Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns

Joseph Doyle John Graves Jonathan Gruber Samuel Kleiner *

August 8, 2012

Abstract

Endogenous patient sorting across hospitals can confound performance comparisons. This paper provides a new lens to compare hospital performance for emergency patients: plausibly exogenous variation in ambulance-company assignment among patients who live near one another. Using Medicare data from 2002-2008, we show that ambulance company assignment importantly affects hospital choice for patients in the same ZIP code. Using data for New York state from 2000-2006 that matches exact patient addresses to hospital discharge records, we show that patients who live very near each other but on either side of ambulance-dispatch boundaries go to different types of hospitals. Hospitals vary along a number of dimensions, and we begin by examining average cost as a summary measure of hospital resource usage. Both empirical strategies show that higher-cost hospitals have significantly lower one-year mortality rates compared to lower-cost hospitals, with a 10% increase in hospital costs associated with a 4% lower one-year mortality rate. In "unbundling" this finding, we find that other summary measures that describe the quality of hospitals inputs, such as their adherence to best practices, whether they adopt the latest technologies, and teaching hospitals, are all causally associated with lower mortality, but have little impact on the estimated mortality-hospital cost relationship. Rather, hospital procedure intensity is a key determinant of this relationship, suggesting that treatment intensity, and not differences in quality reflected in prices, drives much of our findings.

*Doyle: MIT Sloan School of Management 77 Massachusetts Ave, E62-515 Cambridge MA 02139, jjdoyle@mit.edu. Graves: Vanderbilt University School of Medicine 2525 West End Ave. Suite 600 Nashville, TN 37203-1738, john.graves@vanderbilt.edu. Gruber: MIT Department of Economics 50 Memorial Drive Building E52, Room 355 Cambridge MA 02142-1347, gruberj@mit.edu. Kleiner: Department of Policy Analysis and Management Cornell University 108 Martha Van Rensselaer Hall Ithaca, NY 14853, skleiner@cornell.edu Acknowledgements: We are grateful to Amy Finkelstein, Michael Greenstone, Lawrence Katz, Douglas Staiger, Heidi Williams and especially Jonathan Skinner for helpful conversations; seminar participants at the NBER, University of Chicago, MIT, Vanderbilt, New York University, the Brookings Institute, the University of Texas, the University of Pennsylvania, UCLA, USC, RAND Corporation, and the American Society of Health Economists for helpful comments, and to Noam Angrist, Archit Bhise, Shelly Jin, Rex Lam, and Jie Zhao for research assistance.We gratefully acknowledge support from the National Institutes of Health R01 AG41794-01.

Heterogeneity in hospital performance is a fundamental issue in health economics, and a primary concern of healthcare providers and policy makers. For example, hospitals vary enormously in their treatment intensity, yet it is not clear that higher costs are associated with better health outcomes. Meanwhile, policy makers are increasingly providing performance information to encourage quality improvement and reward higher-quality care rather than higher quantities of care (Fung et al. 2008, Dranove and Jin 2010, McClellan et al. 1994). A better understanding of what sets higher-performing hospitals apart from lower-performing ones can inform efforts to improve the quality of care and determine where efficiencies may be found.

A main problem when estimating performance differences is patient selection that can confound hospital comparisons. Patients choose or are referred to hospitals based on the hospital's capabilities: the highest-quality hospital in an area may treat the sickest patients. Alternatively, higher-educated or higher-income patients may be in better health and more likely to choose what is perceived to be a higher-quality hospital. Indeed, efforts to provide "report cards" for hospitals are often criticized for their inability to fully control for differences in patients across hospitals (Ryan et al. 2012).

This paper develops an empirical framework which allows us to compare hospital performance using plausibly exogenous variation in hospital assignment. The key ingredient of our approach is the recognition that the locus of treatment for emergency hospitalizations is, to a large extent, determined by pre-hospital factors – and, in particular, ambulance transport decisions and patient location. To the extent that ambulance companies are pseudo-randomly assigned to patients in an emergency, we can develop convincing measures of the impact of hospital differences on patient outcomes.

We consider two complementary identification strategies to exploit variation in ambulance transports. The first uses the fact that in areas served by multiple ambulance companies, the company dispatched to the patient is effectively random due to rotational assignment or even direct competition between simultaneously dispatched competitors. Moreover, we demonstrate that ambulance companies serving the same small geographic area have preferences as to which hospital they take patients. These facts suggest that the ambulance company dispatched to emergency patients may serve as a random assignment mechanism across local hospitals. We can then exploit ambulance identifiers provided in national Medicare data to develop instruments for hospital choice based on patient ambulance assignment. Finally, we can also use these ambulance data to test and control for any pre-hospital differences in treatment which might independently impact outcomes.

Our second strategy considers contiguous areas on opposite sides of ambulance service area boundaries in the state of New York. In New York, each state-certified Emergency Medical Service (EMS) provider is assigned to a territory via a certificate of need process where they are allowed to be "first due" for response. Other areas may be entered when that area's local provider is busy. We obtained the territories for each EMS provider from the New York State Department of Emergency Medical Services, and we couple these data with a unique hospital discharge dataset that identifies each patient's exact residential address. This combination allows us to compare those living on either side of an ambulance service area boundary. To the extent that these neighbors are similar to one another, the boundary can generate exogenous variation in the hospitals to which these patients are transported.

Hospitals vary along a host of dimensions that can affect outcomes. A natural starting point to characterize hospitals is their average cost. This provides a measure of the resources used by hospitals and begins to answer the question of whether high-cost hospitals may be worth their extra expense. Both of our empirical strategies yield a striking and consistent conclusion: higher-cost hospitals are associated with substantially improved patient mortality outcomes. The results from the ambulance assignment identification strategy imply that a 10% increase in hospital costs is associated with lower patient mortality in the first year after admission by roughly 4% of the baseline mortality rate.

Average costs are correlated with a wide variety of hospital-specific characteristics. In the last section of our paper, we endeavor to explore which hospital characteristics matter most for patient outcomes. Importantly, our empirical strategy provides a causal framework to assess the impact of a range of hospital characteristics, and to assess whether they explain the relationship we have documented between hospital cost and mortality. We study a variety of summary indicators of the quality of hospital inputs, including process-quality scores based on a hospital's adherence to best practices, and measures of "leading edge" hospitals, such as teaching hospitals and hospitals that quickly adopt the latest technologies. These measures are found to be causally associated with better patient outcomes. But none of these measures substantially impact the estimated mortality-hospital cost relationship, suggesting that even among hospitals of similar input quality, spending matters for outcomes. We do, however, find that hospital procedure intensity is a key determinant of the mortality-cost relationship, suggesting that treatment intensity, and not differences in quality, drives much of our findings.

Our paper proceeds as follows. Part I places our project in the context of the previous literature on measuring returns to hospital care and measuring hospital quality. Part II describes the nature of pre-hospital care and how it informs our approach. Part III discusses our data sources and Part IV our empirical strategy. Part V presents the basic results from Medicare data, and Part VI presents comparable results from New York state. Part VII extends the analysis to consider other attributes of the hospital "bundle." Part VIII concludes.

I Previous Literature

Our work is related to a number of cross-cutting literatures which speak to performance differences across hospitals.

I.1 Hospital Cost & Health Outcomes

There is a sizeable literature on spending and outcomes at the hospital level. This literature comes to mixed conclusions about the relationship between hospital spending and health outcomes (Joynt and Jha 2012). Several studies find significant returns to measures of hospital treatment intensity. Stukel et al. (2012) investigate variation in spending across hospitals in Ontario and find that higher spending due to costly interventions such as the use of specialists and more nursing care are associated with significantly lower mortality. Allison et al. (2000) find that those treated for Acute Myocardial Infarction (AMI) at teaching hospitals had roughly 10% lower mortality than non-teaching hospitals, and that this effect persisted for two years after the incident. Romley et al. (2011) document that those treated in the California hospitals with the highest end-of-life spending have much lower inpatient mortality: inpatient mortality in hospitals at the highest quintile of spending is 10-37% lower than at the lowest quintile across a range of conditions. Skinner and Staiger (2009) show that hospitals that were early adopters of "home run" technologies had modestly better outcomes when they accrued higher costs, although slower-adopters did not.

Other studies suggest no returns to higher spending. Glance et al. (2010) study Nationwide Inpatient Sample (NIS) data from 2006 and find that hospitals with low risk-adjusted inpatient mortality rates are associated with lower costs. Rothberg et al. (2010) use these NIS data from 2000-2004 and find that the change in hospitals' mortality rates and their growth in costs are uncorrelated. A middle ground is struck by Barnato et al. (2010), who find small positive returns to higher end-of-life spending in terms of lower mortality, but find that these effects fade quickly and are largely gone by 180 days after admission. These results suggest little long term benefit to higher spending.

Studies of regions within the US show large disparities in spending that are not associated with improvements in health outcomes (Fisher et al. 1994, Pilote et al. 1995, Kessler and McClellan 1996, Tu et al. 1997, O'Connor et al. 1999, Baicker and Chandra 2004, Fuchs 2004, Stukel et al. 2005, Sirovich et al. 2006). Fisher et al. (2003) studied Medicare expenditure data and found that end-of-life spending levels are 60% higher in high-spending areas compared to low-spending ones in the U.S. Nevertheless, no difference is found across regions in 5-year mortality rates following a health event such as a heart attack or hip fracture. This wide variation in spending and similarity of mortality rates were again found when the sample was restricted to teaching hospitals (Fisher et al. 2004). The lack of a relationship between regional variation in spending and health outcomes has been cited in support of reducing Medicare spending by 20-30% without adversely affecting health outcomes (Fisher et al. 2009).

I.2 Existing Quality Measures

A number of initiatives score hospitals on their use of best practices (Chassin et al. 2010, The Joint Commission 2011, Agency for Health Care Research and Quality 2012). For example, these measures include the use of beta blockers at the time of arrival for acute myocardial infarction and the timing of antibiotic administration for pneumonia patients. These process measures have been shown to have little or a negative relationship with spending both across and within markets (Yasaitis et al. 2009), but have been criticized because they tend to have little relationship with risk-adjusted mortality (Werner and Bradlow 2006, Fung et al. 2008).

Another way to characterize hospitals is to compare mortality rates, controlling for observable patient characteristics. Some states publish "report cards" of hospital and physician performance in this way. The report cards hold the potential to inform patients and referring physicians, but have been criticized for unintended consequences (Dranove et al. 2003, Werner and Asch 2005). For example, there are concerns that risk-adjustment fails to adequately address the concern that patients differ across hospitals. One way this manifests itself is that providers may be less likely to provide intensive (and riskier) treatment for patients whose risk is greater than the risk-adjustment would suggest.

I.3 Inference Problem: Patient Selection

A major issue that arises when comparing hospitals is that they may treat different types of patients. For example, greater treatment levels may be chosen for populations in worse health. At the individual level, higher spending is strongly associated with higher mortality rates, even after risk adjustment, which is consistent with more care provided to patients in (unobservably) worse health. At the hospital level, long-term investments in capital and labor may reflect the underlying health of the population as well. Differences in unobservable characteristics may therefore bias the results toward finding no effect of greater spending.

Research on area- or hospital-level variation in costs recognizes the issue of patient selection. To address this concern, the studies tend to focus on diagnoses where patients are likely to present with similar severity levels (e.g. heart attacks). They note that observable patient characteristics are similar across areas.[1] For example, Fisher et al. (2003) found similar predicted mortality rates across areas that varied in their spending levels for hip fracture patients, although somewhat higher predicted mortality in high-spending areas for heart attack patients.

[1] See, for example, O'Connor et al. (1999); Pilote et al. (1995); Stukel et al. (2005); and Fisher et al. (2003)

Further, these studies endeavor to control for patient mix with a variety of indicators of patient severity. But even the best controls based on diagnosis codes and patient characteristics are only imperfect proxies for underlying severity. Advanced risk adjustment techniques explain less than 10% of the year-to-year variation in patient spending in the Medicare program (Garber et al. 1998). While some fraction of the unexplained variation is exogenous and therefore unpredictable, it is likely that patient decisions to seek medical treatment are driven by health factors unobservable to the researcher.

