Paying for Quality in Healthcare

Joseph Doyle — *MIT and NBER*
John A. Graves — *Vanderbilt University*
Jonathan Gruber [‡] — *MIT and NBER*

March 2014

Abstract

Quality measures are playing an increasingly large role in the reimbursement of medical providers in the U.S., despite concerns that quality measures may be confounded by patient selection. This paper aims to estimate the causal relationship between measured hospital quality, reimbursement, and patient outcomes. To compare similar patients across hospitals in the same market, we exploit ambulance company preferences as an instrument for patient assignment. Our primary measure of hospital quality is the hospital's risk-adjusted mortality rate over the prior three years, and we find that this measure is significantly related to subsequent mortality, as well as hospital readmissions. Nevertheless, we find that high-reimbursement hospitals achieve lower mortality rates even conditional on these quality measures, suggesting that "reference pricing"—paying hospitals at rates set by the lowest-cost provider within quality groupings—may have negative implications for patient health. Meanwhile, hospitals that rely on greater levels of post-discharge treatment outside of the hospital setting have higher mortality rates, which is consistent with the current sentiment that efforts to coordinate care outside of the hospital hold the potential to lower costs and improve patient health.

This work is preliminary; Comments Welcome

[‡] 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 1200 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. Acknowledgements: We are grateful to Melinda Buntin, Kitt Carpenter, and Doug Staiger for helpful conversations. We gratefully acknowledge support from the National Institutes of Health R01 AG41794-01.

I. Introduction

Sustained growth in health care expenditures has prompted considerable interest in improving the quality and efficiency of medical care in the United States. This interest is motivated by influential research demonstrating widespread geographic variation in treatment intensity that yields little apparent benefit in terms of patient health outcomes (Fisher, Bynum, and Skinner 2009; Fisher et al. 2003a; Fisher et al. 2003b; Chandra and Skinner 2011). At the same time, a parallel body of research has found consistent gaps between the quality of care patients receive and what the medical system could provide if it were productively efficient and operating at its full potential (Chandra and Staiger 2007; Committee on Quality of Health Care in America 2001; McGlynn et al. 2003).

The contention that the U.S. health care system simultaneously provides too much low value care and too little high quality care lies at the heart of many delivery system reform initiatives. These initiatives have generally focused their efforts in two directions. The first is to create more direct linkages between reimbursement and measures of provider quality. The Medicare Hospital Readmission Reduction Program (HRRP), for example, penalizes hospitals with above-average 30-day readmission rates among acute myocardial infarction (AMI), chronic heart failure (CHF) and pneumonia (PNA) patients (Berenson, Paulus, and Kalman 2012). Another example is "reference pricing," whereby hospitals of equal observed quality are reimbursed by the same amount, and patients are held responsible for any additional charges above the reference price (Robinson and MacPherson 2012).

A second approach is to move away from service-specific reimbursement and towards bundled reimbursement of a package of services. For example, the Affordable Care Act promotes the formation of Accountable Care Organizations (ACOs) which aim to coordinate care across different types of providers. The ultimate objective of ACOs and other bundled payment models is to structure reimbursement around a single capitated amount for either episodes of care or even for all care for a given patient. This approach is distinct from traditional managed care in that it integrates the first approach by targeting payments to quality of care; for example, shared savings payments to Pioneer Accountable Care Organizations (ACOs) are based on provider performance on a set of 33 quality measures (Boyarsky and Parke 2012). Moreover, the earlier results showing that there are low returns to medical care spending across areas suggest that bundled reimbursements may provide a mechanism for lowering health care spending and allowing healthcare providers to decide on where to achieve any savings.

The goal of this paper is to provide a framework for thinking about these issues within the context of hospital reimbursement. A major challenge to doing so is that to create a level playing field for reimbursement based on quality, performance measurements must account for differences in patient acuity across hospitals. This challenge is often difficult to overcome because patients are likely to seek out high quality providers in ways that are unobserved in any administrative database used to measure quality. We therefore develop a framework based on earlier work that aims to purge patient selection by leveraging ambulance company referral patterns as an instrument for hospital assignment (Doyle et al. 2012). We show that ambulance-company assignment is remarkably unrelated to observed patient characteristics and that this assignment affects hospital choice. Thus, our approach provides a compelling lens through which we can evaluate hospital performance, at least for the roughly one-quarter of hospital admissions that arrive via ambulance.

Within this framework, we proceed in three steps. The first considers the fundamental question of whether hospital choice affects survival. We characterize hospitals by their risk-adjusted mortality rates over the prior three years. This measure is commonly used, but it is also controversial due to concerns that risk adjustment fails to fully correct for patient selection. Using our approach that aims to compare similar patients across hospitals, hospitals that score well on this measure have significantly lower mortality and readmission rates among subsequent patients. Indeed, this measure outperforms other common quality measures, including the hospital's readmission rate and measures of the hospital's use of best practices.

We next consider the appropriateness of hospital reference pricing and other quality-based payment reforms. Once selection is corrected and quality is reliably estimated, if patient outcomes are comparable across hospitals of a fixed quality level then there is no reason to pay more to more intensive (and expensive) hospitals. We examine the advisability of reference pricing by estimating outcome regressions as a function of both quality measures and reimbursement levels. Here, we find that conditional on measured quality, higher-reimbursement hospitals have significantly higher survival rates among subsequent patients. This indicates that approaches which reimburse inpatient care solely based on existing quality measures to reign in high-reimbursement hospitals could significantly worsen patient outcomes for this population of emergency hospital patients.

Finally, we find that while there is a positive association between inpatient reimbursement and survival, there is also a *negative* association between measures of post-discharge reimbursement and survival. On balance, the returns to inpatient and post-discharge care offset each other, resulting in a "flat-of-the-curve" finding of no association between total (inpatient and outpatient) Medicare spending and patient outcomes. This is consistent with the larger literature showing little return to area spending differences when inpatient and outpatient care are included in a summary spending measure. Moreover, it is also consistent with recent evidence from the Institute of Medicine identifying post-acute care as the major contributing factor to regional variation in Medicare spending per beneficiary (Joseph P. Newhouse, Alan Garber, and Robin P. Graham 2013; Newhouse and Garber 2013).

Our results provide somewhat subtle implications for attempts to bundle payments for patient care. On the one hand, if the payment is set low to reduce hospital reimbursement, this could inappropriately penalize high-performing hospitals that are improving patient outcomes. On the other hand, our results point to the value of tying hospital reimbursement to downstream treatment, which may depend on hospital decisions and may benefit from greater coordination and care management.

The remainder of this paper proceeds as follows. Section II provides relevant institutional background on hospital quality reporting in the U.S., as well as a review of the literature on the relationship between measured quality and patient outcomes; Section III presents the empirical framework; Section IV discusses the data sources and quality measures we examine. Section V presents the main results. Our first section of empirical results considers hospital quality measures. We then turn to introducing inpatient reimbursement into the outcome regressions, showing that there are sizeable returns to reimbursement levels even conditional on quality measures. We then show that there are offsetting effects of post-acute reimbursement outside the hospital setting. Section VI concludes with a discussion of the implications of our findings for hospital reimbursement policy.

II. Background: Measuring Hospital Quality

Measures of hospital quality have historically been defined across two primary dimensions – process measures and outcome measures -- though in recent years survey-based measures of patient experience have also gained favor. In addition, structural attributes of hospitals, such as teaching status and volume, also have a long history in the hospital literature. These measures are often criticized due to concerns that hospitals treat different patients, and risk adjustment does not fully control for these differences. We discuss the relevant details of these measurement approaches below.

