NBER WORKING PAPER SERIES

UNCOVERING WASTE IN U.S. HEALTHCARE

Joseph Doyle
John Graves
Jonathan Gruber

Working Paper 21050
http://www.nber.org/papers/w21050

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2015

We are grateful to Melinda Buntin, Kitt Carpenter, Amitabh Chandra, Lawrence Katz, Sunil Kripalani, Adam Sacarny, Jonathan Skinner, Doug Staiger, David Stevenson and seminar participants at LSE, MIT, Simon Fraser, Texas A&M, University of Connecticut, University of Warwick, Vanderbilt University, and Wharton for helpful conversations and assistance. We gratefully acknowledge support from the National Institutes of Health R01 AG41794-01 The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Uncovering Waste in U.S. Healthcare
Joseph Doyle, John Graves, and Jonathan Gruber
NBER Working Paper No. 21050
March 2015
JEL No. I10,I18

ABSTRACT

There is widespread agreement that the US healthcare system wastes as much as 5% of GDP, yet little consensus on what care is actually unproductive. This partly arises because of the endogeneity of patient choice of treatment location. This paper uses the effective random assignment of patients to ambulance companies to generate comparisons across similar patients treated at different hospitals. We find that assignment to hospitals whose patients receive large amounts of care over the three months following a health emergency do not have meaningfully better survival outcomes compared to hospitals whose patients receive less. Outcomes are related to different types of treatment intensity, however: patients assigned to hospitals with high levels of inpatient spending are more likely to survive to one year, while those assigned to hospitals with high levels of outpatient spending are less likely to do so. This adverse effect of outpatient spending is predominately driven by spending at skilled nursing facilities (SNF) following hospitalization. These results offer a new type of quality measure for hospitals based on utilization of SNFs. We find that patients quasi-randomized to hospitals with high rates of SNF discharge have poorer outcomes, as well as higher downstream spending once conditioning on initial hospital spending.

Joseph Doyle
MIT Sloan School of Management
100 Main Street, E62-516
Cambridge, MA 02142
and NBER
jjdoyle@mit.edu

John Graves
2525 West End Ave.
Suite 1200
Nashville, TN 37203
john.graves@vanderbilt.edu

Jonathan Gruber
Department of Economics, E17-220
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
gruberj@mit.edu

I Introduction

At the heart of healthcare spending reforms is the idea that 30% of spending in the US, or 5% of GDP, may be wasted (Fisher, Bynum, and Skinner 2009; D. Cutler 2010).[1] This idea stems from the striking amount of geographic variation in treatment intensity in the US that yields little apparent benefit in terms of patient health outcomes (Fisher et al. 2003a; Fisher et al. 2003b; Chandra and Skinner 2011). More broadly, the US is an outlier in terms of healthcare spending per capita at 40% more than the next highest-spending country in the OECD (OECD 2014).

There is less evidence on the crucial question of which types of spending are unproductive. In part, this is due to concerns over selection bias. At the patient level, more intensive treatments are chosen for patients who are healthy enough to undergo them yet sick enough to be candidates for such care. Aggregation to markets or countries may not solve this underlying endogeneity problem. Intensive providers likely invest in physical and human capital to deliver different levels of treatment intensity in part based on the types of patients they typically serve and the expertise they develop (Chandra and Staiger 2007).

The aim of this paper is to circumvent the problem of selection bias in order to identify the sources of waste in healthcare spending around episodes of acute care. We develop a framework based on earlier work that leverages the effective random assignment of patients to ambulance companies to provide plausibly exogenous variation in hospital choice (Doyle et al. 2015). A feature of the instrumental variables strategy that stems from ambulance company assignment is that each community provides its own experiment, with ambulance companies delivering patients to hospitals with different treatment patterns. This enables us to compare patients assigned to hospitals with different combinations of treatment intensities. Given the empirical strategy, our approach focuses on patients entering the hospital on an emergency basis. Using longitudinal Medicare claims data from 2002-2012, we observe treatment provided and paid for across institutional settings. This allows us to characterize each hospital with respect to the sources of spending that accrues to its patients for different types of care. In addition, the data

[1] Skinner and Fisher (2010) note that 20-30% is likely an underestimate and note, "At least three other groups have come to 30% waste estimates: the New England Healthcare Institute, McKinsey, and Thomson Reuters."

are linked to vital statistics records that provide our key outcome: one-year mortality.

We have five primary findings. First, our causal framework largely corroborates the cross-sectional result that hospitals with higher total (inpatient and downstream) spending over the 90 day period after an initial hospitalization do not have better survival outcomes compared to those with lower spending levels. Second, we confirm the finding of Doyle (2015) that patients assigned to hospitals with large average inpatient expenditures have lower mortality rates compared to patients assigned to less intensive hospitals. Third, we square these results by finding that patients assigned to hospitals with high average levels of downstream spending have substantially higher mortality rates compared to those treated in hospitals whose patients receive lower amounts of such care. Fourth, based on these findings we further investigate categories of post-acute care and find that the positive relationship between downstream spending and mortality is concentrated in care provided at Skilled Nursing Facilities (SNF).

Fifth, the positive relationship between SNF spending and mortality could be due to hospitals discharging patients in worse health and/or low-quality care downstream, and either explanation motivates our investigation of a new quality measure that can be incorporated into performance comparisons and payment reform models: the rate at which patients are admitted to SNFs within 30 days of discharge. Specifically, a contribution of the paper is that we compare our risk-standardized SNF admission rate measure to other commonly used (but controversial) measures of hospital productivity based on mortality and readmission rates to provide some of the first causal evidence on whether these influential quality scores are predictive of subsequent outcomes when we control for patient selection. The results demonstrate that patients treated at hospitals with high rates of SNF admission have higher mortality and, conditional on spending during the initial hospitalization, higher spending during the year after discharge. We also demonstrate that a composite "home and alive" rate that incorporates SNF admissions with mortality and readmission also holds promise as a summary measure of hospital productivity based on observed patient outcomes; this measure has the added advantage that it avoids issues with competing risks when mortality and readmissions are compared separately.

Our findings are consistent with recent work that points to post-acute care as a potential culprit for waste in the U.S. system: such care is a major contributing factor to residual geographic variation in healthcare spending among the over-65 population (Newhouse, Garber, and Graham 2013; Newhouse and Garber 2013), a result found for the under-65 population as well (Franzini et al. 2014). Current reform proposals bundle payments to providers for 90-days after a hospitalization, and penalize for hospitals for readmissions that occur within 30 days of discharge. There is hope that such reforms will provide incentives to coordinate care across inpatient and outpatient institutions in a way that reduces costs and improves health. We discuss the implications of our findings for such proposals.

The remainder of this paper proceeds as follows. Section II provides background on the payment reforms and the way hospitals are compared in current policy experiments. Section III describes the empirical framework, while Section IV describes the data. Section V presents the results for the impact of spending on outcomes. Section VI then extends the analysis to integrate post-acute measures into the discussion of hospital quality indicators. Section VII concludes and discusses policy implications.

II Background

II.A Provider Incentives

As noted in the introduction, there is a large literature documenting that, on average, high-spending hospitals and high-spending geographic areas do not have better outcomes compared to lower-spending ones. This raises the fundamental question: where is the unproductive spending in the U.S. healthcare system?

In efforts to control healthcare spending, the main idea in current policy discussions is to remove the incentive to provide too much care created by a system that pays a fee for every service. Instead, proposals call for fixed, rather than marginal, payments and reward or penalize providers for their performance on quality measures to guard against sub-optimal care. For example, the Affordable Care Act promotes the formation of vertically-integrated providers, Accountable Care Organizations (ACOs), with the aim to coordinate care across different types of providers. A related approach (bundled payment) pays providers for an amount for a period of care after a hospitalization. Typically, bundled payments cover up to 90-days of care and vary at the level of the diagnosis.[2] Under both approaches the hope is that providers will know (or will learn) what types of care can be reduced or improved without harming patients.

A related approach penalizes hospitals for poor quality performance. For instance, the Medicare Hospital Readmission Reduction Program (HRRP) lowers all Medicare payments by up to 3% for hospitals with above-average 30-day readmission rates among acute myocardial infarction, chronic heart failure and pneumonia patients (Berenson, Paulus, and Kalman 2012). This has spawned efforts by hospitals to reduce readmissions through outreach to patients and better coordination with primary care physicians (Kripalani et al. 2014; Ahmad et al. 2013).

This paper considers risk-standardized hospital-level measures of Medicare spending per beneficiary to all providers up to 90 days after acute episodes. There are a few advantages of this type of approach. First, there is reason to believe that coordination of care after a hospitalization is cost effective: it provides an incentive for better care transitions and reduces readmissions (Naylor et al. 2011). This junction of the US healthcare system is often the target of suspicion for a major source of waste, as coordination across providers is necessary to achieve the gains, yet is often not reimbursed by payers; indeed, in an effort to improve hospital quality the Centers for Medicare and Medicaid Services began publicly reporting similar hospital-level measures of 30-day Medicare Spending Per Beneficiary.. Second, the care plan after an acute health problem may be more directly under the control of a physician than is total annual patient spending, so that it can provide a natural basis for reimbursement. Third, spending on such care is substantial. Cutler and Ghosh (2012) find that capping episode-based bundled payments to the median level across markets would yield nearly the same savings as capping an annual payment per beneficiary to its median level.