A recent article by Zhang et al. (2010), for example, finds that the unadjusted correlation between pharmaceutical spending and medical (non-drug) spending across high- and low-spending Medicare regions is high (0.6), but that this finding is highly sensitive to patient controls; the correlation falls to just 0.1 when patient health status is taken into account. In addition, Doyle (2011) compared patients in Florida and again found that observable characteristics were similar across areas that had significant variation in hospital spending levels. When the analysis focused on tourists in similar destinations – a group of patients that is arguably more comparable across areas and is unlikely to affect the chosen level of treatment intensity in the area – higher-spending areas were associated with substantially lower mortality.

The use of claims-based diagnoses to control for underlying health may also be problematic because the diagnosis measures themselves could be endogenous. That is, a patient listed with many diagnoses could be in poor health or could have been treated by a provider that tends to diagnose (and record) more illnesses. For example, Song et al. (2010) find that Medicare patients who move to higher intensity regions experience a greater increase in the number of diagnoses over time compared to similar patients in the area from which they moved. In another recent piece, Welch et al. (2011) find an inverse relationship between regional diagnostic frequency rates and case fatality rates, suggesting that the marginal patient diagnosed in a high diagnosis-frequency (and high observation intensity) area may be less sick compared to patients diagnosed with the same condition in low-frequency areas. To control for underlying health differences, another direct measure is the patient's lagged healthcare spending. Yet this too may be problematic when the goal is to describe healthcare systems as high vs. low intensity, as intensity is autocorrelated.

Clearly, with the limitations of standard risk adjustment methods in mind, it is even more critical to develop a methodology that cleanly separates provider assignment from patient health.

One previous source of variation used in health economics is differential distance to the hospital as an exogenous instrument for determining hospital assignment. McClellan et al. (1994) and Cutler (2007) show that patients who live closer to (and are treated by) hospitals that perform cardiac catheterization, relative to hospitals that do not, have improved survival rates. They note that the mechanism for this improvement is likely due to "correlated beneficial care": superior care that is not due to the invasive procedures themselves. Geweke et al. (2003) used differential distance to study pneumonia patients in the Los Angeles area and found that large and small hospitals had better outcomes than medium-sized ones, and, related to the comparisons here, they found suggestive evidence that teaching hospitals had better outcomes than non-teaching hospitals. Chandra and Staiger (2007) employ a Roy model where physicians specialize in more intensive treatments over medical treatments if there are relatively high returns to doing so, and productivity spillovers further enhance the returns to intensive treatment. In this model, the spillover results in potentially worse outcomes for patients who would benefit most from the less-intensive treatment because the region has specialized in the intensive treatment to raise average outcomes. As a result, restricting the intensive hospitals to practice in the less-intensive style would result in worse health outcomes, despite the potential for little difference in outcomes across areas in the cross section. Using differential distance to estimate treatment effects, their empirical results support the model's predictions.

While differential distance has proved useful, it also faces some key limitations. First, patients who live relatively close to "high tech" hospitals could be different than those who do not in ways that are difficult to control. For example, wealthier and healthier areas may demand the latest treatments, and hospitals may locate near certain types of patients. Indeed, Hadley and Cunningham (2004) find that safety net hospitals locate near the poorest patients. Additionally, hospitals may endogenously adopt technologies if they believe their patient population will benefit, and their patient population is primarily composed of those who live close to the hospital. Third, exact distances are difficult to measure in most datasets, with researchers relying on distance from each patient's ZIP code centroid to each hospital. This can affect the precision of the estimates. The current paper presents a new source of variation that is orthogonal to variation based on distance: patients who live near one another but are treated at different hospitals.

II Background on Pre-Hospital Care

The key ingredient of our approach is the recognition that the locus of treatment for emergency hospitalizations is, to a large extent, determined by pre-hospital factors, including ambulance transport decisions and patient location. Among the emergency cases we consider, 61% are brought into the hospital via ambulance. In such cases, the level of care dispatched to the scene (e.g. Advanced Life Support (ALS) using paramedics vs. Basic Life Support (BLS) using Emergency Medical Technicians) may be chosen based on perceived severity (Curka et al. 1993, Athey and Stern 2002). Critically, however, in areas served by multiple ambulance companies, the company dispatched is usually chosen independent of the patient characteristics that can confound the area and hospital comparisons reviewed above.

Rotational assignment of competing ambulances services – as well as direct competition between simultaneously dispatched competitors – is increasingly common in the U.S. For example, two recent articles cite examples from North and South Carolina where the opportunity for ambulance transport is broadcast to multiple companies and whichever arrives there first gets the business.[2] Similarly, large cities such as New York, Los Angeles and Chicago have adopted a hybrid approach under which private ambulance companies work in conjunction with fire departments to provide Emergency Medical Services (EMS) (Johnson 2001). Another report found that of the top 10 cities with the highest population over age 65, 5 contracted with both public and private ambulance carriers, while 2 others contracted exclusively with private carriers (Chiang et al. 2006). In a more recent 2010 survey covering 97 areas, 40 percent reported contracting with private ambulance companies and an additional 23 percent utilized hospital-based ambulance providers (Ragone 2012).

We are aware of no systematic evidence on the basis for rotational assignment of ambulances. To understand the dispatch process, we conducted a survey of 30 cities with more than one am-

[2] See Watson (2011) and Johnson (2001) for examples.

bulance company serving the area in our Medicare data. The survey revealed that patients can be transported by different companies for two main reasons. First, in communities served by multiple ambulance services, 911 systems often use software that assigns units based on a rotational dispatch mechanism; alternatively, they may position ambulances throughout an area and dispatch whichever ambulance is closest, then reshuffle the other available units to respond to the next call. Second, in areas with a single ambulance company, neighboring companies provide service when the principal ambulance units are busy under so-called "mutual aid" agreements. Critically, under either dispatch mechanism, every one of the respondents surveyed reported that ambulance assignment was independent of patient characteristics (within BLS vs. ALS ambulances). Within a small area, then, the variation in the ambulance dispatched is either due to rotational assignment or one of the ambulance companies being engaged on another 911 call. Both sources appear plausibly exogenous with respect to the underlying health of a given patient.

There is some existing evidence that pre-hospital care is an important determinant of hospital choice due to the "preferences" of ambulance companies to take patients to particular hospitals. In the South Carolina example, the article explicitly points out that if an ambulance company associated with a particular hospital gets to the patient first, the patient is much more likely to be transported to that hospital.

Directly relevant to our approach is research by the New York State Comptroller's Office in the wake of a major change in the rotational assignment of private and FDNY ambulances in New York City. Skura (2001) found that patients living in the same ZIP code as public Health and Hospital Corporation (HHC) hospitals were less than half as likely to be taken there when assigned a private, non-profit ambulance (29%) compared to when the dispatch system assigned them to an FDNY ambulance (64%). In most cases, the private ambulances were operated by non-profit hospitals and stationed near or even within those facilities, so they tended to take their patients to their affiliated hospitals.

This point is illustrated in Figure 1, from Skura (2001). This figure shows the location of three hospitals, two of them private hospitals that operate ambulance service (St. Clare's and New York Hospital) and one public (Bellevue hospital). The author examined the rate at which ambulances took patients residing in the Bellevue ZIP code to these hospitals. He found that for those picked up by FDNY ambulances, 61% were brought to Bellevue, and 39% brought to the more distant private hospitals. But for those picked up by private ambulance companies, only 25% were brought to Bellevue, and 75% to the other hospitals (Figure 1). Similar results were found for other ZIP codes within New York City as well.

In summary, ambulance dispatch rules appear to effectively randomize patients to ambulance companies. Previous case studies suggest that these ambulances have preferences about which hospital to choose. Our empirical strategy exploits this plausibly exogenous variation in the hospital choice, as described below.

III Data

Our national data are Medicare claims between 2002 and 2008. The use of these data was previously authorized under a data use agreement with the Centers for Medicare and Medicaid Services (CMS). In particular, the Carrier file includes a 20% random sample of beneficiaries, and from this file we observe the ambulance claim. We then link these claims to inpatient claims, which include standard measures of treatment, such as procedures performed, and up to 10 diagnosis codes. Patient characteristics are also recorded, such as age, race, and sex. The claims data also include the ZIP code of the beneficiary, where official correspondence is sent. In principle, this could differ from the patient's home ZIP code. In addition, vital statistics data that record when a patient dies is linked to these claims, which allows us to measure mortality at different timeframes, such as 30 days or one year.

In addition to these usual controls in claims data, we discovered that the ambulance claims offer a new set of control variables. These data include detailed information on the mode and method of transport (Advanced Life Support vs. Basic Life Support; emergency vs. non-emergency[3]) and on specific pre-hospital interventions administered by ambulance personnel (e.g., intravenous therapy and administered drugs). While previous studies using Medicare data have been limited by their

[3]While we study conditions and admissions that appear to be emergencies, ambulances are reimbursed at a higher rate if they transport the patient under so-called "emergency traffic" (i.e.,"lights and sirens") on the way to the hospital.

inability to control for patient location (beyond using distance from the centroid of the patient's ZIP code), an innovation of our approach is that we also control for "loaded miles": a billing term referring to the exact distance the ambulance traveled to the hospital with the patient on board. Finally, our Medicare claims also include an ambulance company identifier.[4] This allows us to construct empirical referral pattern measures that serve as the basis for our analytic strategy.

The actual costs associated with the treatment of a particular hospitalization are not recorded in hospital inpatient data, and there are several different proxies available for costs. The previous literature has typically used Medicare hospital expenditures, but such expenditures are almost completely driven by diagnosis coding and hospital cost add-ons rather than treatment differences within a diagnosis. An alternative is to measure costs by the facility charges for each hospitalization. This measure overstates costs, however, as hospitals typically mark up list prices well beyond the prices actually paid for their services. Another alternative is to renormalize this charge measure by a hospital-specific cost to charge ratio measured each year, and this is the basis of our cost measure in the main results: average hospital log costs.[5] We report results when alternative measures are used, as well.

Our other major data source is the universe of inpatient hospital discharges from New York State, made available from the New York State Department of Health through the Statewide Planning and Research Cooperative System (SPARCS). These data include detailed information on patient demographic characteristics, diagnoses, and treatments, as well as a unique patient identifier that allows for longitudinal linking across facilities. A unique feature of these data is an address field that allows us to identify the exact patient residence location for 90% of the discharge records in our 2000-2006 sample (approx. 20.6 million records for all patients). These data are matched to vital statistics databases from the entire state of New York, enabling the construction of mortality measures extending up to 1-year.

The SPARCS data complement our analysis in three ways. First, because the SPARCS data

[4]Medicare only reimburses for ambulance transports to a nearby facility; patients who wish to be taken somewhere else must pay the incremental cost of transport to that facility. There are a small number of such episodes in our data, and the results are not sensitive to their inclusion.

[5]A disadvantage of these cost-to-charge ratios is that they can be noisy measures. Centers for Medicare and Medicaid (CMS) recommends replacing the cost-to-charge ratio for outliers with the median for that year. We replaced the extreme 5% of the cost-to-charge ratios in this way.

include a residential address for each hospital patient, we can use narrowly defined geographic areas such as census block groups for our analysis. These smaller areas are likely to be even more homogeneous than the larger ZIP code areas. Second, these addresses allow us to match patients to narrowly defined areas located near the ambulance-dispatch boundaries. Third, the New York data allow us to compare patients throughout the age distribution and study patients not insured under Medicare.