II.1 Outcome-based Measures

We are primarily interested in whether hospital choice matters for patient mortality, and we begin by comparing hospitals that that have been shown to vary significantly in their risk-adjusted mortality rate (Chandra et al. 2013). Along with readmission rates, these outcome-based measures are controversial due to concerns that risk-adjustment does not fully control for differences in patients across hospitals. We aim to understand whether these measures are predictive of subsequent outcomes when we compare similar patients who happen to be treated at different hospitals. This is analogous to studying whether commonly used teacher-quality measures are related to long-term outcomes, even if there are concerns that the measures themselves are problematic (Chetty et al. 2012).

II.1.1 Current Methodology

Outcome-based quality measures compare the predicted or observed number of patients who experience a given outcome (e.g., death or readmission within a 30 days of discharge) to the number expected to experience the outcome based on a national risk model (Ash et al. 2012). That is, these measures ask: for a given hospital, is the number of patients who experience the outcome consistent with what would be expected in a hypothetical hospital of average quality and that has the same patient case mix?

This "indirect standardization" approach is used to construct hospital report cards by the Centers for Medicare and Medicaid Services (CMS) and by other organizations such as U.S. News and World Report. Typically, the basis for these measures is a logistic regression model that includes patient-level measures of clinical acuity (e.g., past diagnoses and comorbidities as recorded in billing claims), demographics (e.g., age, gender) and a hospital-specific intercept that is assumed to be drawn from a known probability distribution. Some important patient-level attributes (e.g., race, ethnicity and socio-economic status) are deliberately excluded from risk adjustment so that the risk-standardized measures do not condition out important racial or socio-economic disparities in care across facilities.[1]

More formally, define Y_{ih} as a binary indicator of whether patient *i* treated at hospital *h* experienced the outcome, and define $\mathbf{x_{ih}}$ as a vector of patient-level characteristics. The current CMS model assumes the following:

$$(1) \qquad Y_{ih}|\alpha_h, \beta, \mathbf{x_{ih}} \overset{ind}{\sim} \text{Bernoulli}(\alpha_h + \beta\mathbf{x_{ih}})$$

$$(2) \qquad \alpha_h|\mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$$

Based on this model, the risk-standardized rate for hospital *h* (RSR_h) is estimated as

$$(3) \qquad RSR_h(\mathbf{x_{ih}}) = \frac{\sum_{i=1}^{n_h} E(Y_{ih}|\alpha_h, \mu, \beta, \mathbf{x_{ih}}, \sigma^2)}{\sum_{i=1}^{n_h} E(Y_{ih}|\mu, \beta, \mathbf{x_{ih}}, \sigma^2)}\bar{Y}$$

where $\bar{Y}$ is the national average outcome and n_h is the number of patients treated in hospital *h* (Ash et al. 2012). This hierarchical modeling approach (equations 1 and 2) both accounts for clustering of patients within hospitals and also "shrinks" estimates of α_h towards $\hat{\mu}$ in cases where there are insufficient observations to reliably estimate the random effects (Ash et al. 2012).

In practice, equation (1) is estimated as a hierarchical Bayes logistic regression model that includes a hospital random effect (α_h). The underlying patient-level data and the estimated model parameters ($\hat{\alpha}_h, \hat{\mu}, \hat{\beta}, \widehat{\sigma^2}$) are used to construct predicted values, which are then fed into equation (3) to obtain the risk-standardized rate for each hospital.[2]

---

[1] Sensitivity analyses based on comparing outcome rates to hospitals that treat large numbers of Medicaid patients and African American patients yield a similar range of performance relative to other hospitals. See, for example https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/MedicareHospitalQualityChartbook2012.pdf (pp. 23-36).

[2] The numerator in equation (3) is simply the sum of predicted values for each patient in the hospital, where the predicted values are based on observed patient values, $\hat{\mu}, \hat{\beta}$, and the estimated hospital random effect ($\hat{\alpha}_h$). The denominator is similarly estimated as the sum of patient-level predictions, but only $\hat{\mu}$ -- the estimated national mean of $\hat{\alpha}_h$ -- is used. This measure is then multiplied by $\bar{Y}$ to place the risk standardized measure on the same scale as

A bootstrap procedure is used to produce a 95% confidence interval around RSR_h. The CMS Hospital Compare program classifies hospitals as "No Different than the U.S. National Rate" if the bootstrapped 95% confidence interval includes $\bar{Y}$. Notably, the current CMS readmissions penalties are simply based on whether a given hospital's point estimate is above $\bar{Y}$ (Epstein, Jha, and Orav 2011). Other approaches to hospital profiling simply use quantile rankings (e.g., top 10% of hospitals).

A related approach used in the health economics literature and some CMS demonstration programs employs a hospital fixed effect to profile outcome quality.[3] This essentially amounts to replacing the fixed variance term (σ^2) in equation (2) with an infinite variance (Gelman and Hill 2007). The fixed effects approach relies less on the functional form assumptions of the random effects, but tend to be less precise. One approach to rate stabilization in this context – which we incorporate in our fixed-effects based approach outlined below – is to utilize an empirical Bayes approach to "shrink" the fixed effects towards the national mean before they are fed into the risk standardization estimator (Equation 3).

II.1.2 Reliability Concerns

One concern is that these shrinkage models assume that the patient-level risk adjustors capture the relevant clinical characteristics such that hospital level attributes (e.g., patient volume, teaching status, etc.) are not independent predictors of patient outcomes.[4] This is a strong assumption, particularly in light of the large medical and health services literature linking hospital attributes like volume to improved patient outcomes (Birkmeyer et al. 2002; Daley 2002; Dudley et al. 2000; Halm, Lee, and Chassin 2002; Hughes, Hunt, and Luft 1987; Luft, Hunt, and Maerki 1987; Shahian and Normand 2003); the relatively limited number of patient characteristics available in billing data; the deliberate exclusion from the model (on substantive

---

the outcome. In effect, outcome-based quality assessments are determined primarily by how much each hospital's estimated random effect ($\hat{\alpha}_h$) deviates from $\hat{\mu}$. The hypothetical reference hospital in the denominator of equation (3) has the same patient case-mix as the hospital in question, however it has average quality, as reflected by the use of only $\hat{\mu}$ to summarize hospital-level factors. In this way each hospital is evaluated against a hypothetical reference hospital with an identical case-mix and with average quality.

[3] For example, see http://www.dialysisreports.org/pdf/esrd/public/SMRdocumentation.pdf

[4] A possible middle-ground in which risk-standardized rates are stabilized by hospital-level attributes has been advocated by some researchers (Dimick et al. 2012; Ryan et al. 2012). In other words, shrinkage occurs towards a hospital model rather than towards the national average. A common approach to a hospital-level model is motivated by the volume-outcome relationship, and simply involves the addition of hospital volume quintiles as shrinkage targets.

grounds) of certain important confounders like race and socio-economic status; and growing evidence on the endogeneity of patient-level diagnoses as recorded hospital billing codes (Song et al. 2010). Given these limitations, it is important to know whether these measures are actually predictive of subsequent mortality using plausibly exogenous variation in hospital assignment.

II.2 Process Measures

Process quality measures quantify the rate at which hospitals provide activities of care with sufficient clinical evidence linking that care to improved patient outcomes. The percentage of AMI patients administered aspirin upon arrival, for example, has long been used to assess whether hospitals regularly incorporate high-value, evidence-based care.

The number of process measures used in hospital report cards has grown considerably in recent years. These measures are a key component of the Centers for Medicare and Medicaid Service's (CMS) and National Quality Forum's Hospital Compare program, the Leapfrog group, and also serve as an important input into U.S. News and World Report's annual hospital rankings. The number of process measures reported on Hospital Compare, for example, has risen from 20 in 2005 (the first year of public reporting) to over 40 by 2014.