[2] In the Episode of Care Payment Demonstration project, a typical hospital would receive $21,000 for 90 days of care after a heart attack hospitalization. See, for example, Table 3-3 in http://www.medpac.gov/documents/reports/jun13_ch03.pdf?sfvrsn=0

II.B Ambulance Referral Patterns

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. Critically, areas are often served by multiple ambulance companies, and the assignment of the ambulance company to the patient is effectively random.[3]

In particular, patients are transported by different companies for two main reasons. First, large cities such as New York, Los Angeles and Chicago contract with private ambulance companies to work in conjunction with fire departments to provide Emergency Medical Services (EMS) (Johnson 2001). Chiang, et al. (2006) 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. 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). In these 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. 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.

In addition to plausibly exogenous assignment, ambulance companies are expected to have preferences for particular hospitals. In survey work described in Doyle et al. (2015), we found that paramedics have developed relationships with local emergency departments. For example, Skura (2001) studied ambulance assignment in the wake of a new system of competition between public and private ambulances in New York City. He found that patients living in the

[3] This section builds on our previous description in Doyle et al. 2015.

same ZIP code as public Health and Hospital Corporation 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.

More broadly, with the exception of acute trauma care for which there are often defined local protocols for hospital assignment (Kahn et al. 2008) transport assignment for other emergencies is more likely to be driven by idiosyncratic preferences. As noted in one Institute of Medicine report, ambulance personnel often "lack the means to determine which hospitals can provide the best care to a patient" (Institute of Medicine 2010). Thus, in this context patient transport decisions are more likely to be made in ways that are not systematically tied to the underlying health of the patient. This combination of exogenous assignment of ambulance companies, coupled with their preference for taking patients to certain hospitals, provides an empirical lens to compare similar patients who live nearby one another but visit different hospitals.

III Empirical Framework

Our main regression of interest is the relationship between spending patterns accrued by patients assigned to particular hospitals, such as average 90-day spending among other patients, H, and outcomes such as mortality, M, for patient i with principal diagnosis $d(i)$ originating from a particular point of origin $o(i)$ (home, nursing home or elsewhere) within ZIP code $z(i)$ in year $t(i)$:

$$M_i = \beta_0 + \beta_1 H_i + \beta_2 X_i + \beta_3 A_i + \gamma_{d(i)} + \theta_{z,o(i)} + \lambda_{t(i)} + \epsilon_i \tag{1}$$

where X_i is a vector of patient controls including age, race, and sex, and indicators for 17 common comorbidities controlled for when the Centers for Medicare and Medicaid Services (CMS) computes quality scores.

Ai 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 has Advanced Life Support (e.g., paramedic) capabilities; whether the transport was coded as emergency (i.e., “lights and sirens”) transport; and whether the ambulance was paid through the outpatient system rather than the carrier system.

We cluster standard errors at the Hospital Service Area (HSA) level, as each local market may have its own assignment rules. In addition to one-year mortality as an outcome, we will also consider one-year Medicare spending downstream of the index event.

Patient selection is likely to confound the structural equation (1), so we estimate it using two-stage least squares. To operationalize ambulance preferences, we calculate a set of instrumental variables based on the characteristics of hospitals where each ambulance company takes other patients – a leave-out mean approach that helps avoid weak instrument concerns (Kolesar et al. 2011; Stock, Wright, and Yogo 2002). For patient i assigned to ambulance $a(i)$, we calculate the average hospital measure H_j (e.g., average 90-day log spending) among the patients in our analysis sample for each ambulance company:

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

This measure is the ambulance company fixed effect in a model that predicts H_j, leaving out patient i.[4]

We use this instrument to estimate the first-stage relationship between hospital spending measures H and the instrument, Z: the hospital measure associated with the ambulance assigned to patient i:

$$H_i = \alpha_0 + \alpha_1 Z_{a(i)} + \alpha_2 X_i + \alpha_3 A_i + \gamma_{d(i)} + \theta_{z,o(i)} + \lambda_{t(i)} + \nu_i \tag{2}$$

This regression, in other words, compares individuals who live in the same ZIP code and are

[4]The first-stage estimates below do not take into account the noise created when estimating this fixed effect, however we expect this to be a small adjustment given that the average number of observations used to calculate this fixed effect is over 1,800.

picked up from similar location (e.g., at home), but who are assigned ambulance companies with different "preferences" across hospitals with different spending patterns.[5] A positive coefficient α_1 would indicate that ambulance company "preferences" are correlated with where the patient actually is admitted.

Doyle et al. (2015) focuses on inpatient spending at the time of the health shock and discusses at length potential limitations with this strategy and various specification checks that begin to address them. In particular, that study finds results that are robust to controls for both patient characteristics and the characteristics of pre-hospital care in the ambulance itself; results that are robust to the level of heterogeneity in the demographics of the zip codes, which suggests that within-ZIP differences in patient assignment is not driving the assignment; that the rate of admission to a hospital among emergent patients is not correlated with the ambulance level instrument, suggesting that selection into hospital admission conditional on ambulance use is not a concern; and the impact of ambulance assignment on health outcomes occurs not in the first day but over longer horizons, which is suggestive that the health of patients upon entry to the hospital does not vary substantially across ambulance companies.

IV Data Description

IV.A Medicare Claims Data

Our primary analytic sample draws on a 20% sample of Medicare beneficiaries aged 66 and older with inpatient admission between 2002 and 2011. We limit this sample to beneficiaries with at least 12 months of continuous enrollment in the fee-for-service Medicare program before and after their index hospitalization, and to those without an admission for the same condition in the year prior to that admission. We use the sample of Medicare claims between 2001 and 2012 to identify comorbidities in the year prior to the indexing admission, and to construct hospital-level spending measures for up to one year after their index hospitalization.

[5] A particular concern is that an ambulance affiliated with a nursing home will systematically pick up older, sicker patients. We hope to avoid this by using ZIP-by-origin fixed effects. Results are similar when we exclude the point-of-origin fixed effects, however.

CMS reimburses ambulance companies using two systems captured by the Carrier file and the Outpatient claims file. We can access Carrier claims for a 20% random sample of beneficiaries, and 100% of outpatient claims. Most ambulance claims are paid via the Carrier claims, and we increase our sample by 6% by including the outpatient claims – claims that are affiliated with a hospital or other facility file. We link each ambulance patients claims to her inpatient claims in the Medicare Provider Analysis and Review (MEDPAR) files, which records pertinent information on date of admission, primary and secondary diagnoses, procedures performed, and payments made by CMS. Diagnoses and procedures recorded in each patients claims for the year leading up to the ambulance-linked admission are then mapped to Condition Codes (CC) to construct a set of comorbidity measures. We also link each ambulance patient to information on age, race, and gender. 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.

The carrier data also include information about the ambulance visit. First, to control for the location where the patient is first contacted by the ambulance company, the data contain a patient origin variable that includes home, nursing home or other non-acute care facility, and scene of an accident. Second, the driving distance from the pick-up location to the hospital is recorded because Medicare reimburses ambulances in part based on distance traveled with the patient. The reimbursement made to the ambulance company also provides a summary of the amount of care provided prior to arriving at the hospital. We also observe whether the patient is assigned to a basic life support ambulance, which provides care administered by Emergency Medical Technicians, or advanced-life support care provided by more highly trained paramedics.

As described in more detail below, we also draw upon the Medicare claims in the construction of additional provider-level spending and outcome quality measures. Finally, vital statistics data that record when a patient dies are linked to these claims. This allows us to measure our main outcome, one-year mortality.

IV.B Sample Construction

We rely on a sample consisting of patients admitted to the hospital with "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, Dobkin, and Maestas (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. In addition to these conditions, we also draw upon the set of non-discretionary emergency conditions based on an expert physician panel as reported in Mulcahy et al. (2013). Appendix Table A1 shows the distribution of admissions across these diagnostic categories, which include diagnoses such as acute myocardial infarction (heart attacks), strokes, and hip fractures. These are the types of diagnoses that are particularly costly and are candidates for episode-based bundled payments (Cutler and Ghosh 2012). Our condition set represents roughly 30% of all hospital admissions among Medicare beneficiaries in 2011, 75% of which arrived through the emergency room.[6] Among those who arrive via the emergency room, we estimate that 40% arrived via ambulance.

We further limit the sample to patients first observed in this ambulance-transport sample of diagnoses, and patients who have not been admitted to the hospital with a principal diagnosis for one of the non-deferrable conditions in the prior year. Last, we limit the sample to individuals in fee-for-service Medicare for at least one year after the index admission so that we can observe uncensored Medicare spending over that time period. Our final analytic sample is comprised of 1,582,421 patients.[7]

The analysis sample is restricted to relatively severe health shocks where there is relatively

[6] Author tabulations of the 2011 National Inpatient Sample.

[7] This sample is larger than previous versions of this paper due to the addition of two new years of Medicare claims (through 2012), additional non-discretionary diagnosis ICD-9 codes based on Mulcahy et al. 2013). In addition, we no longer restrict our sample to patients with no inpatient admissions within the last year. We found that our main results were robust when we relaxed this inclusion criteria to only restrict to patients with no admission for the principal diagnoses studied here within the last year, and results in a larger sample size.

little choice but to seek treatment. The estimates of the effects of hospital types on mortality apply to these types of episodes. We caution against extrapolating our results to other sources of medical spending, such as most treatment for chronic disease; we discuss this point further in the conclusion.

