by-side and where ambulance assignment is as good as random. We confirm the theoretical prediction that private firms can uphold quality if contracted on but produce substantially lower quality on noncontracted health outcomes. Patients serviced by private ambulances have higher mortality within one day and up to three years after being serviced.

We provide evidence in support of the explanation that differences in staff quality are a likely reason for the differences in mortality. We provide additional evidence showing behavioral differences between ambulance firms during each assignment. In particular, we find that private firm ambulances leave substantially more patients at home than public ambulances. Causal estimates are complemented by descriptive evidence suggesting that private firms have a higher turnover, require more hours from their staff, rely more on overtime, and provide less on-the-job training. We also find indications of cognitive skill differences.

When quality is difficult to contract on, theory advises against outsourcing. Policy makers inclined to outsource public services attempt to solve this dilemma by contracting on proxies or on inputs that are assumed to be related to quality. However, this procedure is subject to moral hazard and provides strong incentives for firms to comply with contracted measures while shirking on unmeasured aspects of quality. Moreover, writing overly specific contracts on quality proxies may not be the solution, as this will severely limit the private firms' ability to cost innovate through competition. Contracting on the outcome is thereby the best option to uphold quality and to take advantage of the private firms' inherent superiority in cost innovation.

The literature has described an intermediary solution to this dilemma, namely, engaging nonprofit private firms (Glaeser and Shleifer, 2001). These nonprofit firms can leverage innovative pressure from the private sector while committing to softer incentives. If that is not possible, theory and our evidence show that outsourcing may have very adverse effects. Our estimated effect suggests that private ambulance services led to an additional 420 deaths (within three years) each year. Our back-of-the-envelope cost calculations suggest that the loss due to private ambulance firms is substantial. Even very conservative estimates find that the cost could be approximately one-quarter of the total cost of the ambulance services and almost twice as high as the maximum of cost savings generated by outsourcing. Thus, we are unconvinced that outsourcing to for-profit firms is a superior option in this context.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Working Paper 24003, National Bureau of Economic Research.

AISAB (2020). Ambulanssjukvården i Stockholm AB. `https://www.aisab.nu/om-oss/nyheter/aisab-20-ar/`. Accessed on February 17th 2020.

Andersson, F., Jordahl, H., and Josephson, J. (2019). Outsourcing Public Services: Contractibility, Cost, and Quality. *CESifo Economic Studies*. ifz009.

Armstrong, G. S. and MacKenzie, D. L. (2003). Private versus public juvenile correctional facilities: Do differences in environmental quality exist? *Crime & Delinquency*, 49(4):542–563.

Bales, W. D., Bedard, L. E., Quinn, S. T., Ensley, D. T., and Holley, G. P. (2005). Recidivism of public and private state prison inmates in Florida. *Criminology & Public Policy*, 4(1):57–82.

Barron, D. N. and West, E. (2017). The quasi-market for adult residential care in the UK: Do for-profit, not-for-profit or public sector residential care and nursing homes provide better quality care? *Social Science & Medicine*, 179:137 – 146.

Bartel, A. P., Beaulieu, N. D., Phibbs, C. S., and Stone, P. W. (2014). Human Capital and Productivity in a Team Environment: Evidence from the Healthcare Sector. *American Economic Journal: Applied Economics*, 6(2):231–59.

Bayer, P. and Pozen, D. E. (2005). The effectiveness of juvenile correctional facilities: Public versus private management. *The Journal of Law and Economics*, 48(2):549–589.

Behaghel, L., Crepon, B., and Gurgand, M. (2014). Private and public provision of counseling to job seekers: Evidence from a large controlled experiment. *American Economic Journal: Applied Economics*, 6(4):142–174.

Bennett, J. and Iossa, E. (2009). Contracting out public service provision to not-for-profit firms. *Oxford Economic Papers*, 62(4):784–802.

Bennmarker, H., Grönqvist, E., and Öckert, B. (2013). Effects of contracting out employment services: Evidence from a randomized experiment. *Journal of Public Economics*, 98:68 – 84.

Bergman, M. A., Johansson, P., Lundberg, S., and Spagnolo, G. (2016). Privatization and quality: Evidence from elderly care in Sweden. *Journal of Health Economics*, 49:109 – 119.

Correia, S. (2014). Reghdfe: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. Statistical Software Components S457874.

Cunvik, Z. (2011). Mer arbete oroar ambulanspersonal. *Svenska Dagbladet*. 30th of December.

Doyle, J. J. (2011). Returns to Local-Area Health Care Spending: Evidence from Health Shocks to Patients Far from Home. *American Economic Journal: Applied Economics*, 3(3):221–43.

Doyle, J. J., Graves, J. A., and Gruber, J. (2017). Evaluating Measures of Hospital Quality. Working Paper 23166, National Bureau of Economic Research.

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.

Duggan, M., Gruber, J., and Vabson, B. (2018). The Consequences of Health Care Privatization: Evidence from Medicare Advantage Exits. *American Economic Journal: Economic Policy*, 10(1):153–86.

Duwe, G. and Clark, V. (2013). The effects of private prison confinement on offender recidivism: Evidence from Minnesota. *Criminal Justice Review*, 38(3):375–394.

Ebben, R. H., Vloet, L. C., Speijers, R. F., Tönjes, N. W., Loef, J., Pelgrim, T., Hoogeveen, M., and Berben, S. A. (2017). A patient-safety and professional perspective on non-conveyance in ambulance care: a systematic review. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 25(71).

Figtree, G. A., Vernon, S. T., Hadziosmanovic, N., Sundström, J., Alfredsson, J., Arnott, C., Delatour, V., Leósdóttir, M., and Hagström, E. (2021). Mortality in stemi patients without standard modifiable risk factors: a sex-disaggregated analysis of swedeheart registry data. *The Lancet*, 397:1085 – 1094.

Francois, P. (2003). Not-for-profit provision of public services. *The Economic Journal*, 113(486):C53–C61.

Glaeser, E. L. and Shleifer, A. (2001). Not-for-profit entrepreneurs. *Journal of Public Economics*, 81(1):99–115.

Grabowski, D. C., Feng, Z., Hirth, R., Rahman, M., and Mor, V. (2013). Effect of nursing home ownership on the quality of post-acute care: An instrumental variables approach. *Journal of Health Economics*, 32(1):12 – 21.

Gruber, J., Hoe, T. P., and Stoye, G. (2018). Saving lives by tying hands: The unexpected effects of constraining health care providers. Working Paper 24445, National Bureau of Economic Research.

Gruber, J. and Kleiner, S. A. (2012). Do strikes kill? Evidence from New York State'. *American Economic Journal: Economic Policy*, 4(1):127–57.

Gupta, A. (2017). Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program. Working Paper 07, Becker Friedman Institute for Research in Economics.

Hake, C.-M. (2021). Överlevnaden var sämre i hjärtinfarkt utan riskfaktorer. *Dagens Medicin*. 23rd of March.

Hart, O. and Moore, J. (1990). Property rights and the nature of the firm. *Journal of Political Economy*, 98(6):1119–1158.

Hart, O., Shleifer, A., and Vishny, R. W. (1997). The proper scope of government: Theory and an application to prisons. *Quarterly Journal of Economics*, 112(4):1127–1161.

Hälso och sjukvårdsförvaltningen, H. (2011). Upphandling av ambulanssjukvården. Faktasammanställning från Hälso- och sjukvårdsnämndens förvaltning. Technical report, Stockholms läns landsting.

HSF (2018). Årsrapport 2017 prehospitala verksamheter i SLL. Technical Report 2018:0024, Hälso och sjukvårdsförvaltningen, Stockholms läns landsting.

Jansson, A. and Castwall, H. (2017). Är privat ambulansverksamhet effektivare än offentlig: En analys av de prehospitala ambulansvårdgivarna i Stockholms län under år 2015. Bacheolor thesis, Stockholm University.

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

Korhonen, M. J., Robinson, J. G., Annis, I. E., Hickson, R. P., Bell, J. S., Hartikainen, J., and Fang, G. (2017). Adherence tradeoff to multiple preventive therapies and all-cause mortality after acute myocardial infarction. *Journal of the American College of Cardiology*, 70(13):1543–1554.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions*. *The Quarterly Journal of Economics*, 114(2):497–532.

Landstingsrevisorerna (2013). Prehospital vård – vårdkedjans första insats. Technical Report 08/2013, Stockholms läns landsting.

Laun, L. and Thoursie, P. S. (2014). Does privatisation of vocational rehabilitation improve labour market opportunities? Evidence from a field experiment in Sweden. *Journal of Health Economics*, 34:59 – 72.

Leander, P. (2014). Kris inom ambulansen, "vi klarar inte en stor olycka". *Dagens Arena*. 30th of September.

Lederman, J., Lindström, V., Elmqvist, C., Löfvenmark, C., Ljunggren, G., and Djärv, T. (2020). Non-conveyance of older adult patients and association with subsequent clinical and adverse events after initial assessment by ambulance clinicians: A cohort analysis. Mimeo.

Lee, M. H., Schuur, J. D., and Zink, B. J. (2013). Owning the Cost of Emergency Medicine: Beyond 2%. *Annals of Emergency Medicine*, 62(5):498 – 505.e3.

Lindahl, B., Baron, T., Erlinge, D., Hadziosmanovic, N., Nordenskjöld, A., Gard, A., and Jernberg, T. (2017). Medical therapy for secondary prevention and long-term outcome in patients with myocardial infarction with nonobstructive coronary artery disease. *Circulation*, 135(16):1481–1489.

Lucchese, E. (2020). It could be worse...it could be raining: Ambulance response time and health outcomes. Working Paper 429, University of Milan Bicocca, Department of Economics, Management and Statistics.

Miller, D. L., Shenhav, N., and Grosz, M. Z. (2019). Selection into identification in fixed effects models, with application to head start. Working Paper 26174, National Bureau of Economic Research.

OECD (2019). *Government at a Glance 2019*. OECD publishing, Paris.

Olsson, A. (2017). Falck ambulans kritiseras för att kringgå regler om övertid. *Vårdfokus.* 20th of June.

Orchard, J., Lowres, N., Neubeck, L., and Freedman, B. (2018). Atrial fibrillation: is there enough evidence to recommend opportunistic or systematic screening? *International Journal of Epidemiology*, 47(5):1372–1378.

Patterson, P. D., Weaver, M. D., Frank, R. C., Warner, C. W., Martin-Gill, C., Guyette, F. X., Fairbanks, R. J., Hubble, M. W., Songer, T. J., Callaway, C. W., Kelsey, S. F., and Hostler, D. (2012). Association between poor sleep, fatigue, and safety outcomes in emergency medical services providers. *Prehospital Emergency Care*, 16(1):86–97.

Picard, R. and Stepner, M. (2015). Geo2xy: Stata module to convert latitude and longitude to xy using map projections. Statistical Software Components S457990.

Powers, R. A., Kaukinen, C., and Jeanis, M. (2017). An Examination of Recidivism Among Inmates Released From a Private Reentry Center and Public Institutions in Colorado. *The Prison Journal*, 97(5):609–627.

RiR (2012). Statens insatser inom ambulansverksamheten. Technical Report 20, Riksrevisionen.

Rivard, M. K., Cash, R. E., Chrzan, K., and Panchal, A. R. (2020). The impact of working overtime or multiple jobs in emergency medical services. *Prehospital Emergency Care*, 24(5):657–664.