While the number of process quality measures has increased over time, so too has observed hospital performance on these measures. For example, Figure 1 plots the distribution in each year of a composite process measure constructed from data reported on the Hospital Compare website between 2005 and 2012. This figure summarizes, for a consistent set of 3,027 hospitals, how their performance on a fixed set of 13 process measures evolved over an eight-year period. As can be seen in the figure, in the early years of public reporting there was wide variation in performance among these hospitals. Over time, hospitals have collectively performed better in each year, as evidenced by the fact that the distribution of scores compresses and shifts towards 1 (the highest possible score). By 2012 the average score was 0.94 (out of a maximum of 1), compared to a mean of 0.82 among the same hospitals in 2005.

Given this across-the board improvement, an important question is whether patients treated in hospitals with high current or past achievement on process quality measures achieve better or worse outcomes. We explore exactly these types of questions below.

III. Empirical Strategy

III.1 Ambulance Referral Patterns

Our empirical approach builds on our earlier work that relies on plausibly exogenous sources of hospital assignment determined by ambulance company preferences for certain hospitals (Doyle et al, 2012).[10] 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 ambulance company assignment is effectively random.

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.[11] Similarly, large cities such as New York, Los Angeles and Chicago have adopted

[10] This section draws heavily on the background for this approach described in Doyle et al. (2012).
[11] See Watson (2011) and Johnson (2001) for examples.

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, in Doyle et al. (2012) we conducted a survey of 30 cities with more than one ambulance 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. 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.

Previous case studies suggest that these ambulances have preferences about which hospital to choose. 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 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.[12]

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., 2012). For patient *i* assigned to ambulance *a(i)*, we calculate the average hospital measure H_j (e.g., the readmission rate) among the patients in our analysis sample for each ambulance company:

$$\mathrm{Z}_{\mathrm{a(i)}} = \frac{1}{\mathrm{N}_{\mathrm{a(i)}} - 1} \sum_{j \neq i}^{\mathrm{N}_{\mathrm{a(i)}}-1} H_j$$

This measure is essentially the ambulance company fixed effect in a model for H_j in a model that leaves out patient i. Below, we consider values for H_j that include a variety of quality measures, such as the hospital's publicly reported 30-day readmission rate, its 30-day mortality rate, its average Medicare reimbursement for an inpatient stay, or a composite process measure as described in Section II.2 above. For certain measures like average reimbursement, we exclude the patient from this measure to avoid a direct linkage between Z and the measure in a given hospital. For other quality measures that we construct, we use sample filters and a lagged measure of H_j to ensure that the patient herself is not included in the construction of the quality measure.

III.2 Empirical Model

We use this instrument to estimate the first-stage relationship between hospital quality H and the instrument, *Z*: the hospital measure associated with the ambulance assigned to patient *i* with principal diagnosis *d(i)* living in ZIP code *z(i)* in year *t(i)*:

$$H_i = \alpha_0 + \alpha_1 Z_{a(i)} + \alpha_2 X_i + \alpha_3 A_i + \gamma_{d(i)} + \theta_{z(i)} + \lambda_{t(i)} + \nu_i \qquad (4)$$

where X_i is a vector of patient controls including age, race, and sex, and indicators for 17 common comorbidities controlled for in the CMS quality scores; A_i represents a vector of

[12] Paramedics are in contact with physicians at a local hospital at the scene. Our survey revealed that the ambulance company will often speak to the hospital they are most familiar with, which could lead them to be more likely to transport patients back to their usual hospital.

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. We cluster standard errors at the Hospital Service Area (HSA) level, as each local market may have its own assignment rules.

We also include a full set of fixed effects for principal diagnosis, year and ZIP code.[13] This regression, in other words, compares individuals who live in the same ZIP code, but who are picked up by ambulance companies with different "preferences" across hospitals with different quality scores. A positive coefficient α_1 would indicate that ambulance company "preferences" are correlated with where the patient actually is admitted.

Our main regression of interest is the relationship between hospital quality on outcomes such as mortality, M, for patient i:

$$M_i = \beta_0 + \beta_1 H_i + \beta_2 X_i + \beta_3 A_i + \gamma_{d(i)} + \theta_{z(i)} + \lambda_{t(i)} + \epsilon_i \qquad (5)$$

For this regression we consider various patient outcomes, such as whether they are readmitted to another acute care facility within 30 days of discharge, or whether they died within 30 days or a year. Finally, since patient selection is likely to confound this structural model, we estimate equation (5) using two-stage least squares, with the instrument defined as above.

Doyle et al. (2012) discusses at length potential limitations with this strategy and various specification checks that begin to address them. In particular, that study finds that the impact of ambulance assignment on health outcomes occurs not in the first day but over longer horizons, which is inconsistent with results driven by underlying patient mix or ambulance company quality differences; that the results are robust to the level of heterogeneity of the zip codes, which is inconsistent with potential locational bias in ambulance assignment; that the results are highly robust to controls for both patient characteristics and the characteristics of pre-hospital care in the ambulance; and that admission to a hospital among emergent patients is not correlated with observable ambulance company characteristics.

[13] The principal diagnosis is the 3-digit ICD-9-CM diagnosis code, as shown in Appendix Table A2.

IV. Data

IV.1 Medicare Claims Data

Our primary source of patient-level data are Medicare claims between 1998 and 2010. We use these data to identify an uncensored sample of ambulance patients admitted to an acute care hospital between 2002 and 2009, as well as to construct hospital-level risk-standardized outcome measures for mortality and readmission based on all inpatient admits to acute care facilities between 1998 and 2009.

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 patient's 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, and procedures performed. Diagnoses and procedures recorded in each patient's claims for the year prior to (but not including) the ambulance admission are then mapped to Hierarchical Condition Codes (HCC) to construct a set of comorbidity measures. We also link each ambulance patient to other information on age, race, and gender. Finally, 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 are linked to these claims, which also allow us to measure mortality at different timeframes, such as 30 days or one year.

To construct risk-standardized hospital-level outcome measures for mortality and readmission we again use the Medicare claims, however the sample is not restricted to ambulance patients. To increase the sample size to estimate these measures, we rely on 100% claims files for 1998-2009. We similarly link each of these patients to vital status data, demographic and enrollment data, and their one-year Medicare utilization history to construct 17 comorbidity controls based on HCC categories.

IV.2 Sample Selection

For both our ambulance patient sample and for the construction of outcome quality measures, we rely on a sample consisting of patients admitted to the hospital 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, 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. Appendix Table A1 shows the distribution of admissions across these diagnostic categories. These conditions represent 39% of the hospital admissions via the emergency room, 61% of which arrived by ambulance. Moreover our conditions include sepsis – the most costly inpatient condition in the United States[14] – as well as AMI and Pneumonia, two commonly used conditions to profile quality. 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, and has a higher 365-day mortality rate (34%) 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, so the applicability of our results to less emergent hospitalizations may be limited; we discuss this point further in the conclusion. Our final analytic sample is comprised of 329,218 patients.

IV.3 Hospital Quality Measures

IV.3.1 Hospital Fixed Effects

We first estimate hospital fixed effects for mortality and readmissions within 30 and 365 days of a hospital admission. To the extent possible we mirrored the methodology used by CMS and constructed each measure in each year based on a lagged 3-year sample of 100% Medicare claims in each hospital. Each measure was risk-adjusted based on patient demographics and comorbidities, though like CMS we did not include certain demographic characteristics (e.g., race) in the risk-adjustment model. Our measures are based on claims between 1998 and 2010

[14] See https://www.hcup-us.ahrq.gov/reports/statbriefs/sb160.pdf

that identified index events based on the first observed hospitalization (within a 12 month period) for one of the 29 conditions that we consider.