IV.C Medicare Spending Per Beneficiary Measures

Each beneficiary in our analysis sample has a unique index event associated with the inpatient admission via an ambulance. For each hospital, we compute the average risk-standardized spending (total, inpatient and non-inpatient) for the index episode and over the following 90 days after admission for all other patients treated by the hospital in our analysis sample. Our inpatient spending measure includes all hospital facility payments as well as doctor payments for acute care services and all emergency department spending. The non-inpatient spending measure includes Medicare payment for Part B services not provided in a hospital, all outpatient care (except emergency room use), skilled nursing facilities, home health care, hospice, and durable medical equipment. The only Medicare reimbursement category not included in our data is pharmaceuticals provided outside of the hospital setting. Each spending measure is constructed using a procedure similar to that used to construct the 30-day Medicare Spending Per Beneficiary quality measure for U.S. hospitals, as well as the previous cross-sectional comparisons across hospitals and markets:[8]

(a) Calculate Expected Episode Spending. We utilize an ordinary least squares regression (OLS) model that controls for age, race and gender.

(b) Truncate and Normalize Predicted Values. The predicted values from the OLS regression model are truncated at the 0.5th percentile to reduce the influence of extreme predictions. Predicted values are then normalized so that average 90d spending is the same before and after truncation.

[8] See the Medicare Spending Per Beneficiary methodology report (http://www.qualityforum.org/Projects/c-d/Cost_and_Resource_Project/2158.aspx) for full details.

(c) Calculate Residuals. We calculate the residuals for each patient episode as the difference between the observed 90-day spending and the (truncated) predicted value.

(d) Exclude Outliers. For constructing the measure, we exclude all observations with residuals above the 99th percentile and below the 1st percentile.

(e) Calculate The Spending Measure at the Hospital Level. The hospital-level spending measure is estimated as ratio of the average observed spending for the hospital to the average predicted spending (from (a)) for the hospital, multiplied by average 90-day spending in the sample.

This approach allows us to assess the impact in spending relative to the hospitals underlying mix of patients, e.g. to ask whether a hospital is high spending given its patient mix.

For all of our estimates below, each spending measure enters as a continuous measure that has been demeaned and scaled by 2 standard deviations to facilitate interpretation. Thus, all coefficients reflect a difference of between one standard deviation above vs. below the mean (i.e., coefficient values reflect comparisons between "low" vs. "high" spending hospitals).

IV.D Summary Statistics

Table 1 reports summary statistics for the analysis sample. 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. This sample is slightly older (average age of 82) compared to all Medicare patients. 38% are male, and 90% are white. Common comorbidities measured in the year preceding (and including) the initial episode hypertension (41%), chronic obstructive pulmonary disease (COPD, 21%), diabetes (20%), and pneumonia (19%).

The third column reports the standardized differences in means for the 90-day total spending measure: the difference in the mean of the covariate when the instrument is above versus below its median value computed from a regression model that controls for ZIP code by patient-origin fixed effects, divided by the pooled standard deviation. Relative to commonly-used maximum standardized difference thresholds for assessing sample balance (e.g., a maximum of 0.25 standard deviations; see Rubin, 2001), these standardized differences are remarkably small across the wide range of control variables, consistent with the effective random assignment of ambulance companies to patients. A similar level of balance is found for 90-day spending on inpatient and non-patient categories as well.

V Results

V.A First Stage

Table 2 presents the main results. The main explanatory variable in Panel A is the total 90-day spending measure for the hospital. The first column includes ZIP-by-patient origin and year fixed effects. The second column adds controls for patient characteristics, including principal diagnosis, demographics, co-morbidities, and pre-hospital care. Average 90 day spending for our hospitalized patients is $27,351, and a two standard deviation increase should be regarded as a $8,486 increase.[9] Inpatient spending constitutes the majority of spending, with an average of $15,876 and two standard deviations totaling $6,226. Non-inpatient spending over the 90 days after discharge averages $10,557, and two standard deviations totals $3170.

With or without controls, the first-stage coefficient is 0.19. This means that an increase in the average 90-day spending measure for the hospitals where the ambulance company takes other patients by 1 (a 2 s.d. increase) is associated with a 0.19 increase in the 90-day spending measure where the patient is actually treated, an increase of 0.19*2 = 0.38 standard deviations.

The fact that the relationship between the ambulance and hospital measures is not one-to-one is illuminating about the nature of the variation used in the instrumental-variables results. Consider an ambulance company that ordinarily treats patients in a geographic area and takes seriously-ill patients to a particular hospital. That ambulance company is then called in to a nearby area via a mutual aid agreement. The first-stage results suggest that this patient is much

[9] All dollar amounts are in 2012 dollars using the CPI-U and means and standard deviations for the main explanatory variables are reported in Table A2.

more likely to be transported back to the ambulance companys usual hospital, but at a lower rate than the rate at which it transports its usual patients.

V.B 90-day total spending

Panel B shows the results that relate 90-day total spending to one-year mortality. The OLS results show that patients who are sent to high 90-day spending hospitals have modestly lower mortality rates compared to hospitals with lower 90-day spending: a two standard-deviation increase in the measure is associated with a 1.8 percentage point reduction in the absolute risk of mortality compared to a mean of 43% in models with ZIP-by-patient origin, and year fixed effects. This coefficient falls to 1.1 percentage points when controls are included. That is, the OLS coefficient suggests that a two-standard deviation in area health care spending leads to a 2.6% reduction in mortality compared to the mean. This is consistent with the broader cross-sectional literature across markets, which as noted earlier finds at best a modest association between higher spending levels and lower mortality levels.

The instrumental variable estimates are reported in the third row. Here, the point estimate with the baseline controls is 1.7 percentage point reduction in mortality when patients are transported to hospitals with high spending levels compared to hospitals with lower 90-day total spending levels. The estimate with the full set of controls is almost identical.

Our first conclusion is therefore that the overall lack of meaningful relationship between spending and mortality is not driven by patient selection. Our OLS and instrumental variables estimates are similar, and both show only a very modest relationship between total 90-day spending and mortality. We view these results as corroborative of the cross-sectional results that hospitals with high levels of spending do not have improved outcomes compared to low-spending hospitals (Barnato et al. 2010; Fisher et al. 2003a; Fisher et al. 2003b; Skinner and Fisher, 2010).

V.C Inpatient vs. Non-inpatient Spending

Table 3 extends the analysis by disaggregating 90-day total spending into separate risk-standardized measures of 90-day inpatient and 90-day non-inpatient spending. We begin by showing the first stage estimates for both in Panel A, both with and without additional controls. Unsurprisingly, the first stage coefficient is stronger for inpatient care for than for post-acute care, since the direct assignment to hospital is more impacted by ambulance preferences than is the ultimate disposition of downstream care. Nevertheless, the first stage estimates are again highly statistically significant for both types of care.

Panel B shows OLS results, separately for inpatient and non-inpatient, post-acute spending, first with no controls, then with controls included. The OLS estimates continue to show a lack of a relationship between spending levels and mortality. In a model with full controls, a two standard deviation increase in inpatient spending over the following 90 days is associated with a 1.1 percentage point reduction in mortality. For patients admitted to hospitals with high levels of non-inpatient spending over the subsequent 90 days, the estimated effect is very close to zero.

The magnitude of the estimates changes when we attempt to control for patient selection via instrumental variables. Patients admitted to hospitals with high levels of inpatient spending have substantially lower mortality rates: a point estimate of 4.7% in the model with baseline controls and 4.2% in the model with full controls. That is, mortality is 10% lower than the mean for patients who are transported to high-spending hospitals.[10]

In contrast, the instrumental variable estimates imply that patients admitted to hospitals with high levels of spending on non-inpatient downstream spending have significantly higher mortality rates. A two standard deviation increase in such spending is associated with a 6.3 percentage point increase in mortality with baseline controls and a 5.4 percentage points in a model with full controls. That is, we find that patients who are admitted to hospitals that typically have high post-acute spending are much more likely to die in the subsequent year.[11]

[10] We considered separating our inpatient spending into pre and post-discharge spending due to readmissions. However, post-discharge inpatient spending is so highly correlated with spending levels at the index admission that we cannot separately identify the two measures of resource intensity. Given their policy relevance, we consider readmissions directly below.

[11] One concern is that hospitals that are high-spending on average may not be high spending for a particular

This is a striking finding that confirms the recent Institute of Medicine finding that post-acute hospital spending is a major source of residual variation in U.S. health care spending (Newhouse, Garber, and Graham 2013). As in Doyle et al. (2015), hospitals with higher inpatient spending intensity achieve lower patient mortality. But as in the various studies cited earlier such as Fisher et al. (2003a) and Skinner and Fisher (2010) overall spending does not appear productive. Our results point to lack of productivity of spending after hospital discharge as a possible resolution of these discordant results.

V.D A Closer Examination of Post-Acute Care

The fact that post-acute care is linked with increased mortality risk leads us to further explore the possible sources of inefficiency within this category. Table 4 explores these results in more detail.

The first question we address is whether our results are driven by an underlying correlation between inpatient and post-acute spending. Suppose that certain hospitals are very productive, and they feature high inpatient spending and low post-acute spending; other hospitals are the opposite. Then our finding may reflect not the lack of productivity of post-acute spending, but simply heterogeneity across hospital types.

To investigate this mechanism, we begin by simultaneously allowing both inpatient and post-acute spending to impact patient outcomes; that is, controlling for a hospitals level of inpatient spending, is its reliance on post-acute spending still unproductive? Panel A of Table 4 shows that, in OLS, the results when both measures are included simultaneously are identical to the estimates in Table 3 where they are included separately.