Rogers, A. E., Hwang, W.-T., Scott, L. D., Aiken, L. H., and Dinges, D. F. (2004). The working hours of hospital staff nurses and patient safety. *Health Affairs*, 23(4):202–212.

Rothwell, P. M., Giles, M. F., Chandratheva, A., Marquardt, L., Geraghty, O., Redgrave, J. N., Lovelock, C. E., Binney, L. E., Bull, L. M., Cuthbertson, F. C., Welch, S. J., Bosch, S., Carasco-Alexander, F., Silver, L. E., Gutnikov, S. A., and Mehta, Z. (2007). Effect of urgent treatment of transient ischaemic attack and minor stroke on early recurrent stroke (EXPRESS study): a prospective population-based sequential comparison. *The Lancet*, 370(9596):1432–1442.

Shleifer, A. (1998). State versus private ownership. *Journal of Economic Perspectives*, 12(4):133–150.

SLL (2005). Förfrågningsunderlag upphandling av ambulanssjukvårdstjänst. Stockholms läns landsting.

SLL (2017). Årsprotokoll från uppföljning 2017 – vägburen ambulans. Hälso och sjukvårdsförvaltningen, Stockholms läns landsting.

SOSAAB (2020). Ambulansdirigering, SOS Alarm AB. Accessed November 5th, 2020. `https://www.sosalarm.se/Kund/tjanstekatalog2/ambulansdirigering/`.

Spivak, A. L. and Sharp, S. F. (2008). Inmate Recidivism as a Measure of Private Prison Performance. *Crime & Delinquency*, 54(3):482–508.

SR (2013). Falck ambulans kör med personal utan rätt kompetens - krävs på skadestånd. *Sveriges Radio*. 7th of September.

Szymanski, S. (1996). The impact of compulsory competitive tendering on refuse collection services. *Fiscal Studies*, 17(3):1–19.

Szymanski, S. and Wilkins, S. (1993). Cheap Rubbish? Competitive Tendering and Contracting Out in Refuse Collection – 1981–88. *Fiscal Studies*, 14(3):109–130.

Tanuseputro, P., Chalifoux, M., Bennett, C., Gruneir, A., Bronskill, S. E., Walker, P., and Manuel, D. (2015). Hospitalization and mortality rates in long-term care facilities: Does for-profit status matter? *Journal of the American Medical Directors Association*, 16(10):874 – 883.

TT (2012). Bemannningsföretag kringgår övertid. *Svenska Dagbladet*. Tidningarnas telegrambyrå.

Wübker, A. and Wuckel, C. (2019). The Impact of Private for-Profit Hospital Ownership on Costs and Quality of Care – Evidence from Germany. *CESifo Economic Studies*, 65(4):373–401.

ONLINE APPENDIX TO ACCOMPANY:

THE QUALITY AND EFFICIENCY OF PUBLIC AND PRIVATE FIRMS: EVIDENCE FROM AMBULANCE SERVICES

Daniel Knutsson & Björn Tyrefors*

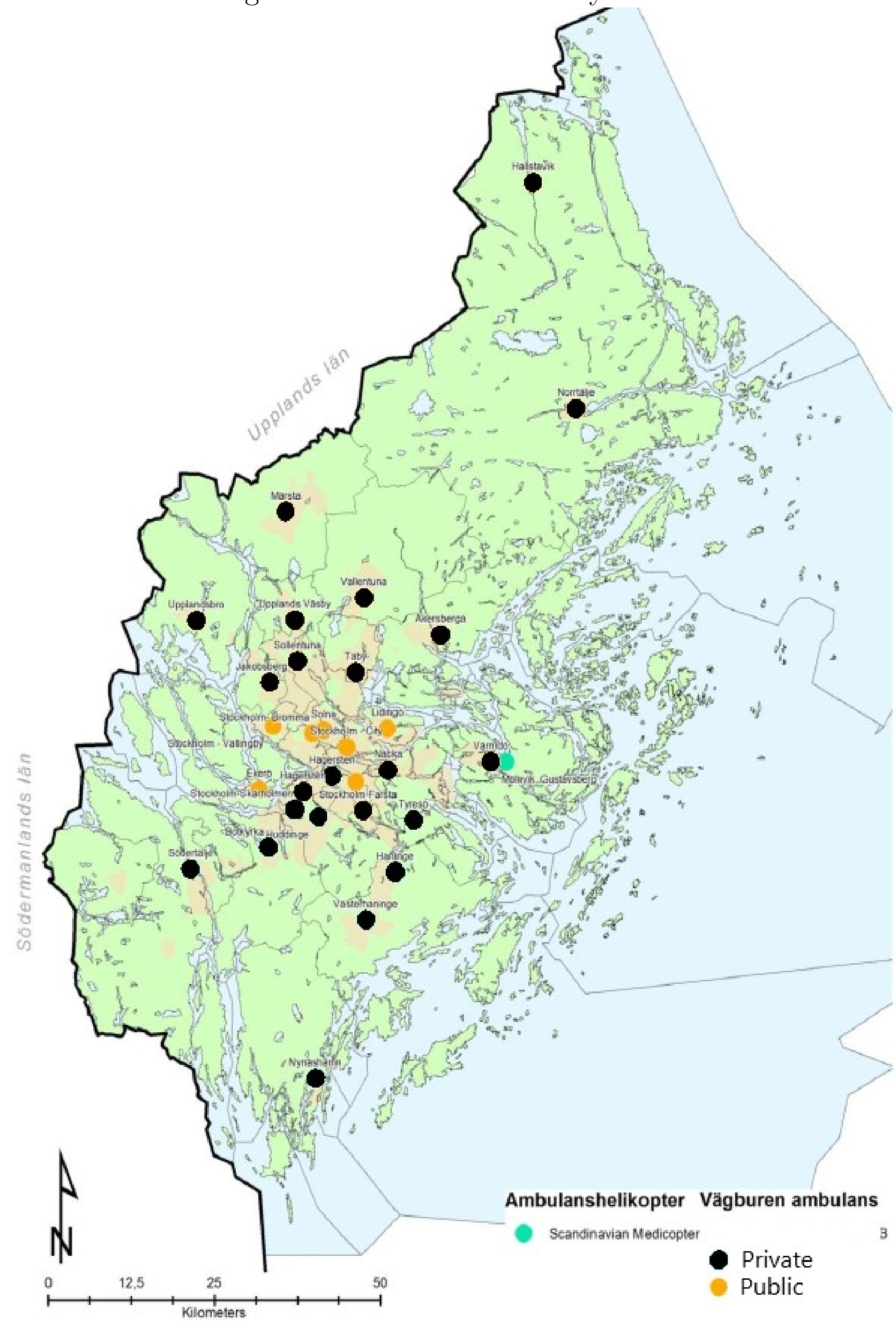
July 1, 2021

Contents

*Corresponding author: Tyrefors: Research Institute of Industrial Economics (IFN), Box 55665, 102 15 Stockholm (e-mail, bjorn.tyrefors@ifn.se; telephone: +46(0)8-665 4500. and Dep. of Economics, Stockholm University; Knutsson: IFN, P.O. Box 55665, SE-10215 Stockholm, Sweden (e-mail, daniel.knutsson@ifn.se; telephone: +46(0)8-665 4500).

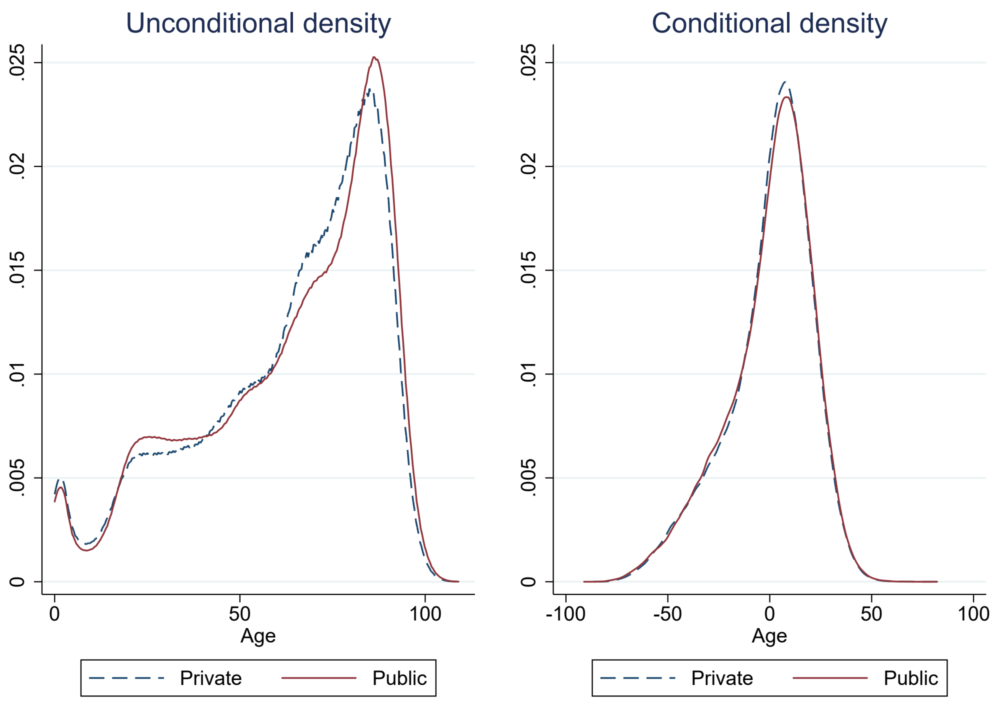
1 Appendix Figures

Figure A1: Stockholm County with ambulance stations



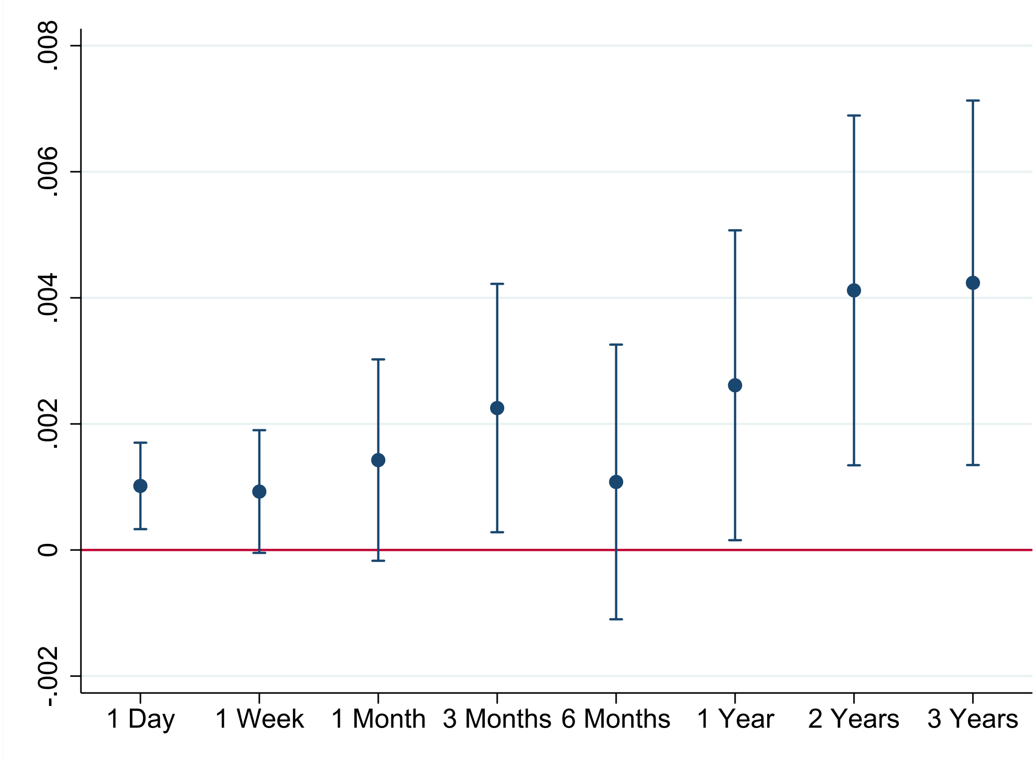
Note: Reproduced from "Årsrapport 2014 prehospitala verksamheter i SLL." The figure shows the geographical reach of Stockholm County with all ambulance stations marked.