As noted above we apply standard empirical Bayes shrinkage methods to the estimated fixed effects before producing the final risk-standardized rates. 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. Estimates that employ the shrinkage are noticeably more precise, as expected. Our constructed 30-day outcome rates correlate well with their CMS analogues.[15]

IV.3.2 CMS Reported Quality Measures: Hospital Compare

We also examine existing hospital quality measures reported by CMS through its Hospital Compare program. Process measures began to be reported in 2005, while 30-day mortality measures were added in 2008 and the 30-day readmission measures in 2009. Process measures are based on hospital self-reports in each year while outcome measures are based on claims from a 3-year pooled sample (with a 1-2 year lag) of fee-for-service Medicare and Veterans Health Administration (VA) patients.

Because the Hospital Compare process measures are a moving target (measures are added and dropped in each year) we adopted the approach of Yasaitis et al. 2009 and constructed a composite measure of process quality based on a consistently reported set of measures. This measure is a volume-weighted average of 13 measures for AMI, Pneumonia and CHF patients treated in 3,027 acute care facilities in each year.[17] Similarly, for outcome measures we construct a composite measure of 30-day mortality and readmission that is a frequency-weighted average of each measure for AMI, Pneumonia and CHF patients.

One difference between our fixed effects estimates and the CMS measures is that the CMS measures are based on three conditions, whereas ours are based on the 29 conditions in our sample. Notably, the penalties for hospitals with above average readmission rates among these

---

[15] For example, among the AMI measures, an unadjusted hospital-level regression of our 30-day mortality rate on the CMS 30-day mortality rate in 2010 yielded a coefficient of 0.91

[17] These measures include: % of AMI Patients Given Aspirin at Arrival; % of AMI Patients Given Aspirin at Discharge; % of AMI Patients Given Beta Blocker at Discharge; % of AMI Patients Given Beta Blocker at Arrival; % CHF of Patients Given Assessment of Left Ventricular Function (LVF); % CHF of Patients Given Oxygenation Assessment; % of Pneumonia Patients Assessed and Given Pneumococcal Vaccination; % of Pneumonia Patients Given Initial Antibiotic(s) within 4 Hours After Arrival; % of Pneumonia Patients Given Smoking Cessation Advice/Counseling; % of Patients Given Discharge Instructions; % of Pneumonia Patients Having a Blood Culture Performed Prior to First Antibiotic Received in Hospital; % of Pneumonia Patients Given the Most Appropriate Initial Antibiotic(s); % of Surgery Patients Who Received Preventative Antibiotic(s) One Hour Before Incision

three conditions is applied by CMS to all diagnosis related group (DRG) payments, not just to the DRG for those individual conditions.[18] Recent research relating CMS measures to all-cause mortality has found a strong positive correlation (McCrum ML et al. 2013). Our results will compare the effectiveness of quality scores based on three conditions to measures based on a wider set of conditions in a causal framework.

Because both AMI and Pneumonia are in our set of nondeferrable conditions we also consider separate condition-specific analyses that focus only on those patients and measures.

IV.4 Standardized Hospital Measures

For all of our regressions the quality measures enter as a continuous measure that has been standardized by 2 standard deviations to facilitate interpretation and comparison across measures. Thus, each has an overall mean of 0 and a standard deviation of 0.5. This standardization procedure is designed so that the coefficients can be interpreted as if they were estimated on a binary low vs. high "quality" measure.

For context, Appendix Table A2 reports means for the unadjusted measures. For the CMS measures (which are based on AMI, CHF and PNA patient sample) a two standard deviation increase represents a 0.13 increase for the composite process quality index for 2005 (mean = 0.82 in our sample), 2.4 percentage points in the 30-day mortality rate (mean = 11.6%), and 3.6 percentage points in the 30-day readmission rate (mean=20.9%). Likewise, for our constructed composite nondeferrable condition-based quality measures a 2 standard deviation increase amounts to roughly 4.0 percentage points for both 30-day mortality (mean = 12.4%), 3.4 percentage points for readmissions (mean = 12.5%) and 6.4 percentage points for the 365-day mortality measures (mean = 31.3%).

V. Results

V.1 Balance

To evaluate the relationship between measured hospital performance and patient outcomes we rely on an instrumental variables approach that assumes patients are quasi-

[18] http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html

randomly assigned to ambulances in an emergency. If this assumption holds then our regression approach should purge endogeneity stemming from patient-level selection into different hospitals. To test whether this is true along observable dimensions, Table 1 shows means of patient-level demographic and health measures across those whose ambulances tend to transport patients to hospitals with high versus low 365-day mortality rates: a measure that we emphasize in our results below. In particular, the data are divided into quartiles based on the distribution of the ambulance instrument for 365-day mortality after it has been de-meaned at the ZIP code level.

The table shows that our sample is remarkably balanced on observable demographic and health characteristics. The average age of patients whose ambulances are more likely to take patients to hospitals in the lowest and the highest quartile of 365-day mortality, for example, is 81.4 years. Likewise, patients in the lowest quartile are transported by ambulance an average of 6.5 miles, versus 6.6 miles in the highest quartile. Balance is similar even across measured dimensions of health as captured by our comorbidity measures: 18.3 percent of patients in the lowest quartile have an indicator of hypertension in their Medicare claims for the previous year – nearly identical to the percentage with hypertension in the highest quartile (18.2 percent). A test of statistical significance comparing patients in the highest and lowest quartiles rejects the null hypothesis of no difference in just 1 out of 30 measures, and even in that instance the difference is negligible. Similar results are found across quartiles of an inpatient reimbursement-based instrument, as documented in Doyle et al., 2012.

V.2 Quality Measures and Patient Outcomes

V.2.1 Quality Measure based on 1-Year Mortality

Our first set of results describes the relationship between patient outcomes and the quality measure based on the 365-day mortality fixed effect. We also examine whether the results are sensitive to controls.[21] The models ask whether a patient assigned to a hospital with a low 365-day mortality rate for other patients is successful in lowering mortality for subsequent patients. Table 2 presents estimates from models with ZIP code fixed effects and various right-hand-side controls (e.g., demographics, diagnosis controls, ambulance controls).

[21] Analogous results for the other quality measures we discuss are in Appendix Table 2.

Panel A reports the first-stage relationship between the quality measure associated with the hospitals where ambulance companies take other patients and the quality measure of the hospital where the patient is actually treated. The point estimates indicate that the first-stage relationship is strong: patients transported by ambulances with preferences for hospitals profiled with higher mortality rates are much more likely to be treated themselves in a hospital with a higher mortality rate. Similarly large first stage estimates are found for every hospital measure we consider (Appendix Table A3). Moreover, looking across the columns it is clear that these relationships are robust to the addition and subtraction of controls.[22]

Panels B and C show that a hospital's 365-day mortality quality measure is indeed positively correlated with subsequent patient mortality (recall that the patient is left out of the calculation of the quality measure). In the OLS comparison, a 2 standard deviation increase in this measure is associated with a 4 percentage point increase in mortality compared to a mean 365-day mortality rate of 33%.

When we move to 2SLS, the point estimates increase from 0.042 to 0.071 with full controls, although the standard errors also increase and we do not reject that the IV estimates are significantly different from the OLS results. While the point estimates suggest that OLS underestimates the relationship for the marginal cases that inform the 2SLS estimate, the results suggest that 365-day mortality rates, risk adjusted in a standard way, provide a useful measure of hospital quality.