Sample Construction

In both data sets, our primary sample consists of patients admitted to the hospital through the emergency room with 29 "nondeferrable" conditions where selection into the healthcare system is largely unavoidable. Discretionary admissions see a marked decline on the weekend, but particularly serious emergencies do not. Following Dobkin (2003) and Card et al. (2009), diagnoses whose weekend admission rates are closest to 2/7ths reflect a lack of discretion as to the timing of the hospital admission. Using our Medicare sample, we chose a cutoff of all conditions with a weekend admission rate that was as close or closer to 2/7ths as hip fracture, a condition commonly thought to require immediate care. These conditions represent 39% of the hospital admissions via the emergency room, 61% of which arrived by ambulance. In New York, these nondeferrable conditions account for 34% of all emergency patients whose primary payer is Medicare.

For our analysis of the Medicare data, we are unable to consider beneficiaries who are part of Medicare Advantage programs, as their claims are not available. These beneficiaries constitute 17% of the Medicare population in 2000 and 22% in 2008 (Kaiser Family Foundation 2010). We further limit the sample to patients during their first hospitalization under the Medicare program, in order to study outcomes after an initial health shock, and in an effort to exclude patients who may have preferences for particular hospitals due to previous hospitalizations. By necessity of the empirical strategy, we limit the analysis to those patients who are brought to the emergency room by ambulance. We further remove a small number of observations with missing ZIP code information, missing ambulance company information, as well as ambulance companies, ZIP codes, or hospitals with fewer than 10 observations. Last, we restrict the sample to hospitals that are within 50 miles of the patient's ZIP code centroid. This results in a sample of 667,143 patients.

The reliance on ambulance transports allows us to focus on patients who are less likely to decide whether or not to go to the hospital. Appendix Table A1 shows that this sample has more comorbidities, is slightly older, and has a higher 1-year mortality rate (37%) compared to all Medicare patients who enter the hospital via the emergency room (20%). These are relatively severe health shocks, and the estimates of the effects of hospital types on mortality apply to these types of episodes. We caution against applying the results to more chronic conditions.

For our analysis of the New York SPARCS data, pre-hospital care is not collected so we cannot identify ambulance transports. To facilitate comparison of results using these data to the Medicare sample, we restrict the analysis to patients who enter inpatient care via the emergency room and have a principal diagnosis considered nondeferrable. Our main results will also focus on the first hospitalization in our data, as well as a restriction to Medicare patients. We also remove a small number of observations with missing address information, patients whose residence is located outside of New York State, and patients whose address could not be matched to a block group. We again restrict the sample to patients receiving care at a hospital located within 50 miles of their residential address. This results in a sample of 142,809 patients within 1 mile of a boundary, 213,968 patients within two miles of a boundary, and 281,036 patients within 5 miles of a boundary.

IV Empirical Strategy

Ambulance Referral Patterns within ZIP Code Areas

Our first approach relies on differences in ambulance referral patterns within ZIP code areas. Ambulance companies have some discretion over hospital choice, with a typical tradeoff between distance and the hospital with the most appropriate level of care. We compare patients picked up by ambulances with different tendencies to favor particular types of hospitals (such as high-cost hospitals) in these decisions. We then assess whether these different preferences lead to meaningful differences in the type of hospital where patient is treated.

We can illustrate that such "preferences" exist by essentially generalizing the New York City example above using variation in hospital shares across ambulance companies serving the same ZIP codes. Specifically, using observed ambulance-hospital frequencies within each ZIP code in our Medicare sample, we estimate a Chi-square test of homogeneity. Consider, for example, a ZIP code served by two hospitals in which we observe emergency patients taken to hospital h_1 75% of the time when they are picked up by ambulance company a_1, but only 33% of the time when they are picked up by company a_2. Since there are only two hospitals, it follows that we would observe 25% of a1's patients, and 66% of a_2's patients, being taken to hospital h_2. Given these observed proportions, we can test whether there is statistical evidence that companies a_1 and a_2 have different patient transport patterns.

In our sample, we calculated test statistics for every ZIP code in our Medicare data with at least 5 ambulance transports by comparing observed ambulance-hospital cell frequencies to those expected under the null hypothesis, which is that ambulances distribute patients across nearby hospitals at the same rates.[6] Among the 9,125 ZIP codes where we can calculate these statistics, 38% have test statistics with $p < 0.1$. This provides evidence that there appear to be differences in where patients are taken based on which ambulance company picks them up that well exceeds pure chance (which would result in less than 10% of zips having test statistics with $p < 0.1$). This type of variation is the basis of our first-stage estimation, which we turn to next.

To operationalize ambulance preferences, we calculate an instrumental variable that measures the treatment intensity of hospitals where each ambulance company takes its patients. For patient i assigned to ambulance $a(i)$, we calculate the average inpatient costs among the patients in our analysis sample for each ambulance company:[7]

$$Z_{a(i)} = \frac{1}{N_{a(i)} - 1} \sum_{j \neq i}^{N_{a(i)}-1} Cost_j$$

[6]The resulting test-statistic (for ZIP z) is distributed χ^2 with $(H_z - 1) * (A_z - 1)$ degrees of freedom, where H_z is the total number of hospitals treating patients from ZIP z, and A_z is the total number of ambulance companies transporting patients from ZIP z.

[7]In practice, some ambulance companies serve large areas including multiple states. To compare patients at risk of receiving the same ambulance company, we compute the instrument at the company-Hospital Referral Region level. Hospital Referral Regions are relatively large areas designed to capture markets for non-emergency care. This allows us to retain information about the ambulance company's preferences across hospitals within and outside the patient's (smaller) Hospital Service Area.

This measure is essentially the ambulance company fixed effect in a model of hospital costs. We exclude the given patient from this measure to avoid a direct linkage between Z and the average cost in a given hospital – a Jackknife Instrumental Variables Estimator (JIVE) that is more robust to weak-instrument concerns when fixed effects are used to construct an instrument (Stock et al. 2002, Kolesr et al. 2011).

We then use this measure to estimate the first-stage relationship between average hospital costs, H, and the instrument, Z: hospital costs associated with the ambulance assigned to patient i living in ZIP code $z(i)$ in year $t(i)$:

$$H_i = \alpha_0 + \alpha_1 Z_{a(i)} + \alpha_2 X_i + \alpha_3 A_i + \theta_{z(i)} + \lambda_{t(i)} + \nu_i \tag{1}$$

where X_i is a vector of patient controls including indicators for each age, race, sex, miles from the ZIP code centroid, and indicators for four common comorbidities: congestive heart failure, chronic obstructive pulmonary disease (COPD), diabetes, and other comorbidity; A_i represents a vector of ambulance characteristics including the payment to the company, which provides a useful summary of the treatment provided in the ambulance; indicators for distance traveled in miles; whether the transport utilized Advanced Life Support (e.g., paramedic) capabilities; whether intravenous therapy was administered; whether the transport was coded as emergency transport; and whether the ambulance was paid through the outpatient system rather than the carrier system.[8]. We cluster standard errors at the Hospital Service Area (HSA) level, as each local market may have its own assignment rules. This choice is relatively conservative compared to clustering at the ambulance company level instead.

We also include a full set of ZIP code fixed effects and year indicators. This regression therefore compares individuals who live in the same ZIP code in the same year, but who are picked up by ambulance companies with different "preferences" across different types of hospitals (excluding the patient herself). A positive coefficient α_1 would indicate that ambulance company "preferences" are correlated with where the patient actually is admitted.

[8]Claims for ambulances owned by institutional providers (e.g., hospitals) are found in the outpatient file, and represent about 10% of all ambulance transports within our file. These data do not include the distance measure, and for these observations the distance indicators were filled with the sample mean.

Relationship Between Hospital Costs and Mortality

Our main regression of interest is the relationship between hospital costs on mortality, M, for patient i:

$$M_i = \beta_0 + \beta_1 H_i + \beta_2 X_i + \beta_3 A_i + \theta_{z(i)} + \lambda_{t(i)} + \epsilon_i \tag{2}$$

This OLS regression parallels the previous literature in modeling mortality of patient i who goes to hospital h, as a function of average hospital cost. Mortality can be measured at intervals such as 30 days, 90 days, or 1 year. As noted earlier, this regression suffers from the fact that patients may be selected into certain hospitals based on characteristics which affect their mortality. To address this, we estimate the model by instrumental variables, where the instrument is the ambulance measure discussed above. That is, we use equation (1) above as a first stage to estimate this model by instrumental variables.

This empirical approach has four main limitations. The first is that ambulance company preferences could be correlated with underlying patient characteristics even within ZIP codes. For example, some ambulance companies could be expert at avoiding complicated cases that are likely to die. Our survey evidence gives us confidence that this is not the case. In addition, the results below investigate whether our results are sensitive to controls for observed patient characteristics.

Second, some ambulance companies may serve only a certain part of a ZIP code, and these ambulance companies may disproportionately take their patients to particular hospitals. For example, an ambulance company that serves a higher-income part of a ZIP code could be more likely to take patients to high-cost hospitals. In that case, we might find that high-cost hospitals are associated with better outcomes through patient selection. We can address this concern to some extent in our specification checks by restricting our sample to particularly homogenous ZIP codes using the Summary File 3 (SF3) file issued by the U.S. Census. By restricting our analysis to ZIP codes with little within-ZIP code variation in demographic characteristics such as household income and racial composition, we hope to minimize the potential for ambulance selection within a ZIP code.

A third concern is that the approach interprets differences in costs and outcomes as stemming from different hospital assignment patterns across ambulance companies, but ambulance companies may have a more direct impact on health. If ambulance companies that tend to take patients to high-cost hospitals provide different levels of pre-hospital care, then the outcome differences would be due to a combination of the two types of treatment. More subtly, part of the variation comes from ambulance companies driving farther to reach a different hospital. If these distances were systematically longer as ambulances travel farther to reach a high-cost hospital, then "high-cost" ambulances may take more time to reach the hospital, resulting in worse outcomes. An innovation in this project is that we study (and control for) differences in care provided by the ambulance company, including the distance traveled to the hospital. We also investigate the timing of any mortality reductions, as the direct effects of treatment delivered by the ambulance company will likely be observed soon after admission.

A final concern is that ambulance company assignment may only affect hospital choice for a subset of patients, and the results would provide a local average treatment effect. For example, we are unable to estimate effects for patients who always insist (and are taken to) high-cost hospitals. Because of these limitations we turn to a second, complementary approach.

Borders Approach

Our alternative approach compares patients along borders that define distinct ambulance dispatch areas. The idea is that patients could live in the same neighborhood yet go to very different hospitals because they reside on opposite sides of a shared border. This parallels the analysis of Black (1999), who compared those living on either side of school district borders to study the impact of school quality on housing prices. For this analysis, we focus on New York state, where we have data on exact patient addresses coupled with a detailed disptch grid we obtained from the New York State Department of Emergency Medical Services.

Each state-certified Emergency Medical Service (EMS) provider in New York is assigned to a territory where they are allowed to be "first due" for response via a certificate of need process, subject to the terms of New York Public Health Law (Article 30). These territories are typically delineated using county, city, town, village and fire district boundaries. Other areas may be entered when the provider is requested for mutual aid.

Using these data, we can identify census block groups in New York state on either side of an ambulance dispatch area boundary. Census block groups are the smallest geographical units defined by the U.S. Census Bureau for which demographic information is publicly available. These block groups have an average population of 1,300 residents. Using the latitude and longitude coordinates of each patient's residential address as recorded in our hospital discharge data, we map each patient to a unique census block group. We then identify individuals whose block group centroid is located within a defined distance of an ambulance service territory border.

Specifically, we include patients residing in block groups located within 1-mile, 2-miles, and 5-miles of an ambulance border. The smaller distance criteria allows us to compare patients who live very near to one another and are likely a better matched comparison. The 5-mile criterion allows us to retain more rural areas, however, as block groups are constructed based on population counts and the centroid in these areas may lie outside the 1- or 2-mile restrictions.