Figure A2: Differences in the age distribution of patients



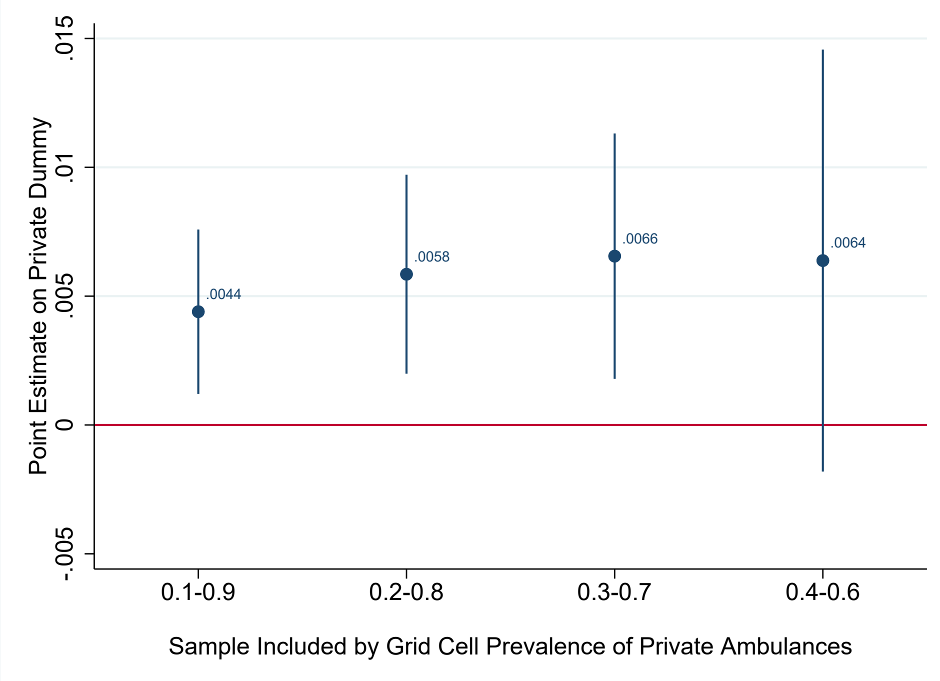
Note: Kernel density plots of the patient age distributions, unconditional (left panel) and conditional (right panel) by private and public ambulance firms. The conditional distribution is residualized by 850x850 grid fixed effects, priority to patient, year, and month dummies according to our preferred specification. In the conditional density distributions, we drop exact zero values, as these come from grids with only one observation, which is more common in private, more rural, areas.

Figure A3: Mortality outcomes



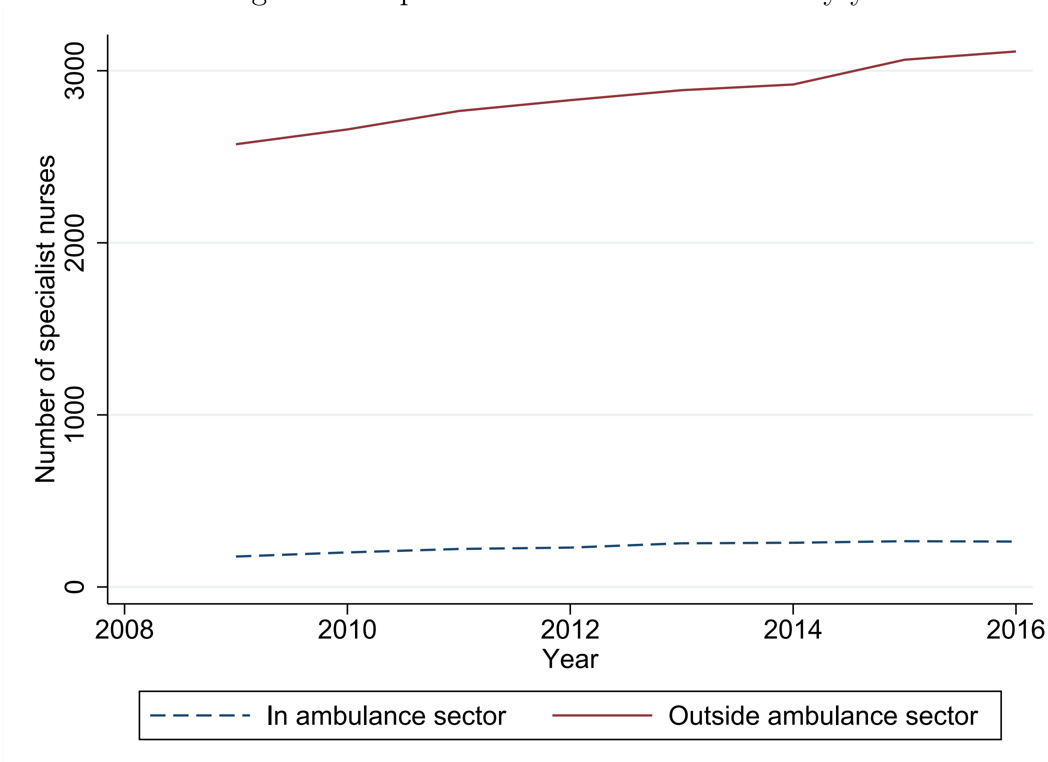
Note: This figure is produced using information from Table 3 in the paper. Each point, with the corresponding confidence interval, displays results from a separate regression. The outcomes are mortality within a time period from each ambulance service, as described on the x-axis. Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190 x 130 meters large, as well as year, month, and priority to patient dummies. The sample is restricted to ground emergency ambulances and excludes extra ambulances. Standard errors are clustered at each grid cell.

Figure A4: Focusing on cells where both public and private ambulances operate



Note: Each estimate in the figure comes from a different regression of 3-year mortality on private ambulance service, following Equation 1 in the paper. The only difference between the models are the samples that are used according to the x-axis. The sample determining the x-axis variable is generated by taking the mean of *Private* at each grid cell (850 x 850) and year. The regression models are then restricted to cells with at least (most) a certain fraction of private ambulance patients. For example, $0.1 - 0.9$ means that only cells for which at least 10 percent, and at most 90 percent, of the patients are serviced by private firms are included in the regression. The number of observations ranges from above 1 million (on the left) to around 100 thousand (the rightmost). Geographical fixed effects in the regressions are based on 850 by 850 grid cells with dummy variables for priority to patient, year, and month included. The sample is restricted to ground emergency ambulances and excludes extra ambulances. Around the point estimates are 95% CIs, where the standard errors are clustered at the grid cell level.

Figure A5: Specialist nurses in Stockholm by year



Note: This figure describes the number of nurses who reside in Stockholm County, have a specialist degree in nursing, and either work in the ambulance sector or in other sectors. This degree makes them eligible to work in ambulances with approximately 10 weeks of introductory training. Using the FAME database, we identify these nurses by their *Sun2000inr* code, describing the type of education each individual in Sweden has. We use the code *723c*, which includes nurses with a degree in pre-hospital care, intensive care, anesthesiology, and surgery. According to the National Board of Health and Welfare Statistics database (`https://sdb.socialstyrelsen.se/if_per/resultat.aspx`), there were 3556 specialist nurses residing in Stockholm in 2012. With our data, we are able to capture more than 3000 of them. Moreover, we separate nurses working in the ambulance sector by conditioning on sector code (SNI2007) *86902*, describing "ambulance transports and ambulance care." To make sure that we are capturing the ambulance firms under study, we further condition on a specific firm having more than 50 employees in any given year.

2 Appendix Tables

Table A1: Summary statistics 1: Health

	Mean	Std.	Min	Max	Obs
Pre-determined characteristics					
Ambulance within 365 d	0.453	0.498	0.00	1.00	1141939
Age	62.251	24.844	0.00	110.00	1138319
Female	0.467	0.499	0.00	1.00	1141939
Nursing home resident	0.070	0.254	0.00	1.00	1141939
Multiple units	0.035	0.184	0.00	1.00	1141939
Diagnoses					
Any health center visit	0.70	0.46	0.00	1.00	1141939
Any hospital visit	0.38	0.48	0.00	1.00	1141939
Health center visits	33.92	47.76	0.00	200.00	1141939
Hospital visits	1.44	3.95	0.00	114.00	1141939
Diabetes	0.10	0.30	0.00	1.00	1141939
Hypertension	0.15	0.36	0.00	1.00	1141939
Heart failure	0.06	0.23	0.00	1.00	1141939
COPD	0.07	0.25	0.00	1.00	1141939
Angina pectoris	0.05	0.22	0.00	1.00	1141939
Opioid disease	0.01	0.09	0.00	1.00	1141939
Depressive disease	0.03	0.18	0.00	1.00	1141939
Alcohol disease	0.06	0.23	0.00	1.00	1141939
Reumatic	0.01	0.09	0.00	1.00	1141939
Cronic pain	0.04	0.20	0.00	1.00	1141939
Anemia	0.02	0.14	0.00	1.00	1141939
Cardiac fibrillation	0.07	0.25	0.00	1.00	1141939
Chronic heart disease	0.05	0.22	0.00	1.00	1141939
Urinary infection	0.06	0.23	0.00	1.00	1141939
Hyperthyroidism	0.02	0.15	0.00	1.00	1141939
Hyperlipidaemia	0.03	0.18	0.00	1.00	1141939
Stroke	0.02	0.14	0.00	1.00	1141939
Pneumonia	0.03	0.17	0.00	1.00	1141939
Benign prostate cancer	0.02	0.12	0.00	1.00	1141939
Back pain	0.04	0.20	0.00	1.00	1141939
Anxiety	0.04	0.19	0.00	1.00	1141939
Malign prostate cancer	0.01	0.12	0.00	1.00	1141939
Group CVD	0.15	0.36	0.00	1.00	1141939
Group Respiratory	0.08	0.28	0.00	1.00	1141939
Group Psychatric/drugs	0.10	0.31	0.00	1.00	1141939
Group Surgery	0.12	0.32	0.00	1.00	1141939
Group Diagnoses >1	0.25	0.43	0.00	1.00	1141939

Source: Data from the Stockholm County VAL database. The "Ambulance within 365 d" variable is a dummy variable indicating any ambulance contacts within 365 days of the index observation. Nursing home residents are pre-registered in the ambulance data. Multiple units refers to ambulance dispatches where more than one unit is assigned to the same patient. Diagnoses and health visits are measured in 2007-2008 from in-patient and out-patient administrative data kept in the VAL database. We create dummy variables indicating any record of each diagnosis. We choose the above diagnoses because they are the most prevalent in the population.