V.2.2 Alternative Quality Measures

We next turn in Table 3 to examine the relationship between various quality measures and patient outcomes. Each cell is from a separate regression of the dependent variable indicated in the column on the quality measure indicated in each row, along with the full set of control variables described above. We examine in particular two relevant outcomes: readmission to any acute care hospital within 30 days and mortality at 365 days. We have also examined the

[22] 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.

impact on mortality at 30 days and all of the results are similar to those shown here; those results are shown in Appendix Table 1.

The top row of Table 3 shows the CMS process-based composite quality score as a summary measure of the extent to which hospitals regularly incorporate and report evidence-based care. We start by examining the hospital's most recent measure from 2010. As shown in Figure 1, this measure offers little variation as most hospitals meet the guidelines. For comparison, we also include the earliest measure from 2005, which exhibits the widest dispersion across hospitals. To the extent that this measure picks up on differences in underlying quality – as opposed to differences in documentation, which may or may not be correlated with quality – it should be apparent in the 2005 score since that score was the first one publicly reported.

The remaining rows of Table 3 extend our analysis to consider other quality measures. The third row shows the results for quality measured as a composite average of publicly reported 30-day readmission rates for AMI, PNA, and CHF patients. These results are then followed by results using our own computation of a composite readmission rate estimated using the sample of non-deferrable conditions and using a hospital fixed effect that has been adjusted using the empirical-Bayes shrinkage as described in section IV.3.1 above. Similarly, the remaining rows consider the composite average of publicly reported CMS 30-day mortality rates, as well as our own 365-day mortality rates.

For each dependent variable, we present two columns, for OLS and then using the 2SLS approach described above. We focus in our discussion on the 2SLS results which we view as more reliable, but given the larger standard errors none of these estimates is significantly different from its OLS counterpart.

The first row of Table 3 shows that a higher CMS process score in 2010 is only very weakly associated with both readmissions and mortality. A two-standard deviation increase in the measure is associated with a 0.7 percent reduction in readmissions and a 0.7 percent reduction in 365-day mortality, compared to means of 15.6 and 33 percent respectively. There is a stronger association, however, between the 2005 process score and mortality outcomes, with our IV estimate suggesting that a 2 standard deviation rise in the 2005 process score is associated with a reduction in mortality of 2.8%, or about 8% of the baseline. The much stronger association with 2005 is not surprising given how much hospitals have converged by 2010. It does suggest that process quality measures may be more effective if targets evolve over time to highlight the highest quality hospitals.

The next set of rows considers both the CMS readmission measure and our own computed fixed effects measure. We find that our measure is a better predictor of readmission, although neither measure significantly predicts mortality outcomes. Of course, the CMS measure is based on only three conditions while ours is based on the 29 conditions that form our sample. This suggests that focusing on three conditions can be only so effective at distinguishing hospitals across a range of other conditions. We return to this point below when we focus on two specific conditions.

The next two rows of the Table consider the CMS measure of 30 day mortality, as well as our own fixed effects estimates of one-year mortality. We find that both admissions are mortality are significantly predicted by our one year mortality measure, while there is little predictive value in the CMS mortality measure in either case. Even when we study 30 day mortality in Appendix Table 1, we find that the one year mortality fixed effect estimate is a better predictor of outcomes.

V.2.3 Condition-specific Measures

As noted above, the CMS measures and our own measures are somewhat difficult to compare because the CMS measures are defined over only three diagnoses, while ours are over the sample of non-deferrable admissions. We next turn to the particular case of quality measures and patient outcomes for AMI and pneumonia, since both diagnoses are included in our sample of patients with non-deferrable admissions. In this case we use the condition-specific measures from CMS, as well as our own AMI- and pneumonia-specific estimates. Given the much smaller samples when studying these particular conditions, we replace the zip code fixed effects used in Table 3 with HSA fixed effects. Appendix Table 1 shows that the results in Table 3 are robust to the use of HSA fixed effects, albeit with somewhat weaker conclusions for the mortality measures.

Table 4 shows the results for AMI and pneumonia in a format that parallels Table 3, where each cell is again the coefficient of interest from a regression of the dependent variables indicated in the columns on the independent variables indicated in the rows. The top panel shows the results for AMI, while the bottom panel shows the results for pneumonia. Our major conclusions from Table 3 are maintained here. First, the 2005 process measure is more predictive than the 2010 measure, although neither is very predictive for pneumonia. This may be due to the fact that the measures for pneumonia, including smoking cessation advice, may be less effective than the more substantiated measures for AMI, including the use of beta blockers. Second, the fixed effects measures are typically more predictive than the CMS measures. Finally, the strongest correlate of mortality is the 365-day mortality fixed effects estimate, which is large both for AMI and pneumonia in both OLS and 2SLS.

We draw four conclusions from this overview of quality measures. First, while the 2SLS measures of quality impacts often differ from OLS, the differences are typically modest and never significant. Second, older process score measures correlate more strongly with outcomes, suggesting the value of updating process score targets. Third, by far the strongest relationship of those estimated here is between fixed effects estimates of mortality and mortality outcomes, either across the broad set of 29 deferrable conditions, or specifically by condition. This suggests that there is enormous value in quality measures of this nature, even when controlling for patient selection. Fourth, he mortality fixed effect is approximately as good as the readmission fixed effect at predicting subsequent readmissions once we control for patient selection, and the mortality measure is, less surprisingly, much better at distinguishing hospitals based on their subsequent mortality outcomes. This suggests that the mortality fixed effect dominates readmissions when it comes to measuring hospital quality.

V.3 Is Quality Measurement Enough? Feasibility of Quality-Based Reference Pricing

The fact that we have estimated a consistent relationship between mortality-based hospital quality measures and patient outcomes suggests that policy makers could potentially use such measures as a basis for hospital reimbursement. One version of this approach is known as reference pricing: payers would reimburse hospitals solely based on their quality, and not based on their actual treatment intensity, to ensure that reimbursement is targeted to the highest quality hospitals. But the appeal of such an approach depends on both a correlated measure of quality *and* the lack of a strong residual relationship with current hospital reimbursement. If hospital reimbursement still impacts outcomes conditional on quality, then reference pricing could lead to worsening outcomes.

V.3.1 Measuring Hospital Reimbursement Levels

To capture Medicare reimbursement for care included in the initial hospital stay, as well as for care received in the days after discharge, we turn to measures of hospital-level average (log) reimbursements for *other* emergency patients treated in the same hospital. Our key reimbursement measure for initial inpatient care includes the average amount paid to the hospital under the prospective inpatient payment system (the Diagnosis Related Group, or DRG, amount) plus any outlier and graduate medical education payments. This measure also includes Medicare Part B payments that reimburse for the physician component of patient care within the hospital. Higher average reimbursements within an area are driven by greater treatment intensity, reflected in more costly DRGs, higher fees paid to physicians, and in outlier payments.

In addition, we consider additional reimbursement measures that include payment for different aspects of care that during the initial hospital stay and in the 90 days post discharge. In short, these measures seek to summarize both inpatient and post-acute reimbursement amounts under the Medicare program for emergency patients. We chose a 90-day cutoff for these measures because the vast majority (95%) of bundled payment demonstrations under Medicare defines a bundle across 90 days of post-acute care following an indexing hospitalization.

V.3.2 Relationship Between Patient Outcomes and Reimbursement Levels

In Table 6 we run "horse race" regressions of one year mortality on both quality measures and on hospital reimbursement; we focus on one year mortality because we view that as a more relevant long term outcome measure, but our results are very similar (albeit with smaller magnitudes) if we use 30 day mortality instead.