The estimating equations parallel the earlier analysis, but now the instrument is constructed across areas rather than ambulance companies. Rather than ZIP code fixed effects, we include matched-pair fixed effects that allow us to compare patients who live on either side of the same boundary. For patient i living in dispatch area $a(i)$, the first-stage model takes the form:

$$H_i = \phi_0 + \phi_1 Z_{a(i)} + \phi_2 X_i + \theta_{p(i)} + \lambda_{t(i)} + \omega_i \tag{3}$$

where H_i represents the average costs in the hospital where the patient is treated, while $Z_{a(i)}$ is the average hospital cost for patients living in the dispatch area where the patient resides, and $\theta_{p(i)}$ is a set of dummies for each matched pair of census block groups. So this regression asks: are patients who live in a dispatch area serviced by relatively high-cost hospitals more likely to be treated at high-cost hospitals themselves compared to those who live nearby but across a dispatch area boundary? Standard errors in these models are clustered at the dispatch-area level.

We can then once again estimate a mortality-cost model of the form:

$$M_i = \gamma_0 + \gamma_1 H_i + \gamma_2 X_i + \theta_{p(i)} + \lambda_{t(i)} + \epsilon_i \tag{4}$$

where we instrument for hospital costs using the first stage relationship in (3).

The borders approach augments the ambulance preference approach because differences in hospital patterns within these small areas are plausibly due to differences in ambulance dispatch patterns and not patient tastes. At the same time, there may be other factors that change at borders that could bias our findings. For example, in New York a common border for dispatch areas is the county, and counties may differ in other factors that impact the choice of residence, such as the quality of public services.[9] Our analysis will control for differences in resident characteristics at the boundary using U.S. Census SF3 data. Of course, differences in unobserved characteristics of patients across a border from one another remain a concern.

In summary, we consider two different identification strategies using two different data sets. Each has advantages and weaknesses, but taken together they can provide insights into whether different types of hospitals achieve better outcomes.

V Ambulance Company Preference Results

Balance

The key underlying assumption of our approaches is that the sources of variation in the hospital type have been purged of patient-specific factors which impact costs or outcomes. To assess whether this is true at least along observable dimensions, Table 1 shows the balance of patient characteristics across those whose ambulances tend to transport patients to relatively high-cost or low-cost hospitals available to a ZIP code area.

The first row of the table shows that costs among those picked up by ambulance companies with preferences for higher-cost hospitals is about 20 log points higher compared to those picked up by ambulance companies with preferences for lower-cost hospitals. The next set of rows shows that these two groups of patients are similar in terms of their overall health and demographic characteristics. For example, the average age for low-cost ambulance companies is 78.5 vs. 78.0

[9] Of the borders delineated by the New York EMS service file, borders that separate adjacent cities and adjacent towns are most common (29.2%, and 18.7% respectively), while 15.5% of the borders divide counties and towns, 13.2% of the borders divide counties and fire districts, and 6.5% divide counties and villages. Other types of borders (e.g. town-village, city-county, etc.) each comprise less than 5% of the border sample.

for high-cost companies. Sex, race, and comorbidity rates are similar as well. Perhaps even more comforting is that the two groups of patients look nearly identical in terms of their pre-hospital treatment, as summarized by the ambulance payment ($291 vs. $296). They administer an IV in 9.7% of cases in both groups. The high-cost ambulances are more likely to be Advanced Life Support ambulances, and we will report estimates of a robustness check that limits the sample to ALS ambulances. In addition, the high-cost areas had a 3% shorter distance, on average. While time to the hospital matters for survival, and we control for distance, a 0.2 mile difference is unlikely to drive the results. That said, we do consider 1-day mortality as an alternative way to investigate effects of treatment by the ambulance company. At least in terms of observable characteristics, our sample appears well balanced.

Basic Results

Table 2 shows the first-stage results for our ambulance company preference instrument, equation (1) above. We begin by estimating the relationship between average hospital cost at the patient's hospital and the average hospital cost associated with the ambulance company assigned to the patient, controlling only for year and ZIP code fixed effects. There is a very strong correlation between the two, suggesting that if the ambulance company that tends to take other patients to 10% more expensive hospitals, the hospital where the patient is taken has 3.1% higher average hospital costs, and this difference is highly statistically significant. The subsequent columns add controls for patient and ambulance characteristics. The result is remarkably robust to these additional controls.

The source of the variation is evident from the coefficient being significantly less than one. Consider the variation that stems from mutual-aid arrangements, for example. The instrument is calculated with heavier weights on areas where the company usually operates. If it is called in to help when another company is busy, the first-stage reflects that the company is more likely to transport the patient back to its usual hospitals, but less so in the mutual-aid area than where it usually operates. The mutual aid area likely has other nearby hospitals in the choice set.

Table 3 shows the results of estimating equation (2) by OLS. We find a strong and highly significant negative correlation between hospital cost and mortality. The results suggest that raising costs by 10% (or about $1000) would lower one year mortality by 0.24 percentage points (or about 0.7% of baseline mortality). As noted earlier, this result suffers from potential selection bias due to patient choice of hospital. It is worth noting, however, that the results are relatively robust to controls, suggesting little bias in the unconditional mortality difference due to observable characteristics; we find quite different results in our OLS estimates for New York, however.

The next set of rows report the 2SLS estimates and the point estimates are much larger in magnitude. In a model with year, diagnosis and ZIP code fixed effects, a 10% rise in costs is associated with a 1.96 percentage-point lower mortality rate, or about 6% of baseline mortality. In a model with full controls, the estimate declines to 1.44 percentage points. While the estimates do decline with controls, the estimates are not statistically significantly different from one another.

To put the estimate in context, this result implies that a two standard deviation increase in average hospital costs, an increase of 0.8 log points, is associated with an 11.5 percentage point reduction in mortality, or 31% of the sample mortality rate. Thus, we find compelling evidence that higher-cost hospitals have significantly lower patient mortality, at least for emergency admissions.

Robustness

We next subject this striking finding to a series of robustness tests. First, we report results across alternative mortality horizons up to two years. We are particularly interested in very short run effects at one day after admission. If we are finding strong effects of differing pre-hospital care, then those should show up right away and perhaps fade over time. More generally, the analysis allows us to examine when the mortality difference appears and whether it extends beyond one year.

The results of using different horizons are presented in Table 4. The OLS results in the top rows show mortality reductions which are significant at every duration, and the magnitude grows with duration along with the sample mean. In contrast, the 2SLS results show significant increases in mortality at one day, which become reductions over time and are only significant by 30 days. These findings are inconsistent with the notion that the main results simply reflect differences in pre-hospital care. To the converse, it appears that the hospitals to which individuals are (effectively) randomly assigned appear to undertake actions which raise mortality in the short run, but reduce it in the longer run. Indeed, when we classify procedures by their 1-day mortality rate, we find expensive hospitals are more likely to perform procedures with high 1-day mortality rates.[10] Moreover, the longer-term mortality reductions do not fade out but continue to have important mortality impacts even at 2 years, although part of the longer-run effects could be due to ongoing care in the subsequent period rather than the effects of the initial hospital stay.

Table 5 reports robustness checks of our findings along several other dimensions. The first panel shows sizeable effects across different age groups. The next set examines the impact of ZIP code heterogeneity on our findings. When we consider the standard deviation of income in the ZIP code, the results are largest in the 2nd and 3rd quartile of this measure. Meanwhile, the effects are remarkably stable across ZIP codes that vary according to racial concentration. This is inconsistent with the notion that ZIP code heterogeneity is driving our results through patient/ambulance co-location within a ZIP code. Last, the smallest geographic unit in our data is the ZIP code, and we find a negative cost-mortality relationship across ZIP codes of different sizes: the largest effects are found in the 2nd and 3rd quartiles of this measure. Again, the effects are not monotonically related to ZIP code size in a way that would suggest systematic confounding variation.

Appendix Table A2 reports further robustness checks. Results were similar when we restricted the sample to: ambulance transports labeled "emergency," patients who were transported from home or a nursing home, patients transported in ALS ambulances, and patients who were not transferred to another inpatient facility. Results were also similar when we used hospital charges rather than those charges deflated by a cost-to-charge ratio.

Another robustness check defines the hospital cost categories in terms of sample quartiles to allow for non-(log)linear effects, with instruments for each quartile calculated at the ambulance company level. The results show that the top three quartiles have statistically significantly lower mortality compared to the bottom quartile. The point estimates also show that the point estimates increase in magnitude for the 2nd and 3rd quartile (-0.033 and -0.054), while the top-quartile point estimate is slightly smaller (-0.045), suggesting diminishing returns. [11] Table A2 also shows that

[10] To characterize procedures, the 1-day mortality rate was calculated for all 3-digit principal procedures with at least 30 observations, a mean of 5.5/1000. Using our IV strategy, a two-standard deviation increase in the hospital cost measure was associated with a 1.99/1000 increase in this principal-procedure mortality rate ($p < 0.001$).

[11] In further exploration, when the hospitals are placed in 7 categories rather than 4, the point estimate increases

the results are similar when we vary the instrument calculation to exclude the patient's ZIP code rather than the patient. Overall, the result is remarkably consistent to all of these changes.

Monotonicity Assumption Specification Check

One issue that arises with this type of instrument is that the monotonicity assumption to interpret the results in a LATE framework need not be satisfied. In this context, ambulance companies could be more likely to take certain types of patients to high cost hospitals, but less likely to take other types of patients to those hospitals. In speaking with EMS technicians, this did not seem to be the case for serious emergencies considered here. Further, this is an insured population, so there are fewer concerns with regard to "dumping" uninsured patients on other hospitals.

To further investigate this issue, we calculated the instrument for each ambulance-company by disease-category cell rather than at the ambulance- company level. This estimation allows ambulances to direct patients differentially by type of illness but retain the LATE interpretation.[12] Table A2 shows that the results are similar when we allow the instrument to vary at the disease-type level.

VI Border Results for New York State

Balance

In this section, we turn to the second empirical strategy, relying on comparisons of individuals living in close proximity but on either side of ambulance dispatch area borders in New York State. Once again, we begin by showing that these samples are similar in terms of observed characteristics. In Table 6 we divide each pair of census block groups in New York into those on the low-cost side of an ambulance border, and those on the high-cost side. We find about an 11 log point difference in hospital cost, on average, between these two groups. The remaining rows again show relatively well-balanced compositions of the groups in terms of our control variables. The largest difference is

in magnitude to -0.064 (s.e. 0.032) at the second-highest cost category, with the coefficient for the top category lower at -0.044 (s.e. 0.055).

[12] We used the 17 broad categories that make up the ICD-9-CM classifications.

the distance from the patient's home census block to the hospital, with a 0.25 mile shorter distance for patients on the high-cost side. This difference fades away, however, and is not seen in the sample of patients within 5 miles of a boundary.

Since we are comparing different areas, and individuals may differ if they choose to live on one side of the boundary the other, we can also consider differences in characteristics of the block groups in which they reside. In the final rows of Table 6, we find that block groups on adjacent sides of these borders are similar in terms of characteristics such as income, share owner-occupied housing, and share urban.

Basic Results

Patients are more likely to attend a hospital in their district. Specifically, for patients with a hospital located in their district: 72% of patients living within one mile of a district border are treated at a hospital in their district, and this rate increases to 75% and 78% for those living within 2 and 5 miles of a border, respectively. This is borne out in the first-stage relationship reported in Table 7. The point estimates range between 0.64 and 0.74, with F-statistics that range from 12.8 to 18.1.

Table 7 also reports the OLS and 2SLS results. One salient difference between the New York results and those for the national sample is that in our New York State sample we find that the OLS results for mortality are much more sensitive to controls. These findings indicate substantial selection bias in OLS when the models only control for year indicators: the coefficient on hospital costs becomes much more negative as we add controls.

As with the national sample, however, the estimates increase in magnitude when we use a 2SLS approach. Using the border strategy, we find that a 10% increase in costs is associated with lower mortality ranging from 0.47 to 0.54 percentage points, or about 2.0-2.3% of baseline mortality rate, depending on the sample.[13] This result implies a two standard deviation increase in average hospital costs is associated with a 4.1-4.5 percentage point reduction in mortality, or 17-19% of

[13] The mortality rate in this sample is lower than in the Medicare analysis, as this sample is composed of all emergency room admissions rather than ambulance transports.

the baseline mortality rate.[14] These results reinforce our findings from the ambulance company preference strategy, as we again find compelling evidence that higher-cost hospitals have lower mortality for emergency admissions.