Table A2: Summary statistics 2: Outcomes

	Mean	Std.	Min	Max	Obs
Treatment					
Private ambulance	0.587	0.492	0.00	1.00	1141939
Controls					
Priority to patient	1.71	0.70	1.00	4.00	1141820
Outcomes: Ambulance responses					
Time to ambulance dispatch (s)	921.98	1872.35	1.00	35987.00	1069862
Response time (s)	106.82	224.00	0.00	1800.00	1059247
Time to patient (s)	779.34	525.68	0.00	3600.00	1073220
Assessed severity	2.84	0.99	0.00	7.00	1125465
Patient left at home	0.12	0.33	0.00	1.00	1141939
Outcomes: Ambulance Diagnoses					
Unspecified/general	0.05	0.21	0.00	1.00	1141939
CVD	0.14	0.35	0.00	1.00	1141939
Alcohol abuse/Psychiatry	0.06	0.23	0.00	1.00	1141939
Respiratory	0.09	0.29	0.00	1.00	1141939
Surgery	0.05	0.23	0.00	1.00	1141939
Cardiac arrest	0.01	0.08	0.00	1.00	1141939
Outcomes: Health					
Mortality 3 years	0.30	0.46	0.00	1.00	1141909
Mortality 2 years	0.25	0.43	0.00	1.00	1141909
Mortality 1 year	0.18	0.38	0.00	1.00	1141909
Mortality 6 months	0.13	0.34	0.00	1.00	1141909
Mortality 3 months	0.10	0.30	0.00	1.00	1141909
Mortality 1 month	0.06	0.24	0.00	1.00	1141909
Mortality 1 week	0.03	0.18	0.00	1.00	1141909
Mortality 1 day	0.02	0.13	0.00	1.00	1141909
Outcomes: Hospital admission diagnoses					
Infectious diseases	0.02	0.13	0.00	1.00	1141939
Cancer	0.01	0.12	0.00	1.00	1141939
Endocrinology (Diabetes)	0.01	0.11	0.00	1.00	1141939
Alcohol abuse/Psychiatry	0.03	0.18	0.00	1.00	1141939
Neurology	0.02	0.15	0.00	1.00	1141939
CVD	0.11	0.31	0.00	1.00	1141939
Respiratory diseases	0.06	0.24	0.00	1.00	1141939
Intestinal diseases	0.04	0.19	0.00	1.00	1141939
Arthritis/Orthopedics	0.02	0.14	0.00	1.00	1141939
Urology	0.02	0.16	0.00	1.00	1141939
Surgery	0.06	0.24	0.00	1.00	1141939
Minor trauma	0.07	0.26	0.00	1.00	1141939
Poisoning/burn/fracture	0.02	0.15	0.00	1.00	1141939
Without medical condition	0.01	0.12	0.00	1.00	1141939

Source: Data from the Stockholm County VAL database and SOS Alarm AB. Further information on the data and variables is available in Section 6 of the Online Appendix.

Table A3: Mortality: Close to Stockholm sample

	Mortality after ambulance transport occurred within:							
	1 day	1 week	1 month	3 months	6 months	1 year	2 years	3 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Private ambulance	0.0013	0.0011	0.0012	0.0019	0.0011	0.0031	0.0049	0.0048
	(0.0004)	(0.0005)	(0.0008)	(0.0010)	(0.0011)	(0.0013)	(0.0014)	(0.0015)
Public outcome mean	0.0148	0.0294	0.0580	0.0954	0.1290	0.1758	0.2477	0.3066
Observations	735956	735956	735956	735956	735956	735956	735956	735956

Note: The sample is restricted to ambulance stations within a reasonable commuting distance from central Stockholm (with access to commuting train or subway with less than 30 minutes travel time). The stations that are excluded are located in Norrtälje, Hallstavik, Märsta, Södertälje, Upplands-Bro, Vallentuna, Värmdö, Handen, Tyresö, Ekerö, Västerhaninge, Upplands Väsby, Botkyrka, and Nynäshamn and comprise 32% of the entire sample. Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters large, as well as year, month, and priority to patient dummies. The sample is further restricted to ground emergency ambulances and excludes extra ambulances. Standard errors are clustered at each grid cell.

Table A4: Private ambulance effect on all available time stamps

	Response time	Time to patient	Time at patient	Time to hospital	Time at hospital
	(1)	(2)	(3)	(4)	(5)
Private ambulance	-8.3122	-61.1357	44.9929	47.4279	86.9303
	(0.8752)	(5.2931)	(2.6164)	(5.0690)	(4.1817)
Public outcome mean	107.6759	716.5211	1079.8855	878.0634	1860.1842
Observations	1000902	1017910	952210	913183	834819

Note: Each column shows results from a different regression of a difference between two time stamps on being serviced by a private ambulance. The results in columns 1 and 2 are reproduced from Table 2 in the paper. Since we document in the paper that private ambulances leave significantly more patients at home, the results in columns 3-5 use a population that has been selected based on the treatment, and conclusions must be drawn with caution. Each model includes 850x850 grid fixed effects as well as year, month, and priority to patient dummies. Standard errors are clustered at each grid cell.

Table A5: Does the effect depend on whether the patient is from a private or public area?

	Public area	Private area
	(1)	(2)
Private ambulance	0.0038	0.0047
	(0.0025)	(0.0017)
Public outcome mean	0.3080	0.3024
Observations	418744	655893

Note: Each column describes the results from a separate regression. The two samples used are mutually exclusive. From each ambulance dispatch, we use information describing in which ambulance district the patient is located and map these districts to the different firms. These districts do not restrict the decisions of the dispatch operators assigning ambulances but normally describe the nearest ambulance station. Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters large, as well as year, month, and priority to patient dummies. Standard errors are clustered at each grid cell.

Table A6: Main outcomes for the contract period 2012 to 2016

	Response time	Time to patient	Mortality 3 yrs
	(1)	(2)	(3)
Private ambulance	-7.4612	-71.8774	0.0058
	(1.0191)	(5.1547)	(0.0016)
Public outcome mean	103.1392	693.4347	0.2995
Observations	672429	685700	729809

Note: Each column presents a different regression estimate. Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters large, as well as year, month, and priority of patient dummies. The sample is restricted to the years 2012-2016. Standard errors are clustered at each grid cell.

Table A7: Ambulance firms and the movements of their employees

	Turnover	Change firm	Exit sector	Public (private) to private (public)	To public hospital
	(1)	(2)	(3)	(4)	(5)
Private ambulance	0.085	0.041	0.044	-0.004	0.009
	(0.011)	(0.005)	(0.009)	(0.004)	(0.006)
Public outcome mean	0.1277	0.0289	0.0988	0.0289	0.0345
Observations	5436	5436	5436	5436	5436

Note: Each model includes year dummies and covers the years 2009-2016 unless otherwise noted below. The data includes all employees with less than 5 years of education and an annual income of less than 80,000 EUR (excluding managers and physicians) in ambulance firms that have more than 50 employees. In column 1, the outcome is turnover. This dummy variable takes the value 1 for an employee in a given year if the employee is observed in a specific ambulance firm in year t but not in $t+1$. This variable is a combination of employees changing firms and employees exiting the sector for other employment. Those outcomes are also the variables displayed in columns 2 and 3. In column 2, changing firms means that an employee works in a specific ambulance firm in year t but in a different ambulance firm in $t+1$. The outcome in column 3 is a dummy variable denoting if an employee exits the ambulance sector. This variable is coded as 1 if an employee is observed in year t in any of the ambulance firms but not in any following years. Although some of the estimated relationship between private firm and turnover could be due to firms switching ambulance districts after the auction in 2011, we find that both the turnover and the likelihood of changing firms are positive and significant when using only the later contract period (i.e., 2012-2016). In column 1, the association between private firm and turnover is approximately 0.04 (se0.013), and the corresponding result in column 3 is close to 0.03 (se=0.006) in this reduced sample. The relationship between private employment and exits is more affected in the reduced sample, where the estimate becomes insignificant and highly attenuated (estimate: 0.008 (0.011)). This could suggest that exits are more common late in each contract period or in connection to an auction. In column 4, the outcome is an indicator variable that is equal to 1 if an employee is observed in a public (private) firm in year t and in a private (public) firm in year t+1. Finally, in column 5, we create a dummy variable that equals 1 if an employee is observed in an ambulance firm in year t and in a public hospital in year $t+1$.

Table A8: Balance of predetermined variables

	Estimate of ***Private***	
	Bivariate	Basic model
Outcome:	(1)	(2)
Age	-1.1992 (0.0844)	0.1163 (0.0680)
Female	0.0192 (0.0025)	0.0001 (0.0014)
Any health center visit	0.0147 (0.0021)	-0.0031 (0.0013)
Any hospital visit	0.0022 (0.0027)	0.0005 (0.0015)
Hospital visits normalized	-0.0172 (0.0070)	0.0016 (0.0030)
Nursing home resident	-0.0130 (0.0061)	0.0006 (0.0005)
Ambulance within 365 d	-0.0111 (0.0039)	0.0025 (0.0015)
Primary care visits normalized	-0.0236 (0.0063)	0.0004 (0.0030)
Multiple units	0.0070 (0.0006)	0.0003 (0.0005)
Diabetes	0.0137 (0.0014)	0.0013 (0.0008)
Hypertension	0.0078 (0.0020)	-0.0002 (0.0010)
Heart failure	-0.0001 (0.0013)	-0.0007 (0.0007)
COPD	0.0013 (0.0013)	0.0004 (0.0007)
Angina pectoris	0.0053 (0.0012)	-0.0007 (0.0006)
Opioid disease	-0.0033 (0.0004)	-0.0003 (0.0003)
Depressive disease	0.0002 (0.0009)	-0.0003 (0.0005)
Alcohol disease	-0.0039 (0.0016)	0.0003 (0.0007)
Reumatic	0.0010 (0.0004)	0.0002 (0.0003)
Cronic pain	0.0034 (0.0011)	-0.0005 (0.0005)
Anemia	0.0003 (0.0007)	0.0008 (0.0004)
Cardiac fibrillation	-0.0009 (0.0013)	0.0005 (0.0007)
Chronic heart disease	0.0037 (0.0011)	-0.0005 (0.0007)
Urinary infection	-0.0009 (0.0013)	0.0010 (0.0007)
Hyperthyroidism	-0.0025 (0.0007)	-0.0002 (0.0004)
Hyperlipidaemia	0.0036 (0.0008)	0.0000 (0.0005)
Stroke	-0.0005 (0.0007)	0.0002 (0.0004)
Pneumonia	0.0017 (0.0008)	0.0003 (0.0005)
Benign prostate cancer	0.0007 (0.0005)	0.0000 (0.0003)
Back pain	-0.0006 (0.0011)	-0.0011 (0.0006)
Anxiety	-0.0015 (0.0011)	-0.0011 (0.0006)
Malign prostate cancer	0.0013 (0.0005)	0.0001 (0.0003)
Group CVD	0.0050 (0.0022)	0.0002 (0.0011)
Group Respiratory	0.0024 (0.0015)	0.0011 (0.0008)
Group Psychatric	-0.0064 (0.0020)	-0.0008 (0.0009)
Group Surgery	0.0048 (0.0016)	-0.0004 (0.0009)
No hospital visits 2007-2008	0.0072 (0.0025)	0.0001 (0.0013)

Note: This table replicates Figure 4 in the paper, using predetermined characteristics as outcome variables estimating Equation 1 from the paper. Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters large, as well as year, month, and priority to patient dummies. Standard errors are clustered at each grid cell.