In the first stage for our 2SLS estimates (shown in the first two columns of appendix table A3), two instrumental variables are used, reflecting the inpatient and non-inpatient spending

condition. This would result in a failure of the monotonicity condition that enables interpretation of the results as local average treatment effects: assignment to a "high-spending" ambulance company need not imply that the hospital is high-spending for the patients condition. As a robustness check, we calculated the instrument at the hospital-major diagnosis level defined by 5 main categories of disease (circulatory, respiratory, injury, digestive, and all other) and found similar results: a coefficient of -0.074 (s.e.=0.011) for 90-day inpatient spending and 0.064 (s.e.=0.014) for non-inpatient spending.

levels of the hospitals where ambulances typically take other patients. We find that the ambulance measure for inpatient spending is highly predictive of inpatient spending itself, and likewise that the measure for post-acute spending is highly predictive of post-acute spending. We find smaller negative cross-effects of both measures: high post-acute spending hospitals are less likely to be high inpatient spending hospitals, for example.

Panel B of Table 4 then illustrates the 2SLS relationship when both inpatient and post-acute spending measures are included. In fact, as with OLS, we find virtually the same results that were estimated when the measures are included separately; given the modest cross-correlation in the first state model, this is not surprising. Therefore, we conclude that the inefficiency of post-acute spending is not driven solely by a negative correlation with productive inpatient spending.

To examine the potential sources of post-discharge care related to higher mortality, we considered each major post-acute spending category. The one category that drives the positive relationship is spending at Skilled Nursing Facilities (SNFs). The remaining columns of Table 4 decompose post-acute spending into spending at SNFs and other post-acute spending. SNF spending is the single largest category of post-acute spending, accounting for 40% of such spending.

The OLS estimates show that SNF spending is not correlated with mortality, and that non-SNF downstream spending is associated with lower one-year mortality. Again, this could be confounded by non-random assignment of patients to this type of spending.

The IV estimates in Panel B of Table 4 show in fact that there is a very strong and positive relationship between SNF spending and mortality, with a more modest negative impact of non-SNF spending. When we incorporate inpatient spending in the last column (3), the inpatient spending coefficient declines, although this model is controlling for downstream SNF spending, which could reflect the productivity of the inpatient spending in a way that competes with the inpatient measure itself. Meanwhile, "other downstream spending" is associated with lower mortality. In summary then, the results are consistent with both inpatient and "other post-acute spending" being productive in terms of reducing patient mortality, but with the latter is offset by the inefficiency of SNF spending. An implication of these findings is a new quality measure beyond the currently used measure of hospital readmissions: SNF admissions.

VI Downstream Care as a Quality Measure

As noted earlier, hospital report card measures of patient outcomes are a key component of efforts to improve and pay for health care quality. Starting in 2012, for example, the U.S. Medicare program began penalizing hospitals based on 30-day readmission rates among acute myocardial infarction (AMI), pneumonia (PNA) and chronic heart failure (CHF) patients (Berenson, Paulus, and Kalman 2012; Shahian et al. 2010; Shahian et al. 2012; Black 2010). Similarly, risk-standardized mortality rates are used to adjust Medicare payments under the Hospital Inpatient Value-Based Purchasing program, and are included in numerous international efforts to profile hospital quality.(Shahian et al. 2010; Shahian et al. 2012; Black 2010; Hawkes 2010; Lilford and Pronovost 2010; Keogh 2013).

The use of quality performance metrics has accelerated despite concerns over the accuracy and reliability of measurement. For example, researchers have questioned whether case-mix adjustment using administrative claims can adequately control for differences in patient acuity across hospitals (Asch et al., 2006; Joynt and Jha, 2013a). Performance measures may also be biased if the recorded diagnoses used for case-mix adjustment also capture the confounding influence of reimbursement system incentives on what is entered (or not) in a billing claim (Song et al. 2010; Werner 2005). Our empirical strategy can shed light on the reliability of these measures.

We also consider a new quality measure, the risk-standardized 30-day SNF admission rate, and compare it to similarly constructed (and more widely used) summary measures of hospital productivity based on 30-day readmission and mortality.

VI.A Constructing Quality Measures

To construct each measure, we first fit a linear probability model that regresses the outcome of interest (e.g., SNF admission within 30 days of discharge) on controls for patient demographics and comorbidities and a hospital fixed effect. More formally, define Y_ih as a binary indicator of whether patient i treated at hospital h experienced the outcome, and define X_ih as a vector of patient-level characteristics including a constant. We then fit the following model:

$$Y_{ih} = \alpha_h + \beta X_{ih} + \eta_i \tag{3}$$

In choosing case-mix adjustment variables (X_ih), as well as the risk-standarized rate estimator described below, we mimic to the extent possible those used in the national risk adjustment model utilized by the Centers for Medicare and Medicaid Services (CMS) in the construction of 30-day outcome rates for the Hospital Compare Program (Ash et al. 2012).

In contrast to the CMS approach, however, we elect to use a hospital fixed effect (α_h) rather than a random effect. That is, we do not assume a probability distribution around α_h. Our fixed effects approach relies less on the functional form assumptions of a random effect approach, but tends to be less precise. To help stabilize outcome rates for low-volume facilities we therefore utilize an empirical Bayes approach to "shrink" the fixed effects (α_h^{eb}). Our results do not materially change in models based on a hospital random effect, or when we estimate a standard fixed effect without any shrinkage; however, estimates that employ the shrinkage are noticeably more precise, as expected.

For outcome Y with mean $\bar{Y}$, the risk-standardized outcome rate for hospital h (RSR$_h$) is based on the following estimator:

$$\text{RSR}_h(X_{ih}) = \frac{E(Y_{ih}|\alpha_h^{eb}, \beta, X_{ih})}{E(Y_{ih}|\beta, X_{ih})}\bar{Y} \tag{4}$$

We estimate the numerator of equation (4) by fitting equation (3) and then summing the predicted values for each patient in the hospital, where the predicted values are based on observed patient values X_{ih}, $\hat{\beta}$, and the empirical Bayes-adjusted hospital fixed effect α_h^{eb}.[12] The denominator is similarly estimated as the sum of patient-level predictions, but only the observed patient-level predictors and the constant term are used.[13] As noted above, each quality measure has been demeaned and scaled by 2 standard deviations to facilitate interpretation.

VI.B SNF Admission Rate as a Quality Measure

Our results thus far indicate that high inpatient spending leads to better patient outcomes, but that going to a hospital where other patients accrue high levels of SNF spending have worse outcomes. This motivates our investigation of the hospitals 30-day SNF admission rate as a quality measure. In particular, we test whether this measure predicts mortality as well as total one-year spending.

SNF admission will have direct implications for Medicare spending, but could also be indirectly related to one-year spending: initial spending at the time of an acute episode (the "index event") may improve health such that patients are able to return home, and SNF use could be a substitute for subsequent hospital readmissions. To investigate these relationships, we report results that include the SNF admission rate on its own in the model, and results that simultaneously estimate the effects of SNF admission and index-event spending within the hospital. Indeed, index spending is negatively associated with SNF admissions, another suggestion that this measure of spending is productive, as shown in our first-stage results in Table A4.

Table 5 shows the main SNF admission results. The first panel of the table shows OLS results, while the second panel uses our instrumental variables framework. The IV framework is the same as that laid out in equation (2), but where hospital spending is replaced with the risk-standardized SNF admission rate. That is, the instrument for the hospitals SNF admission rate

[12] The structural relationship underlying equations (3) and (4) – including the use of empirical Bayes methods to shrink the hospital fixed effects – is similar to that used to estimate the hospital production function in Chandra et al. (2013). Chandra et al. use the hospital fixed effects as an estimate of total factor productivity (TFP) in the production of log(survival days), while in our approach the fixed effects are the underlying source of variation in the risk-standardized outcome rates for one-year mortality, readmission and SNF utilization.

[13] One additional concern arises due to mechanical endogeneity induced when the risk-standardized rate for other patients also includes predicted values for the patient herself. To eliminate this concern in the construction of our outcome rate-based instruments, for each patient in our sample we estimate risk-standardized outcome rate for other patients treated at the same hospital excluding the patient herself.

is the average rate at the hospitals where the ambulance company brings other patients.

The first column of Table 5 shows the impact on mortality. There is a small but significant positive association between the hospitals SNF admission rate and mortality in OLS which grows substantially when estimated using the instrumental variables framework. Our IV results indicate that each two standard deviation increase in the hospitals SNF admission rate leads to a 3.1 percentage point rise in mortality, or 7% of the baseline. Thus, being brought to a hospital with a high SNF admission rate sizably increases the absolute risk of death of the patient over the next 365 days. In the second column of Table 5 we include hospital spending for the index admission. We find that including this spending measure only modestly reduces the coefficient on the SNF admission rate.

We find that hospitals with higher SNF admission rates have lower spending over the year after admission. Once again, however, this result may reflect either hospital heterogeneity (high inpatient spending hospitals have higher total costs, lower mortality, and lower SNF admission rates) or a causal relationship with SNF admissions (hospitals that have high use of SNFs save money while lowering patient care quality). To address this, we include in the regression hospital index admission spending. While this had only a modest impact on the mortality effect of the SNF admission rate in column (2), it has a strong effect on the cost impacts: conditional on index spending, greater SNF use raises total one-year spending.[14] SNFs are expensive in terms of post-hospitalization care, which need not have been the case if SNF use substituted for subsequent inpatient admissions, for example.