Extensions

We undertake two extensions with our New York sample. First, in Table A3, we show the impact of varying mortality horizons. Consistent with the OLS results from the ambulance preference strategy, the OLS results in the first panel show mortality reductions that grow with duration for all three distance samples. In contrast, point estimates for the 2SLS results show no change in mortality at one day for patients. Over time, however, we find that for all of our distance samples, reductions in mortality are significant for the 30-day mortality measure. Although the New York data do not include information that allows us to control for pre-hospital factors (as was done with the Medicare data), this pattern of results is again inconsistent with the notion that our results are driven by differences in pre-hospital care which would likely be found most strongly at the shortest mortality horizons.

Second, one important extension that is available in New York is our ability to look at the impacts of hospital costs on the non-elderly. To do so, we replicate our existing strategy for the non-elderly sample, ages 18-64, and the "near elderly," ages 50-64 (see Appendix Tables A4 and A5). For both samples, we restrict analysis to the same non-deferrable conditions analyzed in the Medicare sample. We find similar first-stage estimates in these samples, and the OLS results are again sensitive to controls. In contrast to the results for the elderly, we find no statistically significant impacts on mortality using 2SLS. As with the elderly, however, point estimates for the 2SLS results are roughly twice as large in magnitude relative to the OLS estimates. In addition, the point estimates imply a substantial mortality-cost relationship. For both groups, a two standard-deviation increase in hospital costs is associated with a 15-30% reduction in mortality compared to the sample mean mortality rate, magnitudes which are within the range of the New York Medicare

[14] The standard deviation of mean hospital log(costs) is equal to 0.415, 0.432 and 0.447 for the one-mile, two-mile and five-mile samples, respectively.

sample.[15]

VII Interpretation

"Unbundling" the Cost Results

Our analysis thus far has considered hospital costs as a summary measure of hospital characteristics that might drive patient outcomes. But hospital costs may proxy for a "bundle" of hospital characteristics that are correlated with costs. As we move from carefully documenting that hospitals matter to understanding why they matter, it is important to begin to unbundle these findings.

Of course, any other hospital characteristic that we relate to outcomes is potentially endogenous due to patient selection, as is the case with average hospital cost. But we can use the same Medicare ambulance preference strategy used for costs to instrument for these measures as well; for example, we can effectively compare patients who are picked up by ambulances that prefer teaching hospitals to ones that prefer non-teaching hospitals. In this way, we can provide selection-free estimates of the impacts of these alternative quality measures. And, by including the instrumented quality measures in the same regression as our instrumented cost regression, we can effectively run a "horse race" to see if controlling for other quality measures drives out the cost result. Unfortunately, given the more limited variation available in the New York data, we are unable to replicate this analysis on that sample as well.

In theory, there is a very large set of hospital characteristics that we could relate to outcomes. We choose to focus on three sets of measures for this analysis. The first set of measures aims to capture "leading edge" hospitals which should be at the forefront of patient treatment and therefore might be both higher cost and have higher quality physical and human capital. The second measure is an independent measure of quality based on the rate at which hospitals provide best-practice care. The third measure focuses on a quantity measure of intensity rather than costs, which combines variation in price and quantity.

[15] For the nonelderly, a 2 standard deviation increase in costs (std. dev=0.42) is associated with a 0.9 to 1.6 percentage point reduction in mortality, or 15-30% of the sample mortality of 5.4%. Meanwhile, a two-standard deviation increase in costs for the near elderly (std. dev=0.43) is associated with a 1 to 1.5 percentage point reduction in mortality, or 18-28% of the sample mortality of 8.2%.

We use two indicators to characterize hospitals as "leading edge." The first is teaching hospital status. Our main measure includes hospitals that are recognized as teaching hospitals by the American Medical Association, although results are similar when we use alternative definitions.[16] 41% of the patients in our sample are treated in a teaching hospital. Second, we define "high-tech" hospitals using a technology-frontier index we constructed using hospital survey data from the American Hospital Association (AHA). Values for this index were defined by identifying new questions in each year's AHA hospital survey between 1995 and 2008. For example, beginning in 2004 the AHA began asking whether hospitals had the capability to deliver shaped beam radiation therapy, a technology that precisely targets radiation at tumors. The addition of such a question would be considered a "new" technology in our measure. For our 2002-2008 data we allowed a "burn-in" period of 7 years from 1995 to 2001 to construct a base set of new technologies, and then beginning in 2002 ranked hospitals based on the number new technologies they had in each year. We identified "high-tech" hospitals as those in the top decile of this measure in a given year, and hospital rankings are similar when we used alternative measures, such as a rolling five-year average of the ranking. This measure is somewhat more selective than the teaching hospital definition, with 30% of patients treated in these high-tech hospitals. Table A6 shows that both teaching and high-tech hospitals are positively correlated with average hospital costs, with teaching hospitals associated with hospital log costs that are 8 points higher than non-teaching hospitals, and high-tech hospitals associated with 12 log point higher costs.

Another hospital characteristic we consider is a process-oriented composite measure of whether best practices are employed for heart failure, heart attack, and pneumonia patients. This measure is constructed from publicly reported CMS data on the Hospital Compare website. We summarize three sets of process scores for these conditions using a denominator-weighted composite index, which has been shown to be highly correlated with "shrinkage" scores based on Bayesian hierarchical modeling (Shwartz et al. 2008). The composite quality score for 2007 is used to characterize the hospitals, and it is standardized to have mean zero and standard deviation of one. [17] Procedural

[16] Appendix Table A2 reports results for the more selective definition of hospitals that are members of the Council of Teaching Hospitals, which represents 30% of our sample admissions.

[17] This measure is available for a wide set of hospitals beginning in 2005. Results were similar when the 2005 measure is used instead.

quality measures have also been shown to be negatively correlated with treatment intensity at the area level (Baicker and Chandra 2004) and shown to be positively correlated with survival gains for heart attack patients during the 1980s and early 1990s (Skinner et al. 2006). In addition, Skinner and Staiger (2009) show that hospitals that were early adopters of these best practices appear to have positive returns to spending. Table A6 shows that the denominator weighted composite score is associated with higher hospital costs in our sample of Medicare patients.

The results for different hospital types are shown in Table 8. The top panel shows the OLS results, first including each measure separately, and then including each alternative measure along with hospital costs in the same model. We find that each of the alternative measures is significantly negatively associated with mortality. Teaching, high-tech hospitals, and those with a one-standard-deviation higher composite quality rating have conditional mortality rates that are 1 percentage point lower than other hospitals. When regressed together with hospital average costs, the coefficient estimates for teaching and high-tech hospitals decline slightly but are qualitatively similar. This slight decline is seen for hospital costs as well: controlling for these alternative measures of quality does not dampen the estimated relationship between hospital cost and mortality.

The second panel shows the same exercise using 2SLS to instrument for each of our quality measures, as well as hospital costs. The first stages for these models have larger coefficients compared to hospital costs and are highly statistically significant (see Appendix Table A7). For teaching hospitals and high-tech hospitals, we find that instrumenting significantly increases the point estimates, suggesting that the OLS results are indeed biased downwards (in absolute value) by patient selection. We find that going to a teaching hospital (due to ambulance preferences) lowers mortality by 3.9 percentage points, which is over 10% of baseline mortality. The effect is even larger for being brought to a high-tech hospital: 4.7 percentage points. For hospitals' procedural quality score, the associated impact is also quite large: a two standard deviation increase in this score is associated with a 5.4 percentage-point reduction in mortality.

The second set of results relate to the horse races. We once again find that the effects of hospital costs on patient outcomes are not reduced by including these other quality measures in the regression either one at a time or all at once. The point estimates for each of the other measures falls by more than half when hospital cost is included in the model, suggesting that part of the effect of being treated at those hospitals is that they treat patients more intensively. Higher-cost hospitals appear to improve outcomes above and beyond these measures of the quality of the hospital inputs.

Finally, we consider the decomposition of our cost measure into its price and quantity components. As a proxy for quantity, we compute for each patient the number of procedures performed during his stay. We then measure the average number of procedures at the hospital (excluding the patient himself). This averages 1.33 with a standard deviation of 0.58.

Table 9 shows that more procedures are negatively correlated with 1-year mortality in the OLS regressions, but the 2SLS coefficient is significantly larger: -0.161.[18] This implies that a two-standard deviation increase in this treatment-intensity measure is associated with an 18 percentage point reduction in mortality. Next, we attempt to decompose the cost result by including average cost and average number of procedures in the same model. The coefficient on the quantity measure declines only slightly, but the coefficient on average cost decreases by more than half to -0.065 compared to -0.144 in the main results. These results suggest that much of the difference in mortality at high-cost hospitals stems from larger quantities of care provided. Cost remains significant, suggesting a remaining effect of spending above and beyond our proxy for quantity, but quantities are substantively more important.

VIII Conclusions and Implications

Given the high and rising costs of health care in the U.S., with the attendant pressures on public and private budgets, it is imperative that policy makers develop strategies for "bending the cost curve" and lowering the ultimate share of GDP which will be consumed by health care. A wide variety of potential solutions have been posed to the cost control problem, but none holds more promise than the notion of saving money by ending over-reimbursement for inefficient treatments. The results of the Dartmouth analysis, and parallel findings for hospitals, have been used to suggest that a sizeable share of hospital spending does little to improve patient outcomes, and so could

[18] See Table A7 for the first-stage relationship, which continues to be economically and statistically significant for the procedure measure.

readily be cut without harming patient health. Our results do not disprove that conclusion - but they should give policy makers some pause before assuming that spending can be easily cut without harming patient health, at least in the context of emergency care.

In particular, using exogenous ambulance assignment to control for the endogeneity of patient hospital choice, our results suggest that high-cost hospitals have better outcomes: 20-30% lower mortality compared to low-cost hospitals. Other summary measures of hospital quality again suggest that hospital choice matters: teaching hospitals and hospitals that are early adopters of new technologies have lower mortality when similar patients are compared. Interestingly, even controlling for teaching status or capital quality, high-cost hospitals have lower mortality. We rule out many alternative explanations for this finding, and show that it is robust to two different identification strategies.

As with previous studies of this relationship, it is not possible to conclude that the treatment intensity-mortality relationship reflects the effects of higher-cost inputs. Other aspects of the hospital, such as physician or staff quality measures beyond those considered here, may be correlated with overall cost and may achieve similar results with smaller budgets. To address that question, a plausibly exogenous within-hospital source of spending variation would need to be found. The fact that our spending results persist when we condition on other summary measures of hospital quality suggests that treatment intensity itself may be driving the improved outcomes. It would be particularly useful to extend this analysis to the non-emergent population, to assess whether spending is productive for that group as well.

References

Agency for Health Care Research and Quality, "National Healthcare Quality Report, 2011," Technical Report 12-0005, AHRQ, Rockville, MD 2012.

Allison, Jeroan J., Catarina I. Kiefe, Norman W. Weissman, Sharina D. Person, Matthew Rousculp, John G. Canto, Sejong Bae, O. Dale Williams, Robert Farmer, and Robert M. Centor, "Relationship of Hospital Teaching Status With Quality of Care and Mortality for Medicare Patients With Acute MI," *JAMA: The Journal of the American Medical Association*, 2000, *284* (10), 1256 –1262.

Athey, Susan and Scott Stern, "The Impact of Information Technology on Emergency Health Care Outcomes," *RAND Journal of Economics*, 2002, *33* (3), 399–432.

Baicker, Katherine and Amitabh Chandra, "Medicare spending, the physician workforce, and beneficiaries' quality of care," *Health Affairs (Project Hope)*, June 2004, *Suppl Web Exclusives*, W4–184–97. PMID: 15451981.

Barnato, Amber E, Chung-Chou H Chang, Max H Farrell, Judith R Lave, Mark S Roberts, and Derek C Angus, "Is survival better at hospitals with higher "end-of-life" treatment intensity?," *Medical Care*, February 2010, *48* (2), 125–132. PMID: 20057328.