Table A9: Mortality conditional on being transported

	Mortality after ambulance service occurred within:							
	1 day	1 week	1 month	3 months	6 months	1 year	2 years	3 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel 1. Deceased and not conveyed								
Private ambulance	0.0006	0.0008	0.0011	0.0019	0.0029	0.0041	0.0059	0.0075
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0005)
Public outcome mean	0.0064	0.0068	0.0078	0.0095	0.0115	0.0146	0.0202	0.0246
Observations	1075977	1075977	1075977	1075977	1075977	1075977	1075977	1075977
Panel 2. Deceased and conveyed								
Private ambulance	0.0004	0.0002	0.0003	0.0003	-0.0019	-0.0014	-0.0018	-0.0033
	(0.0003)	(0.0004)	(0.0008)	(0.0010)	(0.0011)	(0.0012)	(0.0014)	(0.0015)
Public outcome mean	0.0084	0.0226	0.0503	0.0859	0.1176	0.1611	0.2275	0.2821
Observations	1075965	1075965	1075965	1075965	1075965	1075965	1075965	1075965
Panel 3. Deceased all (Panel 1 + Panel 2)								
Private ambulance	0.0010	0.0009	0.0014	0.0023	0.0011	0.0026	0.0041	0.0042
	(0.0003)	(0.0005)	(0.0008)	(0.0010)	(0.0011)	(0.0013)	(0.0014)	(0.0015)
Public outcome mean	0.0148	0.0294	0.0580	0.0954	0.1290	0.1758	0.2477	0.3066
Observations	1075958	1075958	1075958	1075958	1075958	1075958	1075958	1075958
Panel 4. Survived and not conveyed								
Private ambulance	0.0315	0.0314	0.0311	0.0302	0.0292	0.0281	0.0263	0.0247
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0010)	(0.0010)	(0.0010)
Public outcome mean	0.0946	0.0942	0.0932	0.0915	0.0896	0.0864	0.0808	0.0764
Observations	1075977	1075977	1075977	1075977	1075977	1075977	1075977	1075977
Panel 5. Survived and conveyed								
Private ambulance	-0.0325	-0.0323	-0.0325	-0.0325	-0.0303	-0.0307	-0.0304	-0.0289
	(0.0011)	(0.0012)	(0.0012)	(0.0013)	(0.0014)	(0.0014)	(0.0015)	(0.0015)
Public outcome mean	0.8906	0.8764	0.8487	0.8131	0.7814	0.7379	0.6715	0.6169
Observations	1075965	1075965	1075965	1075965	1075965	1075965	1075965	1075965
Panel 6. Survived all (Panel 4 + Panel 5)								
Private ambulance	-0.0010	-0.0009	-0.0014	-0.0023	-0.0011	-0.0026	-0.0041	-0.0042
	(0.0003)	(0.0005)	(0.0008)	(0.0010)	(0.0011)	(0.0013)	(0.0014)	(0.0015)
Public outcome mean	0.9852	0.9706	0.9420	0.9046	0.8710	0.8242	0.7523	0.6934
Observations	1075958	1075958	1075958	1075958	1075958	1075958	1075958	1075958

Note: The mortality outcomes in the panels above are constructed by setting mortality to zero (instead on one) for patients that are conveyed in Panel 1 and for those that are not conveyed in Panel 2. Panel 3 shows the results from Table 3 of the paper for reference and is by construct the sum of the results from the first two panels. Approximately 12% of all patients are non-conveyed, and our results show that private firms leave 3.2 percentage points more patients at home. In panels 4-6, we recode the mortality variables to indicate their opposites, namely, survival, and preform the same exercise. Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters large, as well as year, month, and priority to patient dummies. Standard errors are clustered at each grid cell.

Table A10: Which patients are more affected by private ambulances? Heterogeneity analysis.

	Outcome variable: 3 year mortality								
Interaction variable	Priority = 1	Male	Age> 70	No Hospital visit 2007-2008	Multiple diagnoses	CVD	Respiratory disease	Psychiatric/ Drugs	Surgery
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Private x Variable	0.0163	0.0120	0.0099	0.0083	-0.0004	-0.0010	-0.0074	0.0030	-0.0066
	(0.0023)	(0.0027)	(0.0029)	(0.0031)	(0.0038)	(0.0047)	(0.0062)	(0.0051)	(0.0053)
Private ambulance	-0.0020	-0.0013	-0.0017	-0.0010	0.0043	0.0044	0.0047	0.0039	0.0051
	(0.0021)	(0.0019)	(0.0018)	(0.0024)	(0.0017)	(0.0016)	(0.0015)	(0.0016)	(0.0016)
Variable	-0.0610	0.0385		-0.1186	0.1647	0.2146	0.1755	-0.0410	0.1051
	(0.0020)	(0.0023)		(0.0028)	(0.0034)	(0.0043)	(0.0055)	(0.0047)	(0.0047)
Interaction variable mean	0.4235	0.4633	0.4740	0.6206	0.2469	0.1538	0.0847	0.1060	0.1167
Observations	1076073	1075958	1072718	1075958	1075958	1075958	1075958	1075958	1075958

Note: Each column describes the results from a separate regression estimating the interaction effects of private ambulance on 3-year mortality. The variable used for the interaction is noted above each column number. In column 1, we interact private ambulance with a dummy variable indicating if the dispatch was of the highest priority. Some caution with the interpretation of these results is necessary, as the dispatch priority is potentially part of the treatment allocation mechanism. In column 2, we interact private ambulance with a dummy variable for male sex. In column 3, we interact private ambulance with a dummy variable for age above 70 (and include a full set of age fixed effects). In column 4, we interact private ambulance with a dummy variable indicating if a patient had no hospital visits in 2007-2008. To generate the interaction with private ambulance, columns 5-9 use dummy variables based on diagnoses in 2007-2008. In column 5, the interaction variable is a dummy that denotes having more than 1 of these diagnoses. We map the diagnoses to broader categories by using the first letter of the ICD code (see Section 6 on the data in this Online Appendix). Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters large, as well as year, month, and priority to patient dummies. The sample is restricted to ground emergency ambulances and excludes extra ambulances. Standard errors are clustered at each grid cell.

3 Coefficient stability and standard errors

To make sure that our results are not just valid under the exact model specification that we apply, the Online Appendix, Figure A8, shows how our main result is affected by changing the size of the fixed effects. Simultaneously, we change the level of clustering to match our fixed effects. We start from 200x200 grids, approximately 800 times 550 meter cells, and go down to 1000 by 1000 grids (160 times 110 meters). We find that neither the point estimates or the standard errors are affected in any important way from these alterations. Clustering at the grid cell level increase our standard errors slightly compared to calculating robust only standard errors. In our preferred specification the t-value of the coefficient estimated on the Private dummy decreases from 3.49 with robust standard errors, to 2.87 when clustering at the grid cell level.

Another way to approach coefficient stability is to use the method described in Oster (2019). This approach assumes that the unobserved selection is proportional to the observed selection. By observing how the treatment coefficient changes as controls are included and relating this change to the change in the R^2, an estimate of how much variation needs to be explained to nullify the estimate can be calculated (the delta value). To be able to use this approach, we residualize the outcome, treatment, and all control variables by using our main specification before running the program *psacalc*. We let the R^{MAX} be 0.85 to account for the fact that our fixed effects can explain approximately 15% of the variation in 3-year mortality. We find that unobservables would have to explain 48% of the variation in the outcome to nullify our treatment effect completely.

An other way to approach coefficient stability is to use the method described in Oster (2019). This approach assumes that the unobserved selection is proportional to the observed selection. By observing how the treatment coefficient change as controls are included, and relating this movement to the change in the R^2, an estimate of how much additional variation needs to be explained to nullify the estimate can be calculated (the δ value). To be able to use this approach, we residualize the outcome, treatment, and all control variables by using our main specification (with 850x850 grids) before running the program *psacalc* in STATA. We follow the recommendations in Oster (2019), as well as other papers applying the test (Feigenberg and Miller, 2021), and let the R^{MAX} be 1.3 times the adjusted R^2 from the long regression specification.

We find that selection on unobservables would have to be 6 times as large as the selection on observables to explain our treatment effect completely. This in a data setting where we explain mortality using 31 control variables on age, diseases, and other characteristics that are likely to be able to explain much of the mortality we observe. As the suggested δ value

Figure A6: Three-year mortality and different grid fixed effects

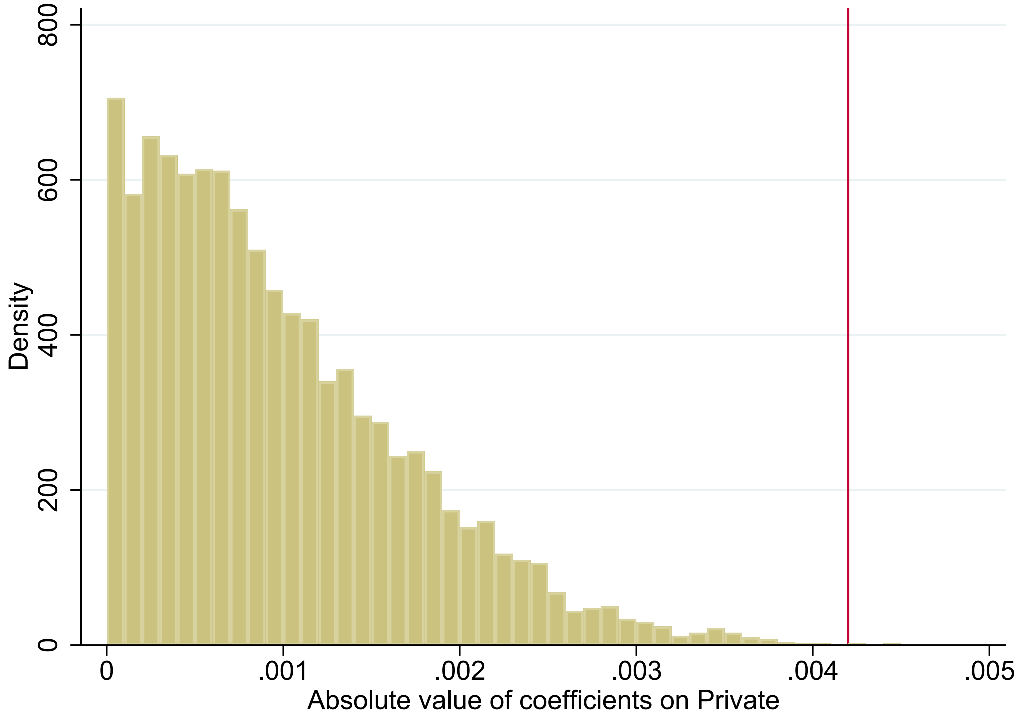
Note: This figure shows estimates of *Private* by using three-year mortality as the outcome. Each connected point represents a different regression. We use different grid sizes to define the fixed effects, from 200x200 grids (to the left) and up to 1000x1000 grids (to the right). The 200x200 grids correspond to cells sized approximately 0.8 km by 0.55 km, while the 1000x1000 grids correspond to about 160 by 110 meters. Standard errors are clustered at the grid cell level. In the figure above, the estimated t-value for the private dummy coefficient estimate obtained by using our preferred specification (850x850 grid cells) is 2.87 (se = 0.0015). The corresponding t-value when clustering at the individual level is 3.46 (se = 0.0012) and with robust standard errors it is 3.49 (se = 0.0012).

for results to be robust is 1, we conclude that our design is robust to unobserved selection in the Oster (2019) sense.