To the extent that hospitals can improve the quality of care such that patients are healthy enough to return home upon discharge, as opposed to relying on the use of SNFs, these results suggest that such a reliance on SNF admissions is indeed wasteful: being randomly assigned to a hospital with a higher SNF admission rate, conditional on hospital spending, leads to worse

[14]When the model includes risk-adjusted index spending but excludes the SNF admission rate, the estimated coefficient for spending is -0.043. Doyle et al. (2015) also find a substantial negative relationship between index spending and mortality but that paper focuses on average log(spending) rather than the risk-standardized measure used here. Both spending measures answer related, but ultimately distinct, questions on the returns to Medicare spending overall vs. the returns to spending above and beyond what the case-mix would predict. The present measure is policy relevant given recent payment reform models that use risk-adjustment; the current results can also be more readily compared to the previous cross-sectional literature, which also risk adjusts.

outcomes and higher costs.

VI.C Comparison to Other Quality Measures

A natural question is how the 30-day SNF admission rate measure – which attempts to capture the sources of inefficiency of health care spending identified earlier – compares with other commonly used summary measures of hospital productivity based on mortality and readmission. We next turn our attention to this question, and also consider a composite measure that combines readmission, mortality and SNF admission into a single summary measure that we label "home and alive" (i.e., the patient has survived and was not readmitted to an acute care hospital or admitted to a SNF within 30 days of discharge). Because it is a composite outcome, the "home and alive" rate measure has an added advantage of avoiding the competing risk problem inherent in separate comparisons of mortality and utilization: mortality prevents subsequent readmission and SNF utilization (Joynt and Jha 2013b; Krumholz 2013; Press et al. 2013).

Results on each outcome quality measure (constructed on the same sample and based on equations 3 and 4 above) are presented in Table 6. The first panel shows OLS results, while the second shows 2SLS; the third panel shows 2SLS conditional on index admission index spending. First stage results for all of these regressions are shown in Appendix Table A5. The first set of columns show results for mortality, while the second set show results for spending. The first row in each panel replicates the results shown previously for the SNF admission rate.

The second and third rows present results on more widely-used outcome measures based on readmission and mortality. We find that the hospital readmission rate is only weakly related to mortality outcomes using OLS, but that once instrumented there is a positive and statistically significant relationship; patients treated in hospitals with higher readmission rates have a slightly higher absolute risk of mortality within 1 year. The results are much stronger for the hospital mortality rate measure: the marginal patient is much more likely to die in hospitals with a higher mortality rate in OLS, and that relationship is only stronger when instrumenting. This suggests that the risk-adjusted mortality rate quality measure is useful in characterizing hospital quality, as opposed to concerns that the measures reflect patient selection that confounds the rates. Finally, we find a negative relationship between the composite "home and alive" measure and one-year mortality outcomes: patients treated at hospitals that score well on this measure have significantly improved survival outcomes.

It is hard, however, to compare these alternative quality measures in terms of mortality without also understanding their impact on spending. So the next set of columns (5-8) replaces mortality as a dependent variable with log(total 365) day spending on the patient.

The most striking finding from these results, in combination with the earlier mortality findings, is that the hospital readmission rate predicts worse outcomes as well as higher spending. We find that a two standard deviation increase in the readmission rate raises log spending over the first year by 5.3%, in addition to the aforementioned 1.4% increase in mortality in our non-deferrable condition sample. The other quality measures all show larger effects on mortality but savings over the first year as well. Hospitals that have high rates of mortality and SNF admission rates, and low rates of "home and alive", have higher mortality but lower spending.

As noted earlier, however, the current findings may reflect a correlation between initial hospital spending and outcomes along these dimensions. To address this point, the third panel of Table 6 replicates the 2SLS results from the second panel, but includes index hospitalization inpatient spending as well. As in Table 5, the inclusion of index hospital spending does not have a very large impact on estimated mortality effects, which are all slightly reduced. The effect on spending is much larger, however. As in Table 5, there is now a positive effect of the SNF readmission rate on spending, consistent with higher SNF readmission rates being associated with inefficient care conditional on initial hospital spending. The effects of the hospital mortality rate measure are quite robust. The home and alive rate now shows that hospitals with a higher rate of individuals discharged home and alive can produce improved health very cost effectively conditional on inpatient spending.

VI.D SNF Quality Measures

Our results thus far indicate that SNF spending and utilization is associated with inefficient patient care. We have not, however, distinguished between two potential explanations for this relationship. The first is that SNF use itself is harmful. The second is that SNF use is not harmful, but instead that use of the SNF is a marker of poor hospital quality.

Absent a direct instrument for SNF use, as opposed to an instrument that works through hospital assignment, we are unable to distinguish these views. But one way to approach this question is to assess whether our results vary by underlying measures of SNF quality. To the extent that SNF quality measures are informative, we would expect larger mortality effects at hospitals that use lower-quality SNFs if downstream care itself were harmful. In contrast, if the true signal comes from hospitals discharging patients in worse health such that they cannot return home, then the mortality effects would be seen regardless of the quality of the SNFs used.

Table 7 reports results when we use the rate at which the hospital sends other patients to different types of SNFs of varying quality measures. First, CMS issues quality rating for SNFs based on beneficiary health measures, where a five-star rating is the highest quality; Second, using the 100% Medicare sample we calculated a risk-standardized 30-day re-hospitalization rate for each SNF for our sample. Third, there are concerns that for-profit SNFs provide lower quality care compared to non-profit SNFs.

The results show only modest (and insignificant) effects of being treated at hospitals based on the types of SNFs used. Moreover, the estimated relationship between the 30-day SNF admission rate and mortality are generally unaffected by including controls for the types of SNFs used by the hospital. This suggests that our findings are driven by the discharge of sicker patients rather than harmful care at the SNF itself, but absent verification of the underlying SNF quality meaures this is only suggestive.

VII Conclusions

One of the key challenges facing health policy makers is how best to redesign hospital reimbursement systems to reflect provider quality and reduce costs. In this paper we have sought to overcome selection bias and characterize the relationship between hospitals spending profiles and patient outcomes. Our findings are consistent with previous evidence of low returns to area-wide, total spending differences, yet high returns to hospital treatment intensity: the resolution to this mystery is the unproductive role played by post-acute care. Higher spending on Skilled Nursing Facilities, in particular, appears to lead to both higher costs for Medicare and higher mortality rates for the elderly.

Our results provide somewhat subtle implications for spending reform efforts. On the one hand, overall 90 day spending does not seem closely related to patient outcomes, which suggests that bundled episode-based payments could be lowered relative to current fee for service levels. On the other hand, inpatient spending does appear highly productive, so lowering such bundled payments could penalize high performing hospitals if the reduced reimbursement is not properly allocated across the bundle participants (hospitals and post-acute care facilities). Alternatively, payment reforms could de-couple inpatient and outpatient spending over the 90 days after the patients episode to retain the accountability for downstream spending without curtailing treatment intensity at the initial hospital stay.

At a minimum, within the context of our current reimbursement system, the last section of our paper suggests that payers focus not just on hospital readmission rates but also on other patient outcomes after discharge. Penalizing hospitals as well for high post-acute spending, and in particular for high rates of SNF use, may deliver better outcomes while reducing Medicare spending.

Of course, a limitation is that the results directly apply only to serious, emergent conditions. A high priority for future work is to find different ways to address patient selection in extending this type of analysis to a broader set of diagnoses. Still, the types of conditions studied here are costly and represent prime candidates for the type of episode-based bundled payments currently under discussion in payment reform models.

While the results are consistent with positive returns to treatment intensity after emergencies in the hospital setting, and negative returns when relying on care outside the hospital setting, it is possible that spending patterns are associated with other factors at these hospitals, such as the quality of the physicians and nurses. Future research should test the returns different types of treatment intensity directly. For example, more randomized controlled trials that encourage the switch from SNF care to home health care, complete with monitoring to avoid subsequent hospitalizations, appear to present opportunities to improve health and lower costs.

The inpatient setting is seemingly more difficult to test; increases or decreases in intensity of treatment may have short run impacts which are difficult to predict from our framework. For example, expertise is likely essential to generate positive returns such that an increase in intensity for todays less-intensive hospitals may not result in gains as large as those implied here (Chandra and Staiger, 2007). A better understanding of this learning continues to be a fertile area for research with important implications for the evolution of hospital treatment that can improve health at sustainable costs.

VIII References

Ahmad, Faraz S., Joshua P. Metlay, Frances K. Barg, Rebecca R. Henderson, and Rachel M. Werner. 2013. "Identifying Hospital Organizational Strategies to Reduce Readmissions." American Journal of Medical Quality 28 (4): 278-85.

Asch, Steven. M., Kerr, Eve A., Keesey, Joan B.A., Adams, John L., Setodji, Claude M., Malik, Shaista, and Elizabeth A. McGlynn. 2006. "Who Is at Greatest Risk for Receiving Poor-Quality Health Care?" New England Journal of Medicine. 354: 1147-1156.

Ash, Arlene S., Stephen E. Fienberg, Thomas A. Louis, Sharon-Lise Normand, Therese Stukel, and Jessica Utts. 2012. "Statistical Issues in Assessing Hospital Performance." Centers for Medicare and Medicaid Services.

Barnato, Amber E, Chung-Chou H Chang, Max H Farrell, Judith R Lave, Mark S Roberts, and Derek C Angus. 2010. "Is Survival Better at Hospitals with Higher 'End-of-Life' Treatment Intensity?." Medical Care 48 (2): 125-32.