Black, Sandra E, "Do Better Schools Matter? Parental Valuation of Elementary Education," *The Quarterly Journal of Economics*, May 1999, *114* (2), 577–599.

Card, David, Carlos Dobkin, and Nicole Maestas, "Does Medicare Save Lives?," *The Quarterly Journal of Economics*, May 2009, *124* (2), 597–636.

Chandra, Amitabh and Douglas O. Staiger, "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks," *Journal of Political Economy*, 2007, *115*, 103–140.

Chassin, Mark R., Jerod M. Loeb, Stephen P. Schmaltz, and Robert M. Wachter, "Accountability Measures Using Measurement to Promote Quality Improvement," *New England Journal of Medicine*, August 2010, *363* (7), 683–688.

Chiang, Arthur, Guy David, and Michael Housman, "The Determinants of Urban Emergency Medical Services Privatization," *Critical Planning*, 2006, *Summer.*

Curka, Peter A., Paul E. Pepe, Victoria F. Ginger, Robert C. Sherrad, Michael V. Ivy, and Brian S. Zachariah, "Emergency medical services priority dispatch," *Annals of Emergency Medicine*, 1993, *22* (11), 1688–1695.

Cutler, David M., "The lifetime costs and benefits of medical technology," *Journal of Health Economics*, December 2007, *26* (6), 1081–1100.

Dobkin, Carlos, "Hospital Staffing and Inpatient Mortality," *Unpublished Working Paper*, 2003.

Doyle, Joseph J, "Returns to Local-Area Health Care Spending: Evidence from Health Shocks to Patients Far From Home," *American Economic Journal: Applied Economics*, July 2011, *3* (3), 221–243.

Dranove, David and Ginger Zhe Jin, "Quality Disclosure and Certification: Theory and Practice," Working Paper 15644, National Bureau of Economic Research January 2010.

——, **Daniel Kessler, Mark McClellan, and Mark Satterthwaite**, "Is More Information Better? The Effects of Report Cards on Health Care Providers," *Journal of Political Economy*, June 2003, *111* (3), 555–588. ArticleType: research-article / Full publication date: June 2003 / Copyright 2003 The University of Chicago Press.

Fisher, E S, J E Wennberg, T A Stukel, and S M Sharp, "Hospital readmission rates for cohorts of Medicare beneficiaries in Boston and New Haven," *The New England Journal of Medicine*, October 1994, *331* (15), 989–995. PMID: 8084356.

Fisher, Elliott S, David E Wennberg, Thrse A Stukel, and Daniel J Gottlieb, "Variations in the longitudinal efficiency of academic medical centers," *Health Affairs (Project Hope)*, 2004, *Suppl Variation*, VAR19–32. PMID: 15471777.

Fisher, Elliott S., David E. Wennberg, Thrse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and toile L. Pinder, "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care," *Annals of Internal Medicine*, February 2003, *138* (4), 288 –298.

Fisher, Elliott S, Julie P Bynum, and Jonathan S Skinner, "Slowing the growth of health care costs–lessons from regional variation," *The New England Journal of Medicine*, February 2009, *360* (9), 849–852. PMID: 19246356.

Fuchs, Victor R., "Perspective: More Variation In Use Of Care, More Flat-Of-The-Curve Medicine," *Health Affairs*, October 2004.

Fung, Constance H, Yee-Wei Lim, Soeren Mattke, Cheryl Damberg, and Paul G Shekelle, "Systematic review: the evidence that publishing patient care performance data improves quality of care," *Annals of internal medicine*, January 2008, *148* (2), 111–123. PMID: 18195336.

Garber, Alan M., Thomas E. MaCurdy, and Mark B. McClallan, "Persistence of Medicare Expenditures among Elderly Beneficiaries," *Forum for Health Economics & Policy*, 1998, *1.*

Geweke, John, Gautam Gowrisankaran, and Robert J. Town, "Bayesian Inference for Hospital Quality in a Selection Model," *Econometrica*, July 2003, *71* (4), 1215–1238.

Glance, Laurent G, Andrew W Dick, Turner M Osler, Wayne Meredith, and Dana B Mukamel, "The association between cost and quality in trauma: is greater spending associated with higher-quality care?," *Annals of Surgery*, August 2010, *252* (2), 217–222. PMID: 20647927.

Hadley, Jack and Peter Cunningham, "Availability of Safety Net Providers and Access to Care of Uninsured Persons," *Health Services Research*, 2004, *39* (5), 15271546.

Johnson, Robin, "The Future of Local Emergency Medical Service: Ambulance Wars or Public-Private Truce?," *Reason Foundation*, 2001.

Joynt, Karen E. and Ashish K. Jha, "The Relationship Between Cost and Quality: No Free Lunch," *JAMA: The Journal of the American Medical Association*, March 2012, *307* (10), 1082.

Kaiser Family Foundation, "Kaiser Medicare Chartbook, Fourth Edition (Figure 4.3)," http://facts.kff.org/chart.aspx?cb=58&sctn=165&ch=1757 2010.

Kessler, Daniel and Mark McClellan, "Do Doctors Practice Defensive Medicine?," *The Quarterly Journal of Economics*, May 1996, *111* (2), 353 –390.

Kolesr, Michal, Raj Chetty, John N. Friedman, Edward L. Glaeser, and Guido W. Imbens, "Identification and Inference with Many Invalid Instruments," *National Bureau of Economic Research Working Paper Series*, 2011, *No. 17519.*

McClellan, M, B J McNeil, and J P Newhouse, "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables," *JAMA: The Journal of the American Medical Association*, September 1994, *272* (11), 859–866. PMID: 8078163.

O'Connor, Gerald T., Hebe B. Quinton, Neal D. Traven, Lawrence D. Ramunno, T. Andrew Dodds, Thomas A. Marciniak, and John E. Wennberg, "Geographic Variation in the Treatment of Acute Myocardial Infarction," *JAMA: The Journal of the American Medical Association*, February 1999, *281* (7), 627 –633.

Pilote, L, R M Califf, S Sapp, D P Miller, D B Mark, W D Weaver, J M Gore, P W Armstrong, E M Ohman, and E J Topol, "Regional variation across the United States in the management of acute myocardial infarction.," *The New England Journal of Medicine*, August 1995, *333* (9), 565–572. PMID: 7623907.

Ragone, Michael, "Evolution or Revolution: EMS Industry Faces Difficult Changes," *Journal of Emergency Medical Services*, 2012, (February).

Romley, John A., Anupam B. Jena, and Dana P. Goldman, "Hospital Spending and Inpatient Mortality: Evidence From California," *Annals of Internal Medicine*, February 2011, *154* (3), 160 –167.

Rothberg, Michael B., Joshua Cohen, Peter Lindenauer, Judith Maselli, and Andy Auerbach, "Little Evidence Of Correlation Between Growth In Health Care Spending And Reduced Mortality," *Health Affairs*, August 2010, *29* (8), 1523–1531.

Ryan, Andrew, James Burgess, Robert Strawderman, and Justin Dimick, "What Is the Best Way to Estimate Hospital Quality Outcomes? A Simulation Approach," *Health Services Research*, 2012, *47* (4), 16991718.

Shwartz, Michael, Justin Ren, Erol A Pekz, Xin Wang, Alan B Cohen, and Joseph D Restuccia, "Estimating a composite measure of hospital quality from the Hospital Compare database: differences when using a Bayesian hierarchical latent variable model versus denominator-based weights," *Medical care*, August 2008, *46* (8), 778–785. PMID: 18665057.

Sirovich, Brenda E., Daniel J. Gottlieb, H. Gilbert Welch, and Elliott S. Fisher, "Regional Variations in Health Care Intensity and Physician Perceptions of Quality of Care," *Annals of Internal Medicine*, May 2006, *144* (9), 641 –649.

Skinner, Jonathan and Douglas Staiger, "Technology Diffusion and Productivity Growth in Health Care," *National Bureau of Economic Research Working Paper Series*, 2009, *No. 14865.*

Skinner, Jonathan S, Douglas O Staiger, and Elliott S Fisher, "Is Technological Change In Medicine Always Worth It? The Case Of Acute Myocardial Infarction," *Health Affairs*, March 2006, *25* (2), w34–w47.

Skura, Barry, "Where Do 911 System Ambulances Take Their Patients? Differences Between Voluntary Hospital Ambulances and Fire Department Ambulances," 2001.

Song, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E Wennberg, and Elliott S Fisher, "Regional variations in diagnostic practices," *The New England Journal of Medicine*, July 2010, *363* (1), 45–53. PMID: 20463332.

Stock, James H, Jonathan H Wright, and Motohiro Yogo, "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, October 2002, *20* (4), 518–529.

Stukel, Therese A, Elliott S Fisher, David A Alter, Astrid Guttmann, Dennis T Ko, Kinwah Fung, Walter P Wodchis, Nancy N Baxter, Craig C Earle, and Douglas S Lee, "Association of hospital spending intensity with mortality and readmission rates in Ontario hospitals," *JAMA: Journal of the American Medical Association*, March 2012, *307* (10), 1037–1045. PMID: 22416099.

Stukel, Therese A., F. Lee Lucas, and David E. Wennberg, "Long-term Outcomes of Regional Variations in Intensity of Invasive vs Medical Management of Medicare Patients With Acute Myocardial Infarction," *JAMA: The Journal of the American Medical Association*, March 2005, *293* (11), 1329 –1337.

The Joint Commission, "Improving America's Hospitals: The Joint Commission's Annual Report on Quality and Safety," Technical Report 2011.

Tu, J V, C L Pashos, C D Naylor, E Chen, S L Normand, J P Newhouse, and B J McNeil, "Use of cardiac procedures and outcomes in elderly patients with myocardial infarction in the United States and Canada," *The New England Journal of Medicine*, May 1997, *336* (21), 1500–1505. PMID: 9154770.

Watson, Stuart, "Two ambulance companies compete for patients and their health dollars | WCNC.com Charlotte," http://www.wcnc.com/home/Two-Ambulance-companies-compete-for-patients-and-their-health-dollars–99140244.html 2011.

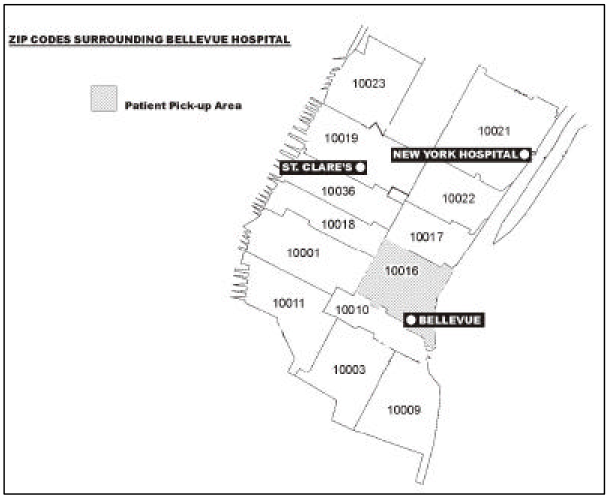
Welch, H. Gilbert, Sandra M Sharp, Dan J Gottlieb, Jonathan S Skinner, and John E Wennberg, "Geographic Variation in Diagnosis Frequency and Risk of Death Among Medicare Beneficiaries," *JAMA: The Journal of the American Medical Association*, March 2011, *305* (11), 1113–1118.

Werner, R. M. and David A. Asch, "The Unintended Consequences of Publicly Reporting Quality Information," *JAMA: The Journal of the American Medical Association*, March 2005, *293* (10), 1239–1244.

——— **and E. T. Bradlow**, "Relationship Between Medicare's Hospital Compare Performance Measures and Mortality Rates," *JAMA: The Journal of the American Medical Association*, December 2006, *296* (22), 2694–2702.

Yasaitis, Laura, Elliott S. Fisher, Jonathan S. Skinner, and Amitabh Chandra, "Hospital Quality And Intensity Of Spending: Is There An Association?," *Health Affairs*, July 2009, *28* (4), w566–w572.