Randomization inference

To validate if the coverage of our standard errors are too small given our design, we perform randomization inference on our main outcome variable. To do this, we make use of the *ritest* package in STATA (Heß, 2017). The concern that we address is that our results might not be very uncommon given the conditionally random treatment allocation but that a large sample size permits for small standard errors or that few observations have large impacts on the estimated coefficient (Young, 2018). Randomization inference generates a counterfactual distribution of estimates under the strict NULL hypothesis of no treatment effect unrelated

Figure A7: Three-year mortality: Randomization inference



Note: This figure shows the density of estimates of *Private* when using three-year mortality as the outcome. To each patient with a given rate defined by the true fraction of private within each grid cell, we randomly allocate Private ambulance service 5000 times. The red vertical line indicates the estimated effect from our preferred specification. We take the absolute value of the coefficients since we have a two-sided test of significance.

to the sample size. This counterfactual distribution informs us on the likelihood of finding large estimated coefficients under alternative random assignments. Now, we are abstracting from sampling variability and considering only the uncertainty stemming from the unknown potential outcomes. We can thereby provide information on the within sample uncertainty under the assumption that the treatment effect is zero for all participants (Cunningham, 2021, chapter 4.2).

We run 5000 permutations, allocating treatment at random within each grid cell (850x850). The probability of treatment within each grid cell is fixed from the original data set. From each permutation, we extract the beta coefficient on our treatment variable (Private ambulance) and construct a new dataset with the 5000 coefficient estimates. To facilitate the visual interpretation of a two-sided test, we take the absolute value of these coefficients and plot them in a histogram (see Online Appendix Figure A7). We find that the p-value of our reported effect (0.0042) from this exercise is 0.0002 and that only one permutation succeeded in producing a larger estimate than ours in absolute terms.

4 The unmeasured cost of private ambulances

Outsourcing public services to private firms helps the public sector reduce costs through private sector cost innovation. However, as our results show that in this context, private firms produce at lower quality, the cost savings from outsourcing will be lower than would be assumed at face value (Mukherjee, 2020).

Assigning a price to mortality requires us to make several unverifiable assumptions. First, we start by assuming that our Table A3 estimates of mortality 1 to 3 years after an ambulance contact are reasonable representations of provider differences in deaths each year. Since these estimates are cumulative differences in deaths, we can sum them to obtain a crude measure of years lost for up to three years after each ambulance contact.[1] However, we must also decide on what to assume about years lost beyond our observed window in time. We start by assuming that mortality converges between providers after 6 years. This means that we observe an increasing mortality difference for three years and then a decrease at the same rate for the following three years. As we know that the population we study is to a large extent old and frail, this seems reasonable. The years of life lost can thereby be calculated as:

$$\text{Years Lost}_t = 2 * (\hat{\beta}_{1Year} + \hat{\beta}_{2Years} + \hat{\beta}_{3Years}) * N_t^{Private}. \tag{1}$$

In Equation 1, we describe how we calculate the number of years lost for year t as twice the sum of yearly lost years of life until year $t + 2$.[2] We multiply by 2 to approximate the mortality convergence over time (4-6 years later), as noted above. As the private firms in our data service approximately 100,000 patients each year, we find that:

$$YearsLost = 2 * 0.01 * 100,000 = 2,000. \tag{2}$$

Using this number of years lost, we add a standard cost of 100,000 EUR on each lost year of life (ICER, 2018).[3] Moreover, we quality adjust each year of life lost by a constant, reflecting that those most affected by ambulance services were old and in poor health, with a rather short life expectancy. This gives us Quality Adjusted Life Years (QALY). Setting

[1] We approximate cumulative mortality as the sum of the estimates. In this way, we somewhat overestimate the cumulative years of life lost, as we do not consider changes within the years. However, we believe that the increased complexity of disposition out weights the gain in precision for this kind of exercise.

[2] The coefficients on the private ambulance dummy for year 1 to year 3 are 0.0026, 0.0041, and 0.0042. The sum equals 0.0109, which we for ease of exposition round to 0.01.

[3] In Sweden, the National Board of Health and Welfare describes a treatment increasing one QALY at a cost of 10,000 EUR as having a low cost, between 10,000-50,000 EUR as having a moderate cost, between 50,000-100,000 EUR as having a high cost, and over 100,000 EUR as having a very high cost (NBHW, 2018).

this weight to 1 would imply that all years lost are years in perfect health, which we find overly optimistic. Instead, we use two arbitrary lower weights, 0.1 and 0.5, and calculate the total cost per year based on the QALYs lost.

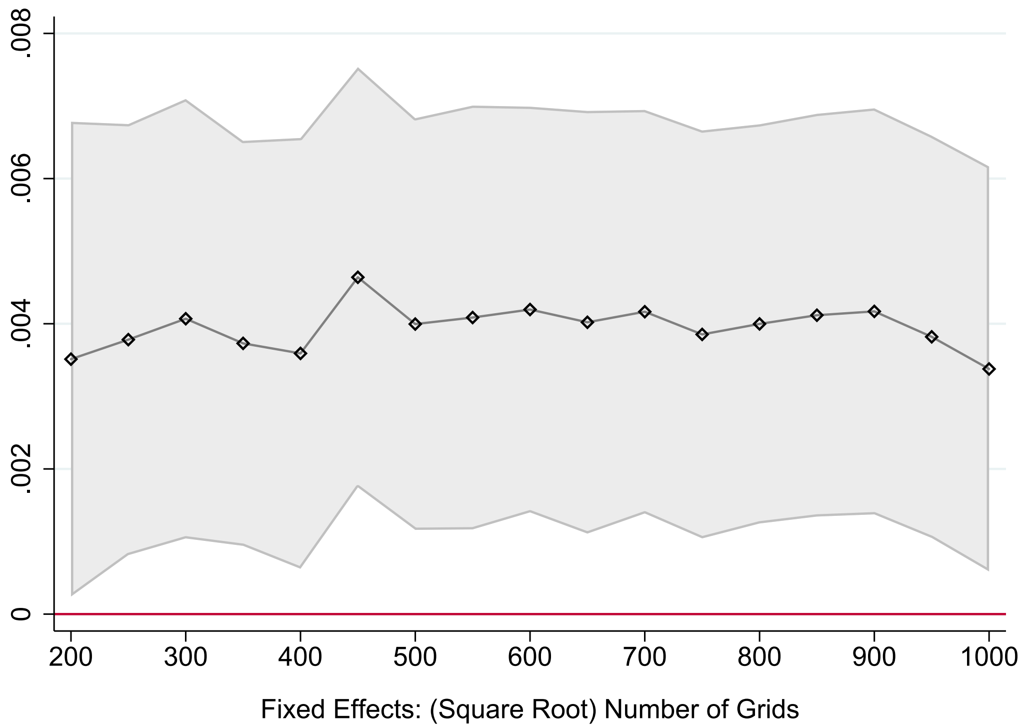
Starting with the highest weight, we assume that each year of life lost has half the quality of years in perfect health. Multiplying the cost per year of life (100,000) by the number of years lost (2,000) and the weight (0.5) yields a total yearly cost of 100 million EUR. Using the more conservative weight of 0.1 yields a total yearly cost of 20 million EUR, which is around half the total cost of the ground-based ambulance services in Stockholm in 2012 (HSF, 2014).

Even with more conservative estimates, the total cost of years of life lost is large. Considering the lowest QALY weight of 0.1 and only half of the years of life lost (assuming that the effect is zero after three years), the yearly cost is still 10 million EUR, a quarter of the total cost of ambulances in Stockholm in 2012 (HSF, 2014). As we have shown previously, outsourcing ambulances in Stockholm could generate savings of up to 6.2 million EUR each year. Even if we adopt a cautious and conservative approach when calculating the cost of increased mortality, our calculations suggest that the cost of reduced quality vastly outweighs the cost savings that private firms generate.

5 Results for 2-year mortality

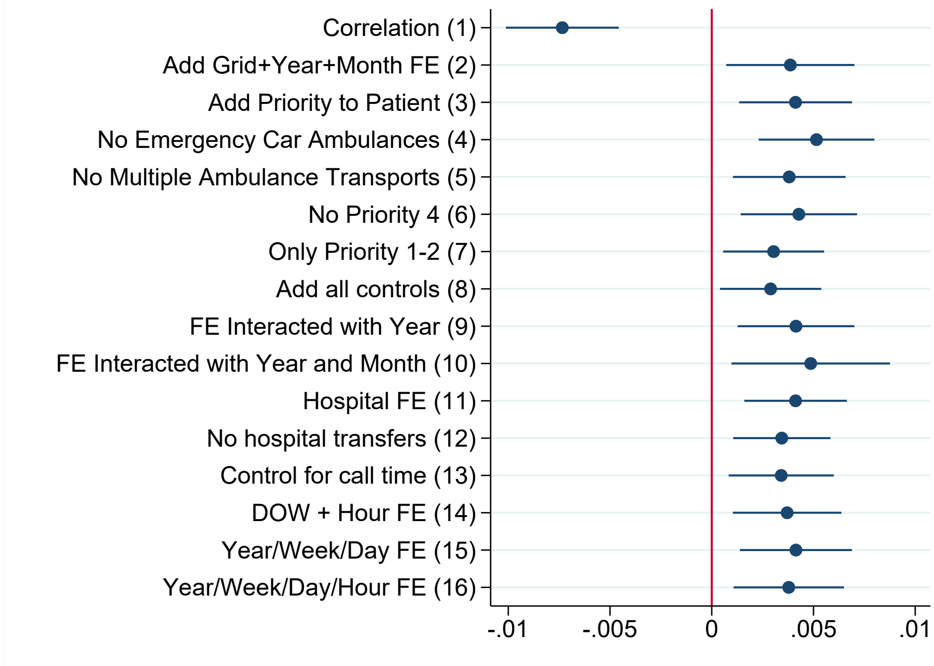
To make credible that our results are not just present in models using our preferred outcome variable, i.e., three-year mortality, we present below robustness tests for two-year mortality. First, we vary the size of the grid cells (and thereby the level of clustering), as in Figure A8. We then replicate the robustness analysis from Figure 4 of the paper. The results are very similar to those where we use three-year mortality as the outcome.

Figure A8: Two year-mortality and different grid fixed effects



Note: This figure shows the estimates of *Private* when using two-year mortality as the outcome. We use different grid sizes to define the fixed effects, from 200x200 grids (to the left) and up to 1000x1000 grids (to the right). The 200x200 grids correspond to cells sized approximately 800 by 550 meters, while the 1000x1000 grids correspond to about 170 by 110 meters. Standard errors are clustered at the grid size level. The estimated t-value for the private dummy coefficient estimate in the figure above using our preferred specification (850x850 grid cells) is 2.91 (se = 0.0014). The corresponding t-value when clustering at the individual level is 3.54 (se = 0.0012), and with robust standard errors, it is 3.57 (se = 0.0012).

Figure A9: Robustness: Deviations from the basic model (2-year mortality)



Note: This figure presents point estimates and 95% confidence intervals for different specifications of Equation 1 in the paper. The first two rows impose fewer restrictions than Equation 1. The first row is the raw correlation between 2-year mortality and private ambulance providers. Row 2 adds grid fixed effects, leading up to our preferred specification in row 3, which also includes priority to patient indicators. Rows 4 and below challenge this specification by adding structure or restricting the sample in different ways, as described in the figure and paper.