Panel A of Table 5 reports results from models estimated with OLS. The hospital-quality results are similar to those above, and higher spending is associated with somewhat lower mortality, although the relationship is not overly large. A two standard deviation increase in log reimbursement is 0.42, and these results suggest that going from a low to a high-spending hospital would result in a 1.6 percentage point reduction in mortality, or 5% of the mean.

A concern is that these estimates may in part be due to patient selection if higher spending hospitals treat healthier patients. For example, the most expensive treatments are less likely to be provided for older patients. Panel B estimates the relationship between mortality, inpatient spending, and the 365-day mortality rate of the hospital employing two instruments: the average hospital reimbursements at hospitals where the patient's ambulance takes other patients, and the average quality (measured by the 365-day mortality fixed effect) of the hospitals where the ambulance company takes other patients. The quality-score coefficients are again similar to those reported above, and the spending coefficients suggest that higher-reimbursement hospitals have significantly better outcomes. The point estimate of 0.18 suggests that moving from a low to a high-reimbursement as described above hospital is associated with a 7.6 percentage point reduction in mortality or 22% of the mean. Our IV results, therefore, suggest that there is substantial value added by hospitals that are reimbursed more highly, even conditional on our preferred quality measure.

Appendix Table A4 extends these results to consider the full range of quality measures shown in Table 3. We find, consistent with Table 5, that the quality measures themselves do not much change when inpatient spending is included; likewise, we find that inpatient spending has a consistent, large and significant effect regardless of the quality measure used.

V.4 Reimbursements for Care Inside vs. Outside the Hospital

The results presented thus far seem to contradict the well-documented finding that higher spending does not deliver better outcomes, at least across areas of the U.S.(Baicker, Chandra and Skinner, 2012). But as we highlight in this section, the major reason for the difference is our focus on inpatient reimbursement, as opposed to total Medicare reimbursements. Indeed, incorporating the impact of post-acute reimbursement, we find little overall impact of *total* reimbursement on patient outcomes.

To investigate this issue, in Table 6 we first move from our measure of inpatient reimbursement during the index episode to total inpatient reimbursement over the 90 days after the index admission common in payment reforms that aim to bundle payment for acute and post-acute care. The source of variation in the instrumental-variable estimation comes from the 90-day reimbursements associated with patients who were initially treated at a given hospital. Some of this care can take place at another hospital, but our goal is to characterize hospitals based on subsequent reimbursement regardless of the location.

The 2SLS point estimate is now larger, -0.31, which suggests that being admitted to a more expensive hospital leads to subsequently productive hospital spending. That is, hospitals that rely on intensive use of hospital care—both at the time of admission and in subsequent care such as planned surgeries in the post-acute phase—have significantly lower mortality.

In the second column of Table 6, we instead use a measure of total *post-acute* reimbursement over the first 90 days post index admission. As noted above, this is the hospital average of total Medicare reimbursement for nondeferrable patients in the 90 days after being discharged from the hospital. Thus, the measure summarizes, for a given hospital, the average amount of care -- including follow-up doctor visits, visits to specialty facilities, and stays in long-term hospitals and skilled nursing facilities -- incurred downstream of their indexing admission. In short, this measure is designed to differentiate between hospitals whose patients tend to receive more vs. less intensive treatment after they are discharged. In the OLS comparison, this measure is modestly positively associated with mortality.

In the 2SLS results, the estimate is large and positive. This does not imply that post-acute treatment has negative impacts, as we are not estimating the effects of this care per se. Instead, this estimate implies that if a patient is admitted to a hospital that is associated with high post-acute reimbursements outside of the hospital setting, they a higher mortality rate. This estimate combines the effects of treatment at the time of acute episodes, which can lead to greater or fewer treatments downstream, and the effects of those downstream treatments.

Column (3) puts these two findings together, so that we can jointly assess the impact of higher inpatient reimbursement and higher post-acute reimbursement. We find that the coefficients are of similar signs and magnitudes, with the post-acute coefficient being larger (but not significantly different in absolute value). It is notable that there is little change in these coefficients when included in the same model. Column (4) then reports simply the total impact of all reimbursement 90 days post admission, and we find a fairly precise zero: being admitted to a hospital that is associated with higher Medicare reimbursement in the 90 days post admission has no impact on patient outcomes.

This finding is very consistent with findings reported in earlier work comparing regions of the country. Indeed, in Column (5) we replicate the primary measure used in that earlier work, patient end of life spending. This measure, which was constructed at the hospital-level by the Dartmouth Atlas of Health Care, summarizes the total amount of Medicare reimbursement among decedents with chronic conditions in the last two years of life.[23] This measure is not

[23] http://www.dartmouthatlas.org/tools/downloads.aspx#reimbursements

restricted to patients with non-deferrable conditions. In keeping with the earlier results, we find no impact of that measure on patient outcomes.

Columns (6), (7) and (8) replicated the previous three columns, including our most predictive measure of hospital quality, our fixed effect 365 day estimate. We find that this variable enters the model significantly and with a similar coefficient to our previous models, and that it has relatively little effect on the reimbursement coefficients. This means that conditional on our quality measure, inpatient reimbursements are associated with better outcomes, and non-hospital reimbursements continue to be associated with worse outcomes, and these effects cancel one another such that total reimbursement is unrelated to mortality.

VI. 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. Major issues to address include how to measure quality, how to address patient selection in the correlation between quality measures and outcomes, and how to incorporate quality with cost-based reimbursement. In this paper we have endeavored to address all three issues.

In particular, we appraised the quality of hospital quality assessments by considering the relationship between hospital report card measures and patient outcomes in a framework purged of patient selection. At least among the marginal cases that make up our ambulance preference-based instrument, process measures are at best weakly informative as to overall hospital quality. Similarly, outcome measures based on readmission rates predict patient-level readmission risk but are seemingly uncorrelated with other important outcomes such as patient mortality. On the other hand, outcome-based quality measures based on mortality are predictive of patient-level mortality and readmission risk, and these relationships hold in both OLS and 2SLS. We view this as good news: despite concerns about the adequacy of risk adjustment to allow outcome comparisons across hospitals, this evidence suggests that the risk-adjusted mortality measures are informative.

A key limitation of outcome quality measures – and one that is most stark when one takes a fixed-effect approach to profiling hospitals – is that they are imprecisely measured. Our instrumental variables approach can help alleviate both attenuation bias from measurement error as well as residual confounding from patient selection. Indeed, we see evidence of this in our results – when coupled with average hospital inpatient reimbursement the coefficient on the 365-day mortality rate estimate doubles and is statistically different than when included on its own (Table 5). That said, the 2SLS estimates have larger standard errors than their OLS counterparts.

Finally, our results also speak to both the enormous difficulties and the great potential of delivery system reforms that more directly link provider reimbursement to measures of quality. If reimbursement were not associated with better outcomes conditional on quality, then reference pricing could be used to pay for quality rather than treatment intensity. As we show here, however, there are strong returns associated with inpatient treatment intensity even conditional on approaches to measuring hospital quality. Interestingly, we find that post-discharge care is positively associated with mortality, which points the way to future research on returns (or lack thereof) to such care and the potential gains from the coordination of care that takes place outside of the hospital.

These findings have implications for the current debate over reimbursement reform. As noted, they suggest that reimbursement based on quality only would harm patients. But they also suggest that hospitals should bear some of the financial consequences for their downstream unproductive care. This would appear to motivate broad "episode based" reimbursement bundles that incorporate both hospital and post-acute care. But the problem with such an approach is that it does not distinguish between (productive) inpatient spending and (unproductive) post-acute spending. An alternative approach would be separate reimbursement bundles within episode-based systems, whereby hospitals are reimbursed for their own costs, and downstream providers for their costs, but where hospitals bear some of the financial implications of downstream spending through "shared savings" system.