Zhang, Yuting, Katherine Baicker, and Joseph P Newhouse, "Geographic variation in Medicare drug spending," *The New England Journal of Medicine*, July 2010, *363* (5), 405–409. PMID: 20538621.



Destination of Patients Picked Up In The Bellevue Hospital Zip Code Area

Destination	All Voluntary Hospital Ambulances	Fire Department Ambulances
Bellevue Hospital (HHC)	25%*	61%**
Any Voluntary Hospital	75%	39%

*157 taken to Bellevue/632 total. **815 taken to Bellevue/1,346 total

Figure 1. New York City Ambulance Referral Patterns: Skura (2001)

Source: Skura, Barry. "Where do 911 System Ambulances Take Their Patients? Differences Between Voluntary Hospital Ambulances and Fire Department Ambulances." City of New York, Office of the Comptroller, 2001.

Table 1—Ambulance Strategy: Patient Characteristics Across Ambulance Companies

	≤Median	>Median
Mean Hospital log(Cost)	8.593	8.795
Age	78.5	78.0
Age < 65	0.091	0.098
Age ≥ 65 & <70	0.086	0.091
Age ≥ 70 & <75	0.120	0.125
Age ≥ 75 & <80	0.173	0.174
Age ≥ 80 & <90	0.391	0.381
Age ≥ 90	0.139	0.131
Male	0.399	0.401
Race: white	0.870	0.852
Race: African American	0.089	0.104
Comorbidity: Congestive Heart Failure	0.230	0.219
Comorbidity: COPD	0.254	0.250
Comorbidity: Diabetes	0.207	0.210
Comorbidity: Other	0.337	0.343
Ambulance Payment	291	296
Ambulance Distance	7.4	7.2
Advanced Life Support	0.673	0.721
Ambulance: IV Administered	0.097	0.097
Ambulance: Outpatient Reimbursement	0.096	0.105
Ambulance: Emergency Transport	0.819	0.843
Observations	667,143	

Note: Some ambulance measures have smaller sample sizes, largely because they are not recorded in the in the outpatient reimbursement system: Ambulance measures for Advanced Life Support and IV administration (N=600,005); Ambulance distance travelled (N=599,825)
Source: 2002-2008 Medicare Part A Claims Data

Table 2—Ambulance Strategy: First Stage

	(1)	(2)	(3)	(4)
Ambulance: Mean Hospital log(Cost)	0.311	0.309	0.303	0.300
	(0.012)**	(0.012)**	(0.012)**	(0.012)**
Diagnosis fixed effects	No	Yes	Yes	Yes
Patient Controls	No	No	Yes	Yes
Ambulance Controls	No	No	No	Yes
Observations	667,143			
Mean of Dep. Var	8.640			

Note: Estimates reported for Equation (1) in the text. All models include ZIP code and year fixed effects. Patient controls include indicators for year of age, race, sex, miles from the ZIP code centroid, and comorbidities. Ambulance controls are listed in Table 1. Standard errors in parentheses, clustered at the HSA level. * significant at 5%; ** significant at 1%
Source: 2002-2008 Medicare Part A Claims Data

Table 3—Ambulance Strategy: One Year Mortality & Hospital Cost

	(1)	(2)	(3)	(4)
A. OLS				
Mean Hospital log(Cost)	-0.032	-0.036	-0.024	-0.024
	(0.0042)**	(0.0038)**	(0.0036)**	(0.0035)**
B. 2SLS				
Mean Hospital log(Cost)	-0.229	-0.196	-0.151	-0.144
	(0.036)**	(0.028)**	(0.024)**	(0.024)**
Diagnosis fixed effects	No	Yes	Yes	Yes
Patient Controls	No	No	Yes	Yes
Ambulance Controls	No	No	No	Yes
Observations	667,143			
Mean of Dep. Var	0.366			

Note: Estimates reported for Equation (2) in the text. All models include ZIP code and year fixed effects. Patient controls include indicators for year of age, race, sex, miles from the ZIP code centroid, and comorbidities. Ambulance controls are listed in Table 1. Standard errors in parentheses, clustered at the HSA level. * significant at 5%; ** significant at 1%
Source: 2002-2008 Medicare Part A Claims Data

Table 4—Ambulance Strategy: Mortality Horizons

	1-day	7-day	30-day	1-year	2-year
A. OLS					
Mean Hospital log(Cost)	-0.002	-0.006	-0.014	-0.024	-0.030
	(0.00097)*	(0.0018)**	(0.002)	(0.0035)**	(0.0039)**
B. 2SLS					
Mean Hospital log(Cost)	0.020	0.006	-0.068	-0.144	-0.173
	(0.0053)**	(0.010)	(0.014)**	(0.024)**	(0.028)**
Observations	760,998	760,998	754,039	667,143	560,929
Mean of Dep. Var.	0.026	0.082	0.168	0.366	0.470

Note: Each cell represents a separate model analogous to those estimated in Table 3. All models include full controls. Outcomes are uncensored, as the estimating sample is restricted to patients observed at least 7-days, 30-days, 1-year, or 2-years from the end of the sample period. Standard errors in parentheses, clustered at the HSA level. * significant at 5%; ** significant at 1%
Source: 2002-2008 Medicare Part A Claims Data

Table 5—Ambulance Strategy: 2SLS Results for Subgroups

Subgroup	Coefficient	S.E.	Obs.	1-year Mortality
A. Age				
< 65	-0.117	(0.043)**	63,329	0.223
65 – 74	-0.102	(0.039)**	141,187	0.295
75 – 84	-0.152	(0.037)**	253,950	0.355
85+	-0.178	(0.038)**	208,677	0.472
B. ZIP Code Characteristics				
Income: Standard Deviation				
Bottom Quartile	-0.118	(0.033)**	162,177	0.368
2nd	-0.193	(0.042)**	162,117	0.365
3rd	-0.161	(0.041)**	162,217	0.366
Top Quartile	-0.111	(0.038)**	162,077	0.365
Race HHI				
Bottom Quartile	-0.145	(0.053)**	162,203	0.369
2nd	-0.140	(0.035)**	162,162	0.371
3rd	-0.142	(0.037)**	162,096	0.365
Top Quartile	-0.154	(0.034)	162,127	0.360
ZIP Code Area (sq miles)				
Bottom Quartile	-0.095	(0.052)	162,175	0.362
2nd	-0.215	(0.051)**	162,141	0.366
3rd	-0.175	(0.039)**	162,159	0.366
Top Quartile	-0.113	(0.0.32)**	162,113	0.372

Note: Each cell represents a separate model. All models include full controls. Standard errors in parentheses, clustered at the HSA level. ZIP code characteristic cells are for ZIP codes with available 2000 US Census data. * significant at 5%; ** significant at 1%
Source: 2002-2008 Medicare Part A Claims Data

Table 6—Border Strategy: Patient Characteristics Across Ambulance Service Area Borders

Sample:	1 Mile		2 miles		5 miles	
	Low	High	Low	High	Low	High
Mean Area log(Cost)	8.658	8.763	8.657	8.757	8.675	8.752
Age	78.6	78.5	78.5	78.4	78.6	78.4
Age < 65	0.060	0.064	0.062	0.066	0.062	0.067
Age ≥ 65 & < 70	0.116	0.117	0.118	0.120	0.117	0.120
Age ≥ 70 & < 75	0.144	0.144	0.145	0.145	0.142	0.144
Age ≥ 75 & < 80	0.186	0.189	0.187	0.188	0.185	0.187
Age ≥ 80 & < 85	0.372	0.364	0.370	0.359	0.369	0.359
Age ≥ 85	0.122	0.122	0.120	0.122	0.125	0.124
Share African American	0.063	0.108	0.069	0.116	0.081	0.119
Share Asian	0.008	0.021	0.011	0.018	0.021	0.018
Share Hispanic	0.028	0.027	0.030	0.028	0.030	0.025
Share Other Race	0.028	0.040	0.029	0.041	0.025	0.037
Share Nat. Amer	0.001	0.001	0.001	0.001	0.001	0.001
Share Male	0.383	0.385	0.385	0.384	0.383	0.382
Distance travelled	3.977	3.727	4.126	3.915	3.958	4.001
Charlson Score	0.911	0.915	0.915	0.921	0.915	0.926
Median Income	33,293	32,092	32,061	30,949	30,680	29,672
Mean Income	65,633	66,722	64,421	65,074	60,878	61,273
Share Owner Occ. Housing	0.866	0.861	0.864	0.871	0.856	0.879
Share Urban	0.971	0.973	0.941	0.940	0.930	0.926
Number of Cross-border Pairs	336		482		583	

Note: Areas represented are distances from US Census Block Group centroids to an ambulance dispatch-area boundary.

Source: SPARCS Data

Table 7—Border Strategy: First Stage, OLS and 2SLS

Sample:	1 Mile		2 Miles		5 Miles	
	(1)	(2)	(3)	(4)	(5)	(6)
A. First Stage: Mean Area log(Cost)						
Mean Area log(Cost)	0.715	0.738	0.671	0.678	0.627	0.632
	(0.207)**	(0.206)**	(0.181)**	(0.177)**	(0.153)**	(0.148)**
Mean of Dep. Var.	8.70	8.70	8.70	8.70	8.72	8.72
B. OLS: 1-Year Mortality						
Mean Area log(Cost)	0.009	-0.015	0.014	-0.012	0.015	-0.016
	(0.010)	(0.008)*	(0.010)	(0.007)	(0.008)	(0.007)*
C. 2SLS: 1-Year Mortality						
Mean Area log(Cost)	-0.046	-0.054	-0.038	-0.047	-0.040	-0.047
	(0.031)	(0.023)*	(0.028)	(0.024)*	(0.024)	(0.02)*
Year Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes	No	Yes
Diagnosis Controls	No	Yes	No	Yes	No	Yes
Observations	142,809	142,809	213,968	213,968	281,036	281,036
Mean of Dep. Var.	0.236	0.236	0.236	0.236	0.236	0.236

Note: Panel A reports estimates from Equation (3) in the text, and Panel B reports estimates from Equation (4). All models include boundary and year fixed effects. Patient controls include indicators for year of age, race, sex, miles from the block-group centroid, comorbidities, and US Census block-group characteristics listed in Table 6. Standard errors in parentheses, clustered at the HSA level. * significant at 5%; ** significant at 1%
Source: SPARCS Data

Table 8—Ambulance Strategy: Alternative Hospital Summary Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. OLS							
Teaching: AMA	-0.012			-0.010			-0.008
	(0.0020)**			(0.0020)**			(0.0021)**
Technology: Top Decile		-0.011			-0.009		-0.005
		(0.0018)**			(0.0019)**		(0.0019)**
CMS Quality			-0.012			-0.010	-0.009
			(0.0020)**			(0.0020)**	(0.0020)**
Mean Hospital log(Cost)				-0.022	-0.021	-0.021	-0.018
				(0.0035)**	(0.0036)**	(0.0037)**	(0.0037)**
B. 2SLS							
Teaching: AMA	-0.039			-0.018			-0.013
	(0.0091)**			(0.010)			-(0.010)
Technology: Top Decile		-0.047			-0.019		-0.009
		(0.0099)**			(0.012)		-(0.013)
CMS Quality Score			-0.027			-0.012	-0.010
			(0.0076)**			-(0.0090)	-(0.008)
Mean Hospital log(Cost)				-0.141	-0.141	-0.149	-0.142
				(0.025)**	(0.026)**	(0.027)**	(0.029)**
Observations	667,143	667,143	651,278	667,143	667,143	651,278	651,278
Mean of Dep. Var.	0.366	0.366	0.366	0.366	0.366	0.366	0.366

Note: Each cell represents a separate model analogous to those estimated in Table 3. All models include full controls. AMA: the American Medical Association; CMS Quality Score is normalized with a mean of zero and a standard deviation of one and is available for a slightly smaller subset of hospitals. Standard errors in parentheses, clustered at the HSA level. * significant at 5%; ** significant at 1%
Source: 2002-2008 Medicare Part A Claims Data