6 Additional notes on the data

In this part of the Online Appendix, we detail much on the variables created and the restrictions we impose that did not fit in the paper.

The VAL database

The main data come from the VAL database, owned by the county of Stockholm (Region Stockholm) and used for evaluation and follow-up. The ambulance part of these data (VAL-ambu) consists of imported data from CAK-net, which is a common database in which all ambulance firms store their medical records for each dispatch. VAL-ambu only imports dispatches for situations where a patient was encountered on site. This means that the ambulance data we use are a secondary source. Since many of the time stamps were missing in these data, we complemented the data with data from SOS Alarm AB, the ambulance dispatch firm that maintains first-hand data on all time stamps. We secured access to the three different time stamps where we had many missing observations (we obtained "dispatch sent," "arrival at patient," and "ready for a new assignments"). As we were informed, these missing data could be due to problems importing data to either CAK-net or to VAL-ambu. We also secured GPS coordinates on the location of each patient from the same source. This location is the place where the dispatch operator determines the patient to be and is normally based on an address provided at the time of the call. This data from SOS Alarm AB could be matched to VAL-ambu by using a unique dispatch ID number.

Time stamps

Some of the time-stamps gave implausible values for the variables that we created, and we had to truncate the values to reduce the influence of extreme outliers that were obviously incorrect. Since these time stamps require staff to actively register each part of the assignment at the due time and as the data had to be transferred to the CAK-net database, there are ample possibilities for errors. That the staff had difficulties reporting these time stamps had been noted by the county in several evaluations (SLL, 2017). We were able to obtain complementary time stamps from the dispatch service, SOS Alarm AB, and for these time stamps, there were fewer problems with data transfer issues. Still, there exists extreme values that are related to the actual registration. As we create times between the different stamps as variables, we make use of reasonable time frames to truncate the variables. Since these time stamps are likely most erroneous, we truncate them by setting them to missing (see Table A11).

First, we restrict the time from a call to a dispatch to 10 hours. This is the time between a call registered at the dispatch center and an ambulance being assigned to the patient. Allowing for 10 hours is an inclusive approach, as few patients have to wait 10 hours or more before an ambulance is assigned to them. However, low priority transports (priority 3 and 4) could have to wait for some time to keep readiness for emergencies, especially at times when ambulances were high in demand. It is also possible that institutions, such as hospitals, would order an ambulance for transporting a patient to, for example, an x-ray at a different location. This could be done well in advance.

Second, we cut the response time to 30 minutes. If the ambulance has not responded sufficiently fast, the dispatch services will contact them by using the radio or telephone. Walking to the car and confirming the assignment will never take 30 minutes, but it could appear to if the staff forgets to confirm, the dispatch call is not received by the ambulance crew, or the confirmation fails to be registered.

Third, we cut the transport time to the patient to one hour. This is the time from confirmation to arrival at the patient's location. Although some transports within the county could take more than an hour, it is extremely rare that an ambulance is assigned a patient with an expected transport time of one hour or more.

The number of truncated observations are presented in Table A11, where we also include later time stamps not used in the paper. These are time at patient, time to hospital, and time at hospital. The reason that we do not use these time stamps in the paper is that, first, they are not followed as closely by the county and not contracted on or discussed in follow-up meetings. Second, since we find that providers differ in their propensity to leave patients at home, they are endogenous. Taken at face value, public ambulances are faster from their patient to the hospital, even though our design forces the distance to be very similar between different patients. Furthermore, we find that private ambulance crews stay longer at the hospital after leaving their patients. If this behavior is due to differences in the required time to resupply the ambulances and to write a report or if the differences are due to shirking is impossible to determine given that their patients are, on average, most likely different (due to differences in conveyance rates).

Fourth, we restrict the time at the patient to 4 hours. Spending more than four hours at a scene of an injury rarely occurs. It is possible at a very large accident scene with many injuries or at a fire where the health care needs are unknown for a long time, but in general, these high values will be errors.

Fifth, we cut the time to hospital to two hours. That we set this truncation higher than the time to the patient is because some patients require specialist care and must thereby be transported to a hospital potentially far away from their current location.

Sixth, and finally, we truncate the time at hospital to two hours. Currently, new rules set ambulances as being available for a new assignment after 25 minutes unless the crew asks for more time. Two hours is a very long time to spend at a hospital after leaving a patient. Sanitizing a car takes some time, but more than two hours will generally come from crews forgetting to register that they are ready for a new assignment or the failure to import data.

Table A11: Truncated time stamps

Variable	Truncated at (minutes)	Truncated observations	Valid observations
Dispatch time	600	72,026	1,069,862
Response time	30	29,080	1,059,247
Time to patient	60	5,688	1,073,220
Time at patient	240	745	1,013,065
Time to hospital	120	2,285	969,998
Time at hospital	120	6,678	884,408

Note: Authors calculations from data.

Data restrictions

Data restrictions come from cleaning the data and from making additional restrictions as robustness checks. These restrictions are detailed below. In cleaning the data, we found that approximately 1000 observations did not have a clear provider reported. Without a provider, we cannot assign treatment status to an ambulance. Furthermore, we cannot use observations that are missing on other important variables that we use in our basic model. We drop observations that are missing on coordinates (n=13,088) and priority to patient (n=118). Moreover, 3,451 observations on patient coordinates are dropped for being outside Stockholm County. We are not using dispatches to nearby counties, although Stockholm ambulances can assist ambulances from nearby counties. The summary statistics tables in the paper show the number of missing observations for the different variables.

We focus attention to units described as emergency ambulances (akutambulans), certified ambulances (cert. ambulans) and back-up ambulances (reservambulans). The units that we exclude are the following: transport ambulances that have lower requirements in terms of staff education level (and that were phased out between 2009 and 2011); mobile intensive care units that transport patients between different intensive care units; ambulances equipped to transport newborn babies in incubators; one out-of-county unit that was only used to transport patients to and from Stockholm County health care facilities; extra ambulances that can be operational on demand from the county at times of exceptionally high demand, such as New Year's Eve; and one psychiatric ambulance that was instituted late in the sample and only responded to patients with a high risk of immediate suicide. Jointly, these constitute approximately 38,500 observations in the data that are excluded.

In the robustness analysis in the paper (Figure 4, row 5), we restrict the sample to single-unit dispatches only. When we restrict the sample to single-unit dispatches, we make three exclusions that the data permit. First, we remove all units where a checkbox was marked indicating whether an emergency car accompanied the ambulance. Second, we exclude transports where any information indicates that a second unit was dispatched (several variables are provided that have this information, e.g., time stamps for second and third units). This is because each patient (ambulance journal) is an observation and if several ambulances are dispatched to the same patient, they end up in the same observation. Capturing these multiple assignments to a single patient will exclude most of the relevant transports. Third, there is a possibility that ambulances are assigned to the same scene but to different patients. If so, they will obtain the same dispatch number but with different last two digits reflecting that each ambulance is assigned to a different patient. We therefore exclude all assignments with dispatch numbers that are not unique. Jointly, these three ways of excluding multiple ambulance dispatches reduce the sample by 71,883 observations.

In addition, in the robustness analysis (Figure 4, row 4), we exclude emergency car ambulances that have higher staff competencies and are targeted to more severe cases. Before 2012, these units existed but did not have transport capacities, meaning that they could assist an ambulance unit but could not act independently. After 2011, these units were upgraded to transport capacity, with the competence and work description intact. To alleviate concerns that by operating these units, private firms are assigned to more severe cases with a higher death expectancy, we exclude these vehicles from the analysis. Unfortunately, individual units are not well-identified in the data, as the unit description was changing during over time due to a new radio system being implemented. Therefore, we can only exclude the entire ambulance station where these units operated from since the ambulance station is well-recorded in the data. We excluded the stations in Sollentuna and Huddinge. Although somewhat crude, this is the best we can do with the given data.

Variables: Whether the patient was transported by the ambulance or was left at home is precoded in the data. For completion of an ambulance journal, this check-box had to be filled in. The severity of each patient's condition also had to be filled in the journal for each patient encounter. Mortality is generated as a dummy variable by using the death date and relating that to each ambulance dispatch date. The mortality data comes from a separate file in the VAL-database. For these variables, we do not have many missing observations. As mortality is obtained from a different data source and is hence not subject to data transfer errors, it is accurately measured as far as we can tell.

Health data 2007-2008

Health data from 2007-2008,were derived from the VAL database on hospital admissions and out-patient visits. We sampled the most frequent diagnoses in the data and translated the ICD code into a disease. We then created for each diagnosis one variable indicating whether each individual had that diagnosis on any occurrence during hospital or primary care visits between 2007-2008 (see Table 6). From these data, we also created variables describing whether each individual had any hospital/out-patient visits and the total number of visits jointly in 2007 and 2008. For some individuals, we obtained very large values, especially out-patient visits. We interpret these as representing regular (daily) contacts with a primary caregiver for certain services, e.g., delivering daily medicines or managing diabetes. We truncate these variables to 200 for out-patient visits (over the two years) and 114 for hospital care visits (such that values above these cutoffs are assigned the maximum value).

Table A12: ICD-10 codes for health status, 2007-2008

Diagnosis	ICD-10 codes	Description
Diabetes	E10-E14	
Hypertension	I10-I15	
Heart failure	I50	
COPD	J40-J47	Chronic obstructive pulmonary disorder
Angina Pectoris	I20	
Opioid	F11	Psychiatric disorders or behavioral disorders caused by opioids
Depression	F32	
Alcohol	F10	Alcohol induced amnesia, abstinence, intoxication, addiction, psychosis.
Arthritis	M05	
Pain	R52	
Anemia	D64	
Fibrillation	I48	
Heart disease	I25	Non-acute ischemic heart disease.
Urinary infection	N39	
Hyperthyroidism	E03	
Hyperlipidemia	E78	
Stroke	I63	
Pneumonia	J18	
Benign prostate cancer	C61	
Malign prostate cancer	N40	
Back pain	M54	
Anxiety	F41	

Note: The use of three-character ICD codes means that all subcategories are included. Information on ICD codes can be found at `http://icd.internetmedicin.se/`.

Ambulance crew diagnoses

The ambulance crew must enter at least one diagnosis ("Bedömt tillstånd") for each patient. This diagnosis is chosen from a preexisting set of approximately 160 choices. From the data, Table 6 reports a sample of these primary diagnoses that have more than 30,000 observations each. From these data, we created several variables to stress the different assessments made by the ambulance crews (see Table 4 in the paper). Table 6 shows what information we have included in the variables.

Table A13: Diagnoses from ambulance journals with more than 30,000 obs.

Diagnosis	Translation	Observations
Andningsbesvär	Respiratory problems	61,829
Buksmärta	Abdominal pain	73,759
Centrala bröstsmärtor - kärlkramp	Central chest pain - angina pectoris	66,389
Fraktur	Fracture	39,521
Medicin ospecificerat	Medicine unspecified	41,175
Skador/olycksfall ospecificerat	Injury/accident unspecified	32,315
Yrsel	Dizziness	35,282

Note: Plain text examples of primary diagnoses set and recorded in the ambulance journals.