Both of these approaches would integrate the financial incentives of hospitals with downstream providers. Which reform is to be preferred depends on whether policy makers think that the appropriate mix of incentives is best decided by the policy maker, who sets the rate of shared savings, or by the joint producers of the care during the episode. That in turn may depend on the relative negotiation power of those producers. Understanding differences between these approaches is an important topic for future research.

Of course, a major limitation of all of the work in this paper is that it only applies to the roughly one-quarter of hospital admissions which are emergent and arrive by ambulance.

Optimal hospital reimbursement systems must devise payment methodologies as well for other patients, and it is not entirely clear whether the results from this subsample apply. 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 hospital outcomes.

Ash, Arlene S., Stephen E. Fienberg, Thomas A. Louis, Sharon-Lise Normand, Therese Stukel, and Jessica Utts. 2012. of hospital admissions which are emergent and arrive by Centers for Medicare and Medicaid Services. http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf.

Baicker, Katherine, Chandra, Amitabh, and Jonathan Skinner. 2012. "Saving Money or Just Saving Lives? Improving the Productivity of US Health Care Spending." *Annual Review of Economics*. 4: 33-56.

Berenson, Robert A., Ronald A. Paulus, and Noah S. Kalman. 2012. n, Robert A., Ronald A. Paulus, and Noah S. Kalman. 2012. ignative." *New England Journal of Medicine* 366 (15): 1364rnal doi:10.1056/NEJMp1201268.

Birkmeyer, John D., Andrea E. Siewers, Emily V.A. Finlayson, Therese A. Stukel, F. Lee Lucas, Ida Batista, H. Gilbert Welch, and David E. Wennberg. 2002. rformance.pdf.by and Surgical Mortality in the United States." *New England Journal of Medicine* 346 (15): 1128rnal doi:10.1056/NEJMsa012337.

Boyarsky, Victoria, and Rob Parke. 2012. y, Victoria, and Rob Parke. 2012. ily V.A. Finlayson, Therese A. Stukel, F. Lee LuMilliman. http://us.milliman.com/uploadedFiles/insight/healthreform/medicare-shared-savings-program.pdf.

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. avid, Carlos Dobkin, and Nic*Quarterly Journal of Economics* 124 (2): 597rnal oChandra, Amitabh, Finkelstein, Amy, Sacarany, Adam and Chad Syverson. 2013. "Healthcare Exceptionalism? Productivity and Allocation in U.S. Healthcare Sector." NBER Working Paper Series Number 19200. 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. , Amitabh, and Douglasrs in Healthcare: Evidence From The Treatment of Heart Attacks.ck*Journal of Political Economy* 115 (1): 103liti

Chetty, Raj, Friedman, John, and Jonah Rockoff. 2013. " Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." NBER Working Paper Series Number 19424. Cambridge, Mass: National Bureau of Economic Research.

Committee on Quality of Health Care in America, Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. The National Academies Press. http://www.nap.edu/openbook.php?record_id=10027.

Daley, Jennifer. 2002. s Press. http://www.nap.edu/openbook.php?record_id=10027.me Relationship-What2. s Press. http://www*Surgery* 131 (1): 16hip-WDimick, Justin B., Douglas O. Staiger, Nicholas H. Osborne, Lauren H. Nicholas, and John D. Birkmeyer. 2012. Justin B., Douglas O. Staiger, Nicholas H. Osborne, LaurenSurgery.Ju*Health Services Research* 47 (5): 1861es Researchas O. Staiger, Nicholas H. OsborDobkin, Carlos. 2003. chaspital Staffing and Inpatient Mortality." *Unpublished Working Paper*.

Doyle, Joseph, John Graves, Samuel Kleiner, and Jonathan Gruber. 2012. Joseph, John Graves, Samuel Kleiner, and JonaEvidence from Ambulance Referral Patternsr, Working Paper 17936. National Bureau of Economic Research. http://www.nber.org/papers/w17936.

Dudley, R. Adams, Kirsten L. Johansen, Richard Brand, Deborah J. Rennie, and Arnold Milstein. 2000. R. Adams, Kirsten L. Johansen, Richard Brand, *JAMA: The Journal of the American Medical Association* 283 (9): 1159nal oEpstein, Arnold M., Ashish K. Jha, and E. John Orav. 2011. , Arnold M., Ashish K. Jha, and E. John Orav. 2011. ah J. Rennie, and Arnold*New England Journal of Medicine* 365 (24): 2287rnal doi:10.1056/NEJMsa1101942.

Fisher, Elliott S, Julie P Bynum, and Jonathan S Skinner. 2009. Elliott S, Julie P Bynum, and Jonathan S Skinner.. ah J. Rennie, and Arnold*The New England Journal of Medicine* 360 (9): 849nd Jodoi:10.1056/NEJMp0809794.

Fisher, Elliott S., David E. Wennberg, Thathan S Sktukel, Daniel J. Gottlieb, F. L. Lucas, and É toile L. Pinder. 2003a. Elliott S., David E. Wennberg, Thathan S Sktukel, Daniel J. G Part 2: Health Outcomes and Satisfaction with Care.uk*Annals of Internal Medicine* 138 (4): 288 rnal MFisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Étoile L. Pinder. 2003b. Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Part 1: The Content, Quality, and Accessibility of Care.l,*Annals of Internal Medicine* 138 (4): 273–87.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/hierarchical Models.* Cambridge: Cambridge University Press.

Halm, Ethan A., Clara Lee, and Mark R. Chassin. 2002. than A., Clara Lee, and Mark R. Chassin. 2002 A Systematic Review and Methodologic Critique of the Literature.s *Annals of Internal Medicine* 137 (6): 511ernalHughes, Robert G., Sandra S. Hunt, and Harold S. Luft. 1987. Robert G., Sandra S. Hunt, and Harold S. Luft. 1987. re.s. J. Gottlieb, F. L. Lu*Medical Care*, 489al CarJohnson, Robin. 2001. dra S. Hunt, and Harold S. Luft. 1987. re.s. J. Gottlieb, F. L. Lucas, and Étoile L. Pin*Reason Foundation*. http://reason.org/news/show/the-future-of-local-emergency.

Joseph P. Newhouse, Alan Garber, and Robin P. Graham, ed. 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.

Luft, Harold S., Sandra S. Hunt, and Susan C. Maerki. 1987. arold S., Sandra S. Hunt, and Susan C. Maerki.enbook.php?record_id=18308.erral Patterns?, *Health Services Research* 22 (2): 157.