Table 9—Ambulance Strategy: Treatment Intensity Measured by Number of Procedures

	OLS (1)	OLS (2)	2SLS (3)	2SLS (4)
One-Year Mortality				
Mean Hospital Number of Procedures	-0.013	-0.010	-0.161	-0.149
	(0.0019)**	(0.0019)**	(0.015)**	(0.014)**
Mean Hospital log(Cost)		-0.018		-0.065
		(0.0035)**		(0.021)**
Observations	667,143	667,143	667,143	667,143
Mean of Dep. Var.	0.366	0.366	0.366	0.366

Note: Each cell represents a separate model analogous to those estimated in Table 3. All models include full controls. Standard errors in parentheses, clustered at the HSA level. * significant at 5%; ** significant at 1%
Source: 2002-2008 Medicare Part A Claims Data

TABLE A1—SUMMARY STATISTICS: AMBUALANCE SAMPLE

	ER Admissions: Full Sample		ER Admissions: Nondeferrable Sample		ER Admissions: Nondeferrable Sample Ambulance Arrivals		Analysis Sample	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Hospital Cost from ER admissions	7,649	4,257	8,515	4,818	9,257	13,706	9,235	13,677
Mean Hospital log(Cost) from ER admissions	8.555	0.406	8.555	0.404	8.694	0.404	8.694	0.401
Teaching Hospital: AMA Recognized	0.411	0.492	0.404	0.491	0.409	0.492	0.410	0.492
Technology Index (top decile)	0.302	0.459	0.297	0.457	0.300	0.458	0.302	0.459
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1-year Mortality	0.201	0.401	0.246	0.431	0.365	0.481	0.367	0.482
Age	75.051	12.179	76.493	11.665	78.200	11.500	78.200	11.500
Male	0.424	0.494	0.423	0.494	0.401	0.490	0.400	0.490
Race: white	0.840	0.366	0.856	0.351	0.862	0.345	0.861	0.346
Race: African American	0.108	0.311	0.096	0.294	0.095	0.293	0.097	0.296
Comorbidity: Congestive Heart Failure	0.151	0.358	0.175	0.380	0.223	0.416	0.224	0.417
Comorbidity: COPD	0.198	0.398	0.232	0.422	0.249	0.432	0.252	0.434
Comorbidity: Diabetes	0.219	0.414	0.207	0.405	0.208	0.406	0.209	0.406
Comorbidity: Other	0.294	0.455	0.312	0.463	0.339	0.473	0.341	0.474
Distance from ZIP centroid to hospital	8.553	16.821	8.452	16.523	8.060	14.000	7.050	7.980
Ambulance Payment	.	.	.	.	299	205	293	163
Ambulance Distance	.	.	.	.	7.540	10.310	7.300	9.390
Advanced Life Support	.	.	.	.	0.646	0.478	0.648	0.478
Ambulance: IV Administered	.	.	.	.	0.097	0.295	0.097	0.295
Ambulance: Emergency Transport	.	.	.	.	0.828	0.377	0.831	0.374
Ambulance: Outpatient Claim	.	.	.	.	0.101	0.301	0.101	0.301
Observations	2,729,156		1,204,291		738,167		667,143	

Note: The full sample includes the top 100 3-digit ICD-9-CM diagnoses among emergency room (ER) admissions, which compres 89% of all such admissions. Distance from ZIP to the hospital was calculated for the 50 closest hospitals, and the number of observations for distance in the full sample is 2,563,145. The nondeferrable sample includes diagnoses most likely to require immediate medical care, as described in the text. The number of observations for distance in this sample is 1,128,411. The ambulance sample includes measures of ambulance inputs, some of which are only available for data from the Centers for Medicare and Medicaid's Carrier file, as opposed to outpatient reimbursement. These include ambulance distance, which is available for 693,376 observations. The analysis sample excludes observations that are not linked to a hospital service area (HSA), patients who were treated more than 50 miles from the ZIP code of their mailing address, patients with missing cost information, and a minumum of 10 observations in the analysis sample for each ZIP code, hospital, and ambulance company.
Source: 2002-2008 Medicare Part A Claims Data

Table A2—Ambulance Strategy: One-Year Mortality Robustness

	Sub-samples				Instrument Calculation				
	Ambulance transports labeled "Emergency"	Exclude patients who were transferred	Patient origin: home or nursing home	Advanced Life Support Ambulances	Varies at the Ambulance Company X Disease Level	Exclude patient's ZIP	Hospital Charges	Cost Quartiles	Teaching Hospital Definition
Mean Hospital log(Cost)	-0.155	-0.161	-0.201	-0.181	-0.139	-0.123			
	(0.029)**	(0.026)**	(0.039)**	(0.028)**	(0.033)**	(0.025)**			
Mean Hospital log(charges)							-0.128		
							(0.020)**		
Teaching: Member COTH									-0.061
									(0.013)**
Hospital Cost: 2nd Quartile								-0.033	
								(0.010)**	
3rd Quartile								-0.054	
								(0.012)**	
Top Quartile								-0.045	
								(0.014)**	
Observations	554704	652174	463892	432349	660305	666110	667143	667143	667143
Mean of Dep. Var.	0.363	0.369	0.381	0.373	0.365	0.366	0.366	0.366	0.366

Note: Each cell represents a separate model analogous to those estimated in Table 3. All models include full controls. COTH: Councile on Teaching Hospitals. Standard errors in parentheses, clustered at the HSA level. * significant at 5%; ** significant at 1%
Source: 2002-2008 Medicare Part A Claims Data

Table A3—Border Strategy: One Year Mortality Robustness

	OLS			2SLS		
	1-Day	7-Day	30-Day	1-Day	7-Day	30-Day
	(1)	(2)	(3)	(4)	(5)	(6)
Sample: Census Block Groups Within 1 Mile of Border						
Mean Hospital log(Cost)	0.000	-0.005	-0.023	0.002	-0.004	-0.043
	(0.002)	(0.003)	(0.006)**	(0.007)	(0.011)	(0.013)**
Observations	142,809	142,809	142,809	142,809	142,809	142,809
Mean of Dep. Var.	0.015	0.049	0.103	0.015	0.049	0.103
Sample: Census Block Groups Within 2 Miles of Border						
Mean Hospital log(Cost)	-0.001	-0.008	-0.022	0.000	-0.007	-0.040
	(0.001)	(0.003)*	(0.005)**	(0.007)	(0.013)	(0.017)
Observations	213,968	213,968	213,968	213,968	213,968	213,968
Mean of Dep. Var.	0.015	0.049	0.104	0.015	0.049	0.104
Sample: Census Block Groups Within 5 Miles of Border						
Mean Hospital log(Cost)	-0.001	-0.011	-0.026	-0.003	-0.019	-0.048
	(0.001)	(0.004)**	(0.006)**	(0.006)	(0.011)	(0.013)**
Observations	281,036	281,036	281,036	281,036	281,036	281,036
Mean of Dep. Var.	0.015	0.049	0.104	0.015	0.049	0.104

Note: Each cell represents a separate model analogous to those estimated in Table 7. All models include full controls. Standard errors in parentheses, clustered at the ambulance district level. * significant at 5%; ** significant at 1%
Source:

Table A4—Border Strategy: First Stage, OLS and 2SLS for Nonelderly (Age 18-64)

Sample:	1 Mile		2 Miles		5 Miles	
	(1)	(2)	(3)	(4)	(5)	(6)
A. First Stage: Mean Area log(Cost)						
Mean Hospital log(Cost)	0.727	0.713	0.730	0.712	0.662	0.644
	(0.150)**	(0.151)**	(0.154)**	(0.150)**	(0.139)**	(0.137)**
Mean of Dep. Var.	8.46	8.46	8.46	8.46	8.46	8.46
B. OLS: 1-Year Mortality						
Average Hospital log(Cost)	0.006	-0.008	0.010	-0.005	0.013	-0.003
	(0.004)	(0.005)	(0.003)**	(0.004)	(0.004)**	(0.003)
C. 2SLS: 1-Year Mortality						
Average Hospital log(Cost)	0.004	-0.019	0.007	-0.011	0.012	-0.010
	(0.008)	(0.013)	(0.011)	(0.013)	(0.013)	(0.012)
Year Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes	No	Yes
Diagnosis Controls	No	Yes	No	Yes	No	Yes
Observations	114,648	114,648	175,161	175,161	233,616	233,616
Mean of Dep. Var.	0.053	0.053	0.054	0.054	0.055	0.055

Note: Each cell represents a separate model analogous to those estimated in Table 7. All models include full controls. Standard errors in parentheses, clustered at the ambulance district level. * significant at 5%; ** significant at 1%
Source: SPARCS Data

Table A5—Border Strategy: First Stage, OLS and 2SLS for Nonelderly (Age 50-64)

Sample:	1 Mile		2 Miles		5 Miles	
	(1)	(2)	(3)	(4)	(5)	(6)
A. First Stage: Mean Area log(Cost)						
Mean Hospital log(Cost)	0.645	0.645	0.603	0.602	0.550	0.549
	(0.179)**	(0.174)**	(0.179)**	(0.170)**	(0.160)**	(0.153)**
Mean of Dep. Var.	8.59	8.59	8.59	8.59	8.60	8.60
B. OLS: 1-Year Mortality						
Average Hospital log(Cost)	0.005	-0.009	0.001	-0.006	0.012	-0.004
	(0.005)	(0.005)	(0.004)*	(0.005)	(0.005)*	(0.004)
C. 2SLS: 1-Year Mortality						
Average Hospital log(Cost)	-0.015	-0.012	-0.008	-0.013	-0.003	-0.018
	(0.023)	(0.022)	(0.025)	(0.020)	(0.029)	(0.021)
Year Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	No	Yes	No	Yes
Diagnosis Controls	No	Yes	No	Yes	No	Yes
Observations	56,788	56,788	86,884	86,884	115,955	115,955
Mean of Dep. Var.	0.081	0.081	0.082	0.082	0.083	0.083

Note: Each cell represents a separate model analogous to those estimated in Table 7. All models include full controls. Standard errors in parentheses, clustered at the ambulance district level. * significant at 5%; ** significant at 1%
Source: SPARCS Data

Table A6—Ambulance Strategy: Correlates with Hospital Cost

	(1)	(2)	(3)	(4)
Teaching: AMA Recognized	0.077			
	(0.012)**			
Technology Adoption: Top Decile		0.115		
		(0.011)**		
CMS Quality Score			-0.008	
			-0.008	
Average number of procedures				0.166
				(0.018)**
Observations	667,143	667,143	454,648	667,143
Mean of Dep. Var.	8.694	8.694	8.683	8.694

Note: Each cell represents a separate model analogous to those estimated in Table 2, but with alternative measures of hospital characteristics. All models include full controls. AMA: the American Medical Association; CMS Quality Score is normalized with a mean of zero and a standard deviation of one. Standard errors in parentheses, clustered at the HSA level. * significant at 5%; ** significant at 1%
Source: 2002-2008 Medicare Part A Claims Data

Table A7—Ambulance Strategy: Alternative Hospital Types: 1st Stages

	Dependent Variable: Hospital Level			
	(1)	(2)	(3)	(4)
Ambulance Level:				
Teaching: AMA Recognized	0.563			
	(0.014)**			
Technology Adoption: Top Decile		0.577		
		(0.013)**		
CMS Quality			0.567	
			(0.019)**	
Average number of procedures				0.277
				(0.012)**
Observations	667,143	667,143	454,648	667,143
Mean of Dep. Var.	0.410	0.459	-0.005	1.340

Note: Each cell represents a separate model analogous to those estimated in Table 2, but with alternative measures of hospital characteristics. All models include full controls. AMA: the American Medical Association; CMS Quality Score is normalized with a mean of zero and a standard deviation of one. Standard errors in parentheses, clustered at the HSA level. * significant at 5%; ** significant at 1%
Source: 2002-2008 Medicare Part A Claims Data