Table A14: Variables on diagnoses created from ambulance journal information

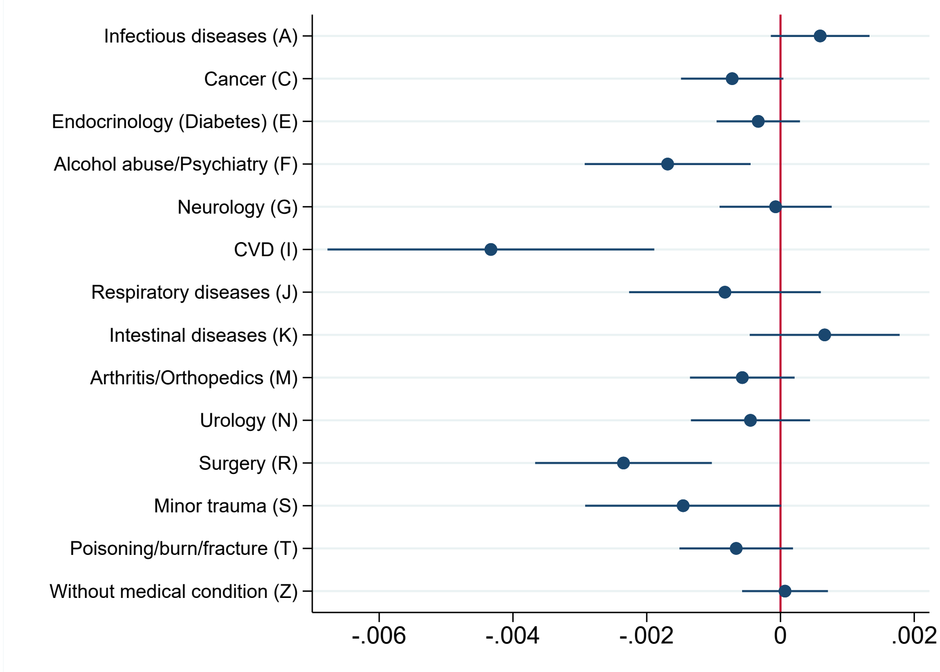
Unspecified/general	
Allmänt ospecificerat	
Allmänt övrigt	
Psychatric/drugs	**Respiratory diseases**
Drog-/läkemedelsmissbruk	Andning ospecificerat
Förgiftning med alkohol	Andning övrigt
Förgiftning med alkohol och läkemedel	Andningsbesvär
Förgiftning med födoämne inkl svamp	Andningsbesvär m främmande kropp
Förgiftning med läkemedel	Andningsbesvär m pip astma
Förgiftning med narkotika	Andningsstillestånd
Förgiftning med tobak	Pseudokrupp
Förgiftning petroleumprod inkl etylen..	Pneumothorax
Förgiftning ospecificerat	Lunginflammation
Förgiftning övrigt	Lungemfysem
Hallucinationer	Kronisk obstruktiv lungsjukdom
Krisreaktion	Kronisk bronkit
Psykiatri ospecificerat	Influensa
Psykiatri övrigt	Epiglottit
Psykiska symtom och sjukdomar	
CVD	**Surgery**
Aortaaneurysm	mage-tarm
Centrala bröstsmärtor - kärlkramp	Sjukdomar i bukspottskörteln
Cirkulation	Magsår
Cirkulation, ospecificerat	Kirurgi övrigt
Cirkulation, övrigt	Kirurgi ospecificerat
Embolier tromboser - artär	Bröstkorgssmärtor
Embolier tromboser - ven	Bråck
Hjärnblödning/hjärninfarkt	Blödning från mun, svalg, mage-tarm
Hjärtinfarkt	Ileus
Hjärtstopp - asystoli - PEA	
Hjärtstopp - med framgångsrik HLR	
Hjärtstopp-ventrikelflimmer	
Hjärtsvikt	
Hypertoni	
Lungödem	
Hypotoni	**Cardiac arrest**
Subarachnoidalblödning	Hjärtstopp - asystoli - PEA
Transitorisk ischemisk attack	Hjärtstopp - med framgångsrik HLR
Rytm- eller retledningshinder	Hjärtstopp-ventrikelflimmer

Note: Plain text primary diagnoses set and recorded in the ambulance journals and included in the variables we create. The variables are binary and indicate if any of the above diagnoses have been assigned to a patient.

Primary cause of hospital admissions

We use the primary diagnosis for approximately 631,000 hospital admissions and match these diagnoses to the same day, or the day after, each ambulance service. Since admissions to a hospital often take some time at the ED, intake registration can also occur in the following day. These ICD-10 codes are separated by the leading letter, which are the large groups of admission causes that we use as binary outcome variables. To reduce the number of these categories, we exclude categories with less than 8,000 records in the data. There are 11 of these rare intake diagnoses of which 6 have less than 800 records. The ones we use have between 14,208 (endocrine diseases, including for example diabetes) and 121,779 (CVDs, including AMI) records.

Figure A10: The effect of private ambulance on all hospital intake diagnoses groups



Note: This figure presents point estimates and 95% confidence intervals for estimating Equation **??** with dummy variables and with hospital intake diagnoses as outcomes. Each row presents a separate regression and displays each estimated coefficient on the private ambulance dummy. Each variable corresponds to the first letter in each ICD-10 code as described in the figure in parenthesis after each row label. Each model includes 850x850 grid fixed effects covering the entire county of Stockholm, where each grid is approximately 190x130 meters large, as well as year, month, and priority to patient dummies. Standard errors are clustered at each grid cell.

Firm data

To understand more regarding the staff differences between private and public firms, we make use of the FAME database housed at the Institute for Industrial Economics. This database provides rich administrative data on matched employer-employees for the full population of Sweden. We create our sample by taking out the full population residing in Stockholm County between 2009-2017. Moreover, we sample all firms that have the sector code (SNI2007) *86902* representing "ambulance transports and ambulance care" and that have more than 50 employees in a given year. This restriction in number of employees comes from the existence of smaller firms (or subsidiaries) working specific units (e.g., helicopter or emergency physician vehicle) or providing services to the ambulance firms. We keep the ambulance firm employees and add a sector code (private or public) by using a 2-digit firm type code.

Unfortunately, we have difficulties capturing only the ambulance staff that are operational in the field, in relation to other staff, for example, senior managers. We take out physicians and managers by restricting the employees to having a wage income less than 80,000 EUR and an education less than 5 years in length. Since employees can have multiple employments, we restrict our data to each employee's main employer (based on the primary source of taxed wage income). We end up with, on average, 680 ambulance firm employees in each year. This number seems plausible given that the firms had 703 regular employees in total in 2017 (SLL, 2017).

With this data, we create several variables. *Turnover* is defined as either exiting the ambulance sector or changing firms. Exiting the ambulance sector is defined as being observed in an ambulance firm in year t but not being observed in an ambulance firm in any of the periods after t. Changing firms is defined as working in ambulance firm a in year t but not in a in the following time period. An employee has to be observed in at least one period following t, which makes a change of firm distinct from an exit.

To distinguish between all employees and specialist nurses (ambulance clinicians), we used educational codes with detailed information on the type of education (*Sun2000inr*). We used the code *723c*, which includes nurses with a specialist degree in anesthesiology, intensive care, surgery, and emergency/accident care. These are also the education specializations that ambulance employees had (especially pre-hospital care and anesthesiology). Although this is just a degree and does not inform us on the actual tasks an employee preforms, we found it superior to professional codes (*ssyk*). The professional codes changed during the sample period (in 2014), and we were not getting very good coverage by using the new professional codes. The old ones, on the other hand, included a more general population of nurses than what we needed (also nurses working in EDs).

Using the FAME database, we are able to add information on different skills. First, we use population wide conscription data (for males only) for cohorts born between 1950 and 1990. The conscription rates declined rapidly during the 1980s, which means that we are under-sampling cohorts born later during the 80s. Essentially, we can capture the cognitive skills of males that are somewhat older (in 2009, the youngest individuals with conscript data are 33 years old). From this data, we create a summary measure of cognitive skills by using four tests of logical, spatial, verbal, and technical skills. These tests are standardized by each cohort to a standard nine scale, and we take the average test result based on the four tests. Approximately 1850 of our observations are matched to the cognitive scores.

Moreover, we complement the direct cognitive measure with broader information on scholastic success by using educational results from the ninth grade. This information is available for the full population but only from 1988 and onward. This means that we are losing information on older ambulance employees when using this data. In 1988, ninth grade graduates were born in 1972, meaning that we have no information on cohorts born before that. We normalize the average grades (GPA) to mean zero and a standard deviation of one by each cohort. We do the same for the grade in math after appropriately re-coding the different grades depending on the grading scale used at the time. This data can be matched to both men and women, and we can match approximately 2300 observations in the data.

We use a measure of experience in the ambulance sector. To create this variable, we matched our target population to the yearly population-wide information on employment from 1990 and then saved information on the first occurrence of a job in any of the ambulance firms. This means that we are not capturing ambulance experience from other regions in Sweden or from other firms than those that constitute our main sample. Since we don't have specific information on if the firms were in the ambulance sector before 2007, we cannot generate this measure in a better way. This variable does not inform us on the number of years working in the ambulance sector but only of the number of years since first employed in one of the ambulance firms.

In Online Appendix, Table A7, we include 2 more outcome variables. The first is a dummy variable that indicates if an ambulance employee changed ambulance firms and at the same time switched from either the private to public sector or from the public to private sector. This variable will capture the net flows of employees from one sector to the other. We create this variable by setting the variable that describes if an employee changed firms between two years in the data to zero if the firm was private in year t and private in year $t+1$.

Finally, we create a variable denoting if an ambulance employee switched his or her workplace to a public hospital. To create this variable, we merged our ambulance population to the full population data between 2009-2017. We then saved all matches to our ambulance population with an industry code (SNI2007) of *86.102*, meaning "hospital somatic care." We then restricted the sample further by excluding employment in private firms and added the data to our data set on ambulance employees in Stockholm. We could then observe if an individual worked at an ambulance firm in year t but at a public hospital in $t+1$ (or the following available time period). A total of 3.4 % of the public ambulance employees started working in a public hospital in a given year.

Wage income is the tax filed wage income in 10 SEK (1 EUR) based on the variable *LoneInk*. Age is calculated using information on the year of birth and the year of observation.

References

Cunningham, S. (2021). *Causal inference: the mixtape*. Yale university press, New Haven and London.

Feigenberg, B. and Miller, C. (2021). Racial divisions and criminal justice: Evidence from southern state courts. *American Economic Journal: Economic Policy*, 13(2):207–40.

Heß, S. (2017). Randomization inference with stata: A guide and software. *Stata Journal*, 17(3).

HSF (2014). Årsrapport 2014: Prehospitala verksamheter i SLL. Technical report, Hälso och sjukvårdsförvaltningen, Stockholms läns landsting.

ICER (2018). Icer value framework overview of conceptual elements and procedures related to value assessment reports and appraisal committee voting at public meetings. Technical report, Institute for Clinical and Economic Review.

Mukherjee, A. (2020). Impacts of Private Prison Contracting on Inmate Time Served and Recidivism. *American Economic Journal: Economic Policy*. Forthcoming.

NBHW (2018). Nationella riktlinjer för hjärtsjukvård. Hälsoekonomiskt underlag. Bilaga. Technical report, National board of health and welfare.

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

SLL (2017). Årsprotokoll från uppföljning 2017 – vägburen ambulans. Hälso och sjukvårdsförvaltningen, Stockholms läns landsting.

Young, A. (2018). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*. *The Quarterly Journal of Economics*, 134(2):557–598.