Berenson, Robert A., Ronald A. Paulus, and Noah S. Kalman. 2012. "Medicare's Readmissions-Reduction Program ? A Positive Alternative." New England Journal of Medicine 366 (15): 1364-66.

Black, N. 2010. "Assessing the Quality of Hospitals." BMJ 340 (apr19 2): c2066-c2066.

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

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2013. "Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector." NBER Working Paper Series, no. w19200. Cambridge, Mass: National Bureau of Economic Research.

Chandra, Amitabh, and Jonathan Skinner. 2011. Technology Growth and Expenditure Growth in Health Care. NBER Working Paper Series, no. w16953. Cambridge, Mass: National Bureau of Economic Research.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." Journal of Political Economy 115: 103-40.

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

Cutler, David. 2010. "How Health Care Reform Must Bend The Cost Curve." Health Affairs 29 (6): 1131-35.

Cutler, David M., and Kaushik Ghosh. 2012. "The Potential for Cost Savings through Bundled Episode Payments." New England Journal of Medicine 366 (12): 1075-77.

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

Doyle, Joseph J., John A. Graves, Jonathan Gruber, and Samuel A. Kleiner. 2015. "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns." Journal of Political Economy 123 (1): 170-214.

Fisher, Elliott S, Julie P Bynum, and Jonathan S Skinner. 2009. "Slowing the Growth of Health Care Costs–Lessons from Regional Variation." The New England Journal of Medicine 360 (9): 849-52. .

Fisher, Elliott S., David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003a. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." Annals of Internal Medicine 138 (4): 288-98.

Fisher, Elliott S., David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003b. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care." Annals of Internal Medicine 138 (4): 273-87.

Franzini, Luisa, Chapin White, Suthira Taychakhoonavudh, Rohan Parikh, Mark Zezza, and Osama Mikhail. 2014. "Variation in Inpatient Hospital Prices and Outpatient Service Quantities Drive Geographic Differences in Private Spending in Texas." Health Services Research, June, 1944-63.

Hawkes, N. 2010. "Patient Coding and the Ratings Game." BMJ 340 (apr23 2): c215-2153.

Institute of Medicine. 2010. Regionalizing Emergency Care. http://www.iom.edu/Reports/2010/Regionalizing-Emergency-Care-Workshop-Summary.aspx.

Johnson, Robin. 2001. "The Future of Local Emergency Medical Service: Ambulance Wars or Public-Private Truce?" Reason Foundation. http://reason.org/news/show/the-future-of-local-emergency.

Joynt, Karen E., and Ashish K. Jha. 2013a. Characteristics of Hospitals Receiving Penalties Under the Hospital Readmissions Reduction Program. JAMA 309: 342-343.

——. 2013b. "A Path Forward on Medicare Readmissions." New England Journal of Medicine 368 (13): 1175-77.

Kahn, Jeremy M., Charles C. Branas, C William Schwab, and David A. Asch. 2008. "Regionalization of Medical Critical Care: What Can We Learn from the Trauma Experience?*:" Critical Care Medicine 36 (11): 3085-88.

Keogh, Bruce. 2013. "Review into the Quality of Care and Treatment Provided by 14 Hosptial Trusts in England: Overview Report." National Health Service. http://www.nhs.uk/nhsengland/bruce-keogh-review/documents/outcomes/keogh-review-final-report.pdf.

Kolesar, Michal, Raj Chetty, John N. Friedman, Edward L. Glaeser, and Guido W. Imbens. 2011. "Identification and Inference with Many Invalid Instruments." National Bureau of Economic Research Working Paper Series No. 17519. http://www.nber.org/papers/w17519.

Kripalani, Sunil, Cecelia N. Theobald, Beth Anctil, and Eduard E. Vasilevskis. 2014. "Reducing Hospital Readmission Rates: Current Strategies and Future Directions." Annual Review of Medicine 65 (1): 471-85.

Krumholz, Harlan M. 2013. "Relationship Between Hospital Readmission and Mortality Rates for Patients Hospitalized With Acute Myocardial Infarction, Heart Failure, or Pneumonia¡alt-title¿Hospital Performance and Readmission/Mortality Rates¡/alt-Title¿." JAMA 309 (6): 587.

Lilford, R., and P. Pronovost. 2010. "Using Hospital Mortality Rates to Judge Hospital Performance: A Bad Idea That Just Won?t Go Away." BMJ 340 (apr19 2): c2016-c2016.

Mulcahy, Andrew, Katherine Harris, Kenneth Finegold, Arthur Kellermann, Laurel Edelman, and Benjamin D. Sommers. 2013. "Insurance Coverage of Emergency Care for Young Adults under Health Reform." New England Journal of Medicine 368 (22): 2105-12.

Naylor, Mary D., Linda H. Aiken, Ellen T. Kurtzman, Danielle M. Olds, and Karen B. Hirschman. 2011. "The Importance Of Transitional Care In Achieving Health Reform." Health Affairs 30 (4):

746-54.

Newhouse, Joseph P., and Alan M. Garber. 2013. "Geographic Variation in Health Care Spending in the United States: Insights From an Institute of Medicine Report." JAMA 310 (12): 1227.

Newhouse, Joseph P., Garber, Alan, and Robin P. Graham, eds. 2013. Interim Report of the Committee on Geographic Variation in Health Care Spending and Promotion of High-Value Health Care: Preliminary Committee Observations. The National Academies Press. http://www.nap.edu/openbook.php?record_id=18308.

OECD. 2014. Total Expenditure on Health per Capita 2014/1. Health: Key Tables from OECD. Paris. http://www.oecd-ilibrary.org/content/table/hlthxp-cap-table-2014-1-en.

Press, Matthew J., Dennis P. Scanlon, Andrew M. Ryan, Jingsan Zhu, Amol S. Navathe, Jessica N. Mittler, and Kevin G. Volpp. 2013. "Limits Of Readmission Rates In Measuring Hospital Quality Suggest The Need For Added Metrics." Health Affairs 32 (6): 1083-91.

Ragone, Michael. 2012. "Evolution or Revolution: EMS Industry Faces Difficult Changes." Journal of Emergency Medical Services 37(2).

Rubin, Donald. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." Health Services and Outcomes Research Methodology 2 (3-4): 169-88.

Shahian, David M., Gregg S. Meyer, Elizabeth Mort, Susan Atamian, Xiu Liu, Andrew S. Karson, Lawrence D. Ramunno, and Hui Zheng. 2012. "Association of National Hospital Quality Measure Adherence with Long-Term Mortality and Readmissions." BMJ Quality & Safety 21 (4): 325-36.

Shahian, David M., Robert E. Wolf, Lisa I. Iezzoni, Leslie Kirle, and Sharon-Lise T. Normand. 2010. "Variability in the Measurement of Hospital-Wide Mortality Rates." New England Journal of Medicine 363 (26): 2530-39.

Skinner, Jonathan, and Fisher, Elliott S. 2010. "Reflections on Geographic Variation in U.S. Health Care." The Dartmouth Institute. http://www.dartmouthatlas.org/downloads/press/Skinner_Fisher_DA_05_10.pdf.

Skura, Barry. 2001. “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.

Song, Yunjie, Skinner, Jonathan, Bynum, Julie, Sutherland, Jason, Wennberg, John E., and Elliott S. Fisher. 2010. “Regional variations in diagnostic practices.” New England Journal of Medicine. 363:45-53.

Stock, James H, Jonathan H Wright, and Motohiro Yogo. 2002. “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments.” Journal of Business and Economic Statistics 20 (4): 518-29.

Werner, R. M. and D.A. Asch. 2005. “The Unintended Consequences of Publicly Reporting Quality Information.” JAMA 293: , 1239-1244.

Table 1—Sample Characteristics and Balance

	Mean	Standard Deviation	Standardized Difference $1(Instrument > Median)$
Age	81.614	7.596	−0.029
Male	0.376	0.484	−0.002
Race: Black	0.073	0.264	0.017
Race: Other	0.031	0.174	0.006
Hypertension	0.409	0.490	−0.028
Stroke	0.068	0.253	0.004
Cerebrovascular Disease	0.101	0.300	−0.004
Renal Failure Disease	0.155	0.353	−0.053
Dialysis	0.015	0.122	−0.001
COPD	0.212	0.408	−0.008
Pneumonia	0.191	0.391	−0.015
Diabetes	0.196	0.396	−0.009
Protein Calorie Malnutrition	0.070	0.252	−0.014
Dementia	0.176	0.381	0.000
Paralysis	0.085	0.278	−0.002
Peripheral Vascular Disease	0.131	0.336	−0.010
Metastatic Cancer	0.034	0.181	0.000
Trauma	0.137	0.342	−0.012
Substance Abuse	0.067	0.247	−0.014
Major Psych. Disorder	0.056	0.229	−0.004
Chronic Liver Disease	0.010	0.097	0.000
Ambulance: Miles Traveled with Patient	7.051	8.216	0.006
Ambulance: Advanced Life Support	0.688	0.467	−0.041
Ambulance: Emergency Transport	0.912	0.293	−0.048
Ambulance: Outpatient File	0.072	0.242	−0.074
Ambulance: Payment	340.510	192.316	−0.108

Note: N=1,582,421. Balance statistics report the standardized difference based on the the difference in the mean of the covariate when the instrument is above versus below its median value computed from a regression model that controls for year and ZIP code × patient origin fixed effects, divided by the pooled standard deviation. Average age reported here, however in all regression models age controls are included as dummy variables for 5 year age bins starting at age 66.
Source: 2002-2012 Medicare Part A and B Data

Table 2—One Year Mortality: First Stage, OLS and 2SLS Estimates, by Risk-Standardized 90D Hospital Spending Measure

		(1)	(2)
Panel A. First Stage			
	Ambulance Average Total 90D Spending	0.185	0.185
		(0.0043)	(0.0043)
Panel B. OLS and 2SLS			
	OLS: Hospital Average Total 90D Spending	-0.018	-0.011
		(0.0025)	(0.0023)
	2SLS: Hospital Average Total 90D Spending	-0.017	-0.016
		(0.0120)	(0.0113)
	Sample Size	1,582,421	1,582,421
	Outcome Mean	0.426	0.426
	Patient Controls	No	Yes

Note: Within each panel and regression type (OLS and 2SLS), each column reports model results based on hospital measures of total spending over 90 days after the index admission. Spending measures are trimmed of outliers and then risk-standardized by age, race and gender; sub-totals may not add to total due to outlier adjustment in the risk-standardization process. Risk-standardized spending measures have been demeaned and scaled by 2 standard deviations. Thus, the reported coefficients reflect a difference of ± 1 standard deviations from the mean (i.e., “low” vs. “high” spending). Means (SDs) for spending variables: 90D Total = \$27,351 (4,243). Models include all patient and ambulance controls listed in Table 1, ZIP code $\times$ patient origin fixed effects, year fixed effects, and primary diagnosis fixed effects (see Table A1 for a full list of categories). Standard errors, clustered at the Health Service Area (HSA) level, are reported in parentheses.
Source: 2002-2012 Medicare Part A and B Data

Table 3—One Year Mortality: First Stage, OLS and 2SLS Estimates, by Risk-Standardized 90D Hospital Spending Measure

		90D Inpatient Spending		**90D Non-Inpatient Spending**	
		(1)	(2)	(3)	(4)
Panel A. First Stage					
	Ambulance Average 90D Inpatient Spending	0.248	0.248		
		(0.0056)	(0.0057)		
	Ambulance Average 90D Non-Inpatient Spending			0.133	0.133
				(0.0049)	(0.0049)
Panel B. OLS					
	90D Inpatient Spending	-0.018	-0.011		
		(0.0023)	(0.0020)		
	90D Non-Inpatient Spending			0.003	-0.001
				(0.0026)	(0.0025)
Panel C. 2SLS					
	90D Inpatient Spending	-0.047	-0.042		
		(0.0100)	(0.0092)		
	90D Non-Inpatient Spending			0.063	0.054
				(0.0153)	(0.0136)
	Sample Size	1,582,421	1,582,421	1,582,421	1,582,421
	Outcome Mean	0.426	0.426	0.426	0.426
	Patient Controls	No	Yes	No	Yes

Note: Each column reports model results based on 90-day measures based on spending in inpatient and non-inpatient settings. All spending measures are trimmed of outliers and then risk-standardized by age, race and gender; sub-totals may not add to total due to outlier adjustment in the risk-standardization process. Risk-standardized spending measures have been demeaned and scaled by 2 standard deviations. Thus, the reported coefficients reflect a difference of ± 1 standard deviations from the mean (i.e., "low" vs. "high" spending). Means (SDs) for spending variables: 90D Inpatient Total = \$15,876 (3,113); 90D Non-Inpatient Total = \$10,557 (1,585). Models include all patient and ambulance controls listed in Table 1, ZIP code $\times$ patient origin fixed effects, year fixed effects, and primary diagnosis fixed effects (see Table A1 for a full list of categories). Standard errors, clustered at the Health Service Area (HSA) level, are reported in parentheses.
Source: 2002-2012 Medicare Part A and B Data

Table 4—365D Mortality: OLS and 2SLS Estimates by 90D Spending Measure

Outcome: 365D Mortality		(1)	(2)	(3)
Panel A. OLS				
	90D Inpatient Spending	-0.011		-0.009
		(0.0020)		(0.0020)
	90D Non-Inpatient Spending	-0.000		
		(0.0025)		
	90D SNF Spending		0.008	0.008
			(0.0023)	(0.0023)
	90D Non-SNF Spending		-0.015	-0.014
			(0.0023)	(0.0023)
Panel B. 2SLS				
	90D Inpatient Spending	-0.050		-0.024
		(0.0093)		(0.0102)
	90D Non-Inpatient Spending	0.060		
		(0.0139)		
	90D SNF Spending		0.095	0.095
			(0.0134)	(0.0136)
	90D Non-SNF Spending		-0.033	-0.028
			(0.0118)	(0.0120)

Note: N=1,582,421. Outcome Mean=0.426. Columns (1) and (3) report model results based on measures of spending that, when summed across types, equal total spending over 90 days beginning with the index admission. Column (2) reports on spending measures that comprise the 90D Non-Inpatient spending measure. All spending measures are trimmed of outliers and then risk-standardized by age, race and gender; sub-totals may not add to total due to outlier adjustment in the risk-standardization process. Risk-standardized spending measures have been demeaned and scaled by 2 standard deviations. Thus, coefficients reflect a difference of ± 1 standard deviations from the mean (i.e., "low" vs. "high" spending). Models include all patient and ambulance controls listed in Table 1, ZIP code $\times$ patient origin fixed effects, year fixed effects, and primary diagnosis fixed effects (see Table A1 for a full list of categories). Standard errors, clustered at the Health Service Area (HSA) level, are reported in parentheses.
Source: 2002-2012 Medicare Part A and B Data

Table 5—SNF Admission Rate: 365D Mortality and Total Spending Outcomes

		Outcome: 365D Mortality		**Outcome: log(365D Spending)**	
		(1)	(2)	(3)	(4)
Panel A. OLS					
	30D SNF Admission Rate	0.011	0.009	-0.031	0.007
		(0.0018)	(0.0019)	(0.0038)	(0.0034)
	Index Hospital Spending		-0.006		0.135
			(0.0019)		(0.0037)
Panel B. 2SLS					
	30D SNF Admission Rate	0.031	0.023	-0.041	0.014
		(0.0058)	(0.0068)	(0.0104)	(0.0118)
	Index Hospital Spending		-0.025		0.163
			(0.0100)		(0.0171)

Note: N=1,582,421. Average 365D spending=$44,857. 365D mortality mean =0.426. Each column reports model results based on a separate regression with full controls. Controls include all patient and ambulance measures listed in Table 1, ZIP code × patient origin fixed effects, year fixed effects, and primary diagnosis fixed effects (see Table A1 for a full list of categories). All hospital measures are risk-standardized by age, race and gender. These risk-standardized measures are then demeaned and scaled by 2 standard deviations. Thus, coefficients reflect a difference of ±1 standard deviations from the mean (i.e., "low" vs. "high" hospitals). Standard errors, clustered at the Health Service Area (HSA) level, are reported in parentheses.

Source: 2002-2012 Medicare Part A and B Data

Table 6—Risk-Standardized Quality Measures: 365D Mortality and Total Spending Outcomes

		Outcome: 365D Mortality				Outcome: log(365D Spending)			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. OLS									
	30D SNF Admission Rate	0.011				-0.031			
		(0.0018)				(0.0038)			
	30D Readmission Rate		0.006				0.047		
			(0.0018)				(0.0039)		
	30D Mortality Rate			0.033				-0.036	
				(0.0015)				(0.0037)	
	30D Home and Alive Rate				-0.020				0.029
					(0.0018)				(0.0033)
Panel B. 2SLS									
	30D SNF Admission Rate	0.031				-0.041			
		(0.0058)				(0.0104)			
	30D Readmission Rate		0.014				0.053		
			(0.0056)				(0.0101)		
	30D Mortality Rate			0.054				-0.069	
				(0.0052)				(0.0098)	
	30D Home and Alive Rate				-0.042				0.041
					(0.0055)				(0.0097)
Panel C. 2SLS									
	Index Hospital Spending	-0.025	-0.033	-0.035	-0.022	0.163	0.162	0.160	0.158
		(0.0100)	(0.0089)	(0.0093)	(0.0098)	(0.0171)	(0.0153)	(0.0156)	(0.0165)
	30D SNF Admission Rate	0.023				0.014			
		(0.0068)				(0.0118)			
	30D Readmission Rate		0.012				0.061		
			(0.0055)				(0.0098)		
	30D Mortality Rate			0.056				-0.074	
				(0.0053)				(0.0099)	
	30D Home and Alive Rate				-0.036				-0.002
					(0.0062)				(0.0105)

Note: N=1,582,421. Average 365D Spending=$44,857. Each column reports model results based on a separate regression with full controls. Controls include all patient and ambulance measures listed in Table 1, ZIP code × patient origin fixed effects, year fixed effects, and primary diagnosis fixed effects (see Table A1 for a full list of categories). All hospital measures are risk-standardized by age, race and gender. These risk-standardized measures are then demeaned and scaled by 2 standard deviations. Thus, coefficients reflect a difference of ±1 standard deviations from the mean (i.e., "low" vs. "high" hospitals). "Home and Alive" corresponds to being alive, not readmitted, and not admitted to a skilled nursing facility. Index Hospital Spending is risk-standardized Medicare spending per beneficiary for the index admission. Standard errors, clustered at the Health Service Area (HSA) level, are reported in parentheses.
Source: 2002-2012 Medicare Part A and B Data

Table 7—2SLS Results: SNF Quality Measures

Outcome: 365D Mortality	(1)	(2)	(3)	(4)	(5)
30D SNF Admission Rate	0.031	0.031	0.033	0.030	0.033
	(0.0058)	(0.0058)	(0.0058)	(0.0061)	(0.0058)
Five Star Rating		0.001			
		(0.0047)			
Rehospitalization Rate: Highest Quintile			-0.006		
			(0.0060)		
For-Profit				0.003	
				(0.0053)	
Non-Profit					-0.011
					(0.0052)

Note: Notes: N=1,582,421. Each column reports model results based on a separate regression with full controls. Controls include all patient and ambulance measures listed in Table 1, ZIP code × patient origin fixed effects, year fixed effects, and primary diagnosis fixed effects (see Table A1 for a full list of categories). All hospital measures are risk-standardized by age, race and gender. These risk-standardized measures are then demeaned and scaled by 2 standard deviations. Thus, coefficients reflect a difference of ±1 standard deviations from the mean (i.e., "low" vs. "high" hospitals). Standard errors, clustered at the Health Service Area (HSA) level, are reported in parentheses.
Source: 2002-2012 Medicare Part A and B Data

Supplemental (Online) Appendix

Table A1—Sample Characteristics and Balance: Principal Diagnosis

	Mean	Standardized Difference $1(Instrument > Median)$
038 Septicemia	0.129	−0.029
162 Malignant neoplasm of trachea, bronchus, and lung	0.008	0.002
197 Secondary malignant neoplasm of respiratory and digestive systems	0.005	0.009
410 Acute myocardial infarction	0.077	0.020
431 Intracerebral hemorrhage	0.011	0.010
433 Occlusion and stenosis of precerebral arteries	0.008	0.004
434 Occlusion of cerebral arteries	0.070	−0.002
435 Transient cerebral ischemia	0.027	0.004
482 Other bacterial pneumonia	0.017	0.003
486 Pneumonia, organism unspecified	0.128	0.007
507 Pneumonitis due to solids and liquids	0.048	−0.002
518 Other diseases of lung	0.057	−0.013
530 Diseases of esophagus	0.011	0.006
531 Gastric ulcer	0.009	0.010
532 Duodenal ulcer	0.007	0.004
557 Vascular insufficiency of intestine	0.007	−0.002
558 Other and unspecified noninfectious gastroenteritis and colitis	0.008	−0.001
560 Intestinal obstruction without mention of hernia	0.024	−0.002
599 Other disorders of urethra and urinary tract	0.084	−0.012
728 Disorders of muscle, ligament, and fascia	0.006	−0.012
780 General symptoms	0.080	0.013
807 Fracture of rib(s), sternum, larynx, and trachea	0.005	−0.001
808 Fracture of pelvis	0.014	−0.001
820 Fracture of neck of femur	0.112	0.007
823 Fracture of tibia and fibula	0.004	0.003
824 Fracture of ankle	0.008	0.001
959 Injury, other and unspecified	0.002	0.001
965 Poisoning by analgesics, antipyretics, and antirheumatics	0.002	−0.008
969 Poisoning by psychotropic agents	0.002	−0.002
Other Non-Discretionary†	0.030	−0.002

Note: N=1,582,421. Balance statistics report the standardized difference (i.e.,the difference in means divided by the pooled standard deviation) based on splitting the sample on whether patients were admitted to a hospital profiled above or below the median for the total 90D risk-standardized spending measure. †Based on diagnoses used in Mulcahy, et al. *N Engl J Med* 2013; 368:2105-2112.
Source: 2002-2012 Medicare Part A and B Data

Table A2—Sample Characteristics: Hospital Measures

	Mean	Standard Deviation
Total 90D Spending	27,351	4,243
Inpatient Spending	15,876	3,113
Non-Inpatient Spending	10,557	1,585
SNF Spending	5,164	1,114
Non-SNF Spending	4,775	974
30D SNF Admission Rate Rate	45.8	8.4
Five-Star Discharge Rate	4.5	5.3
Rehospitalization Rate: Top Quintile	5.2	4.5
For-Profit Discharge Rate	21.0	9.2
Non-Profit Discharge Rate	7.4	7.5
Government Facility Discharge Rate	2.4	4.1
30D Readmission Rate	14.3	2.7
30D Mortality Rate	18.2	2.6
30D Home and Alive Rate	34.9	6.7

Note: N=1,582,421. All spending measures have been risk-standardized by age, race and gender. "Home and Alive" corresponds to being alive, not readmitted, and not admitted to a skilled nursing facility.
Source: 2002-2012 Medicare Part A and B Data

Table A3—First Stage Estimates: 90D Spending Measures

Hospital Spending Measure:	**(1)**		**(2)**		**(3)**		
	Inpatient	Non-Inpatient	SNF	Non-SNF	Inpatient	SNF	Non-SNF
Amb.: 90D Inpatient	0.264	-0.015			0.262	-0.044	0.019
	(0.0060)	(0.0038)			(0.0059)	(0.0050)	(0.0047)
Amb.: 90D Non-Inpatient	-0.044	0.136					
	(0.0033)	(0.0052)					
Amb.: 90D SNF			0.138	-0.011	-0.031	0.141	-0.012
			(0.0041)	(0.0027)	(0.0032)	(0.0042)	(0.0028)
Amb.: 90D Non-SNF			-0.022	0.159	-0.030	-0.011	0.154
			(0.0030)	(0.0048)	(0.0029)	(0.0032)	(0.0050)

Note: N=1,582,421. Within each panel, each column reports first stage results based on various measures of spending. Total 90-day spending is decomposed into 90-day inpatient and 90-day non-inpatient spending; total 90-day non-inpatient spending is further decomposed into 90-day SNF and 90-day non-SNF spending. All spending measures are trimmed of outliers and then risk-standardized by age, race and gender. These risk-standardized measures are then demeaned and scaled by 2 standard deviations. Thus, coefficients reflect a difference of ± 1 standard deviations from the mean (i.e., "low" vs. "high" spending). All models includes full controls. Controls include all patient and ambulance measures listed in Table 1, ZIP code $\times$ patient origin fixed effects, year fixed effects, and primary diagnosis fixed effects (see Table A1 for a full list of categories). Standard errors, clustered at the Health Service Area (HSA) level, are reported in parentheses.
Source: 2002-2012 Medicare Part A and B Data

Table A4—First Stage Estimates: SNF Admission Rate and Index Spending

Hospital Measure:	**(1)**	**(2)**	
	Hospital 30D SNF	Hospital 30D SNF	Inpatient Spending
Ambulance Avg. 30D SNF Admission Rate	0.464	0.244	-0.012
	(0.0077)	(0.0053)	(0.0046)
Ambulance Avg. Index Hospital Spending		-0.098	0.461
		(0.0057)	(0.0078)

Note: N=1,582,421. Each column reports first stage results based on a separate regression with full controls. Controls include all patient and ambulance measures listed in Table 1, ZIP code × patient origin fixed effects, year fixed effects, and primary diagnosis fixed effects (see Table A1 for a full list of categories). All measures are risk-standardized by age, race and gender. These risk-standardized measures are then demeaned and scaled by 2 standard deviations. Thus, coefficients reflect a difference of ±1 standard deviations from the mean (i.e., "low" vs. "high" hospitals). Standard errors, clustered at the Health Service Area (HSA) level, are reported in parentheses.
Source: 2002-2012 Medicare Part A and B Data

Table A5—First Stage Estimates: Risk-Standardized Quality Measures

Hospital Spending Measure:	**(1)**		**(2)**		**(3)**		**(4)**	
	Inpatient	30D-SNF	Index	30D-R	Index	30D-M	Index	30D-H&A
Panel A.								
Amb. 30D SNF Admission Rate	0.464							
	(0.0077)							
Amb. 30D Readmission Rate			0.480					
			(0.0085)					
Amb. 30D Mortality Rate					0.476			
					(0.0073)			
Amb. 30D Home and Alive Rate							0.445	
							(0.0075)	
Panel B.								
Amb. Avg. Index Spending	0.244	-0.012	0.260	-0.010	0.260	0.000	0.249	0.012
	(0.0053)	(0.0046)	(0.0052)	(0.0043)	(0.0052)	(0.0054)	(0.0052)	(0.0055)
Amb. 30D SNF Admission Rate	-0.098	0.461						
	(0.0057)	(0.0078)						
Amb. 30D Readmission Rate			-0.023	0.479				
			(0.0049)	(0.0085)				
Amb. 30D Mortality Rate					0.014	0.475		
					(0.0048)	(0.0073)		
Amb. 30D Home and Alive Rate							0.074	0.442
							(0.0049)	(0.0075)

Note: N=1,582,421. Each column reports first stage results based on a separate regression with full controls. Controls include all patient and ambulance measures listed in Table 1, ZIP code × patient origin fixed effects, year fixed effects, and primary diagnosis fixed effects (see Table A1 for a full list of categories). All measures are risk-standardized by age, race and gender. These risk-standardized measures are then demeaned and scaled by 2 standard deviations. Thus, coefficients reflect a difference of ± 1 standard deviations from the mean (i.e., "low" vs. "high" hospitals). 30D-R = 30-day readmission rate; 30D-M = 30-day mortality rate; 30D-H&A = 30-day home and alive rate. "Home and Alive" corresponds to being alive, not readmitted, and not admitted to a skilled nursing facility. Standard errors, clustered at the Health Service Area (HSA) level, are reported in parentheses.
Source: 2002-2012 Medicare Part A and B Data