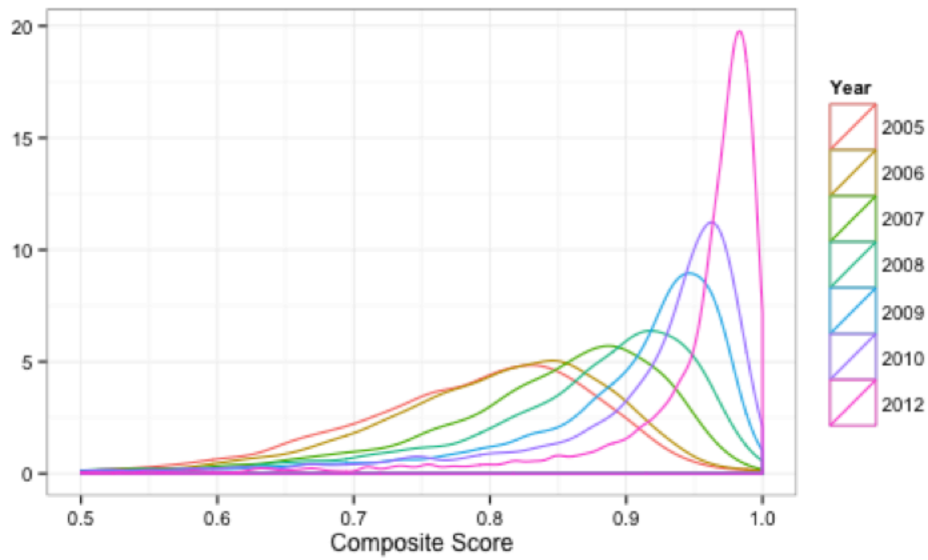
McCrum ML, Joynt KE, Orav E, Gawande AA, and Jha AK. 2013. ML, Joynt Kfor Publicly Reported Conditions and Overall Hospital Mortality Rates.” *JAMA Internal Medicine* 173 (14): 1351edicinePublicly Reported Conditions and OvMcGlynn, Elizabeth A., Steven M. Asch, John Adams, Joan Keesey, Jennifer Hicks, Alison DeCristofaro, and Eve A. Kerr. 2003. , Elizabeth A., Steven M. Asch, John Adams, Joan Kethe United States. A*New England Journal of Medicine* 348 (26): 2635rnal doi:10.1056/NEJMsa022615.

Newhouse, Joseph P., and Alan M. Garber. 2013. e, Joseph P., and Alan M. Garber. 2013. msing in the United States: Insights From an Institute of Medicine Report.” *JAMA* 310 (12): 1227. doi:10.1001/jama.2013.278139.

Robinson, James C., and Kimberly MacPherson. 2012. n, James C., and Kimberly MacPhAnd Centers Of Excellence To Steer Patients To Low-Price And High-Quality Providers." *Health Affairs* 31 (9): 2028scellence To Steer Patients To Low-PRyan, Andrew, James Burgess, Robert Strawderman, and Justin Dimick. 2012. ndrew, James Burgess, Robert Strawderman, and Justin Dimick. A Simulation Approach." *Health Services Research* 47 (4): 1699es Researchess, Robert Strawderman, and JustiShahian, David M., and Sharon-Lise T. Normand. 2003. , David M., and Sharon-Lise T. Normand. 2003. Justin Dimic*The Annals of Thoracic Surgery* 75 (3): 1048 ThorSong, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E Wennberg, and Elliott S Fisher. 2010. unjie, Jonathan Skinner, Julie Bynum, Jason Sut*The New England Journal of Medicine* 363 (1): 45and Jdoi:10.1056/NEJMsa0910881.

Yasaitis, Laura, Elliott S. Fisher, Jonathan S. Skinner, and Amitabh Chandra. 2009. s, Laura, Elliott And Intensity Of Spending: Is There An Association?on*Health Affairs* 28 (4): w566sSpending: Is There An Association?on?

Figure 1: Distribution of Composite Process Score By Year (N=3,027 Hospitals)



The above figure plots, for a consistent set of 3,027 hospitals, the distribution in each year of a composite process score based on a fixed set of 13 process measures reported between 2005 and 2010 on the Hospital Compare website. Source: Authors' analysis of Hospital Compare data.

Table 1: Covariate Balance Across Quartiles of the Risk Standardized 365-Day Mortality Rate Instrument

	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile	
Patient Age	81.5	81.5	81.3	81.5	
Male	0.378	0.382	0.376	0.378	
Race: White	0.877	0.877	0.878	0.877	
Race: Black	0.081	0.082	0.081	0.083	
Race: Other	0.040	0.039	0.039	0.039	
Miles Transported	6.534	6.407	6.506	6.624	
Ambulance: Emergency Transport	0.862	0.870	0.865	0.863	
Ambulance IV Fluids Administered	0.054	0.055	0.051	0.049	
Ambulance: Intubation Performed	0.002	0.003	0.002	0.002	
Patient Origin: Home	0.640	0.697	0.694	0.648	
Ambulance Payment	309	310	308	305	**
Hypertension	0.183	0.181	0.178	0.182	
Stroke	0.021	0.019	0.018	0.021	
Cerebrovascular Disease	0.030	0.029	0.029	0.030	
Renal Failure Disease	0.049	0.049	0.047	0.049	
Dialysis	0.005	0.004	0.004	0.004	
COPD	0.086	0.089	0.082	0.086	
Pneumonia	0.058	0.058	0.053	0.057	
Diabetes	0.082	0.081	0.078	0.081	
Protein Caloria Malnutrition	0.018	0.017	0.015	0.018	
Dementia	0.055	0.050	0.050	0.057	
Paralysis	0.024	0.022	0.022	0.024	
Peripheral Vascular Disease	0.047	0.046	0.045	0.047	
Metastatic Cancer	0.025	0.025	0.025	0.024	
Trauma	0.034	0.032	0.033	0.035	
Substance Abuse	0.023	0.024	0.023	0.023	
Major Psych. Disorder	0.014	0.013	0.013	0.015	
Chronic Liver Disease	0.004	0.004	0.004	0.004	

Columns correspond to quartiles of the constructed risk-standardized 365-Day mortality quality measure based on a hospital fixed effect. Results net out a ZIP code fixed effect. The last column reports significance test for difference between 1st and 4th Quartile means. N=334,045 (* $p < 0.05$; ** $p < 0.01$)

Table 2: OLS, First Stage and 2SLS Estimates for One-Year Mortality Rate Measure

		(1)	(2)	(3)	(4)
Outcome: Non-Def. 365-Day Mort. Rate (FE)					
Panel A. First Stage					
	Ambulance Avg. Non-Def. 365-Day Mort. Rate (FE)	0.457	0.457	0.459	0.459
		(0.014)**	(0.014)**	(0.014)**	(0.014)**
Outcome: 365-Day Mortality					
Panel B. OLS					
	Non-Def. 365-Day Mort. Rate (FE)	0.043	0.040	0.038	0.042
		(0.003)**	(0.003)**	(0.003)**	(0.003)**
Panel C. 2SLS					
	Non-Def. 365-Day Mort. Rate (FE)	0.068	0.061	0.068	0.071
		(0.022)**	(0.020)**	(0.019)**	(0.017)**
Outcome Mean		0.330			
Sample Size:		329,218			
Demographic Controls		No	Yes	Yes	Yes
Ambulance Controls		No	No	Yes	Yes
Comorbidity Controls		No	No	No	Yes

Source: 2002-2010 Medicare Claims
Each estimate is from a separate regression with diagnosis controls and ZIP fixed effects
Quality Measures have been standardized by 2 standard deviations so they
can be interpreted like binary (low-to-high) indicators
FE=Empirical Bayes Fixed Effect
* p¡0.05; ** p¡0.01

Table 3: Quality Measures

	OLS 30-Day Readmission (1)	2SLS 30-Day Readmission (2)	OLS 365-Day Mortality (3)	2SLS 365-Day Mortality (4)
CMS 2010 Composite Process Score	-0.006	-0.007	-0.010	-0.007
	(0.002)**	(0.010)	(0.002)**	(0.014)
CMS 2005 Composite Process Score	-0.005	-0.002	-0.015	-0.028
	(0.002)*	(0.009)	(0.003)**	(0.012)*
CMS 30-Day Readmission Rate	0.011	0.004	0.004	0.004
	(0.003)**	(0.011)	(0.003)	(0.014)
Non-Def 30-Day Readmission Rate (FE)	0.026	0.030	0.019	0.023
	(0.003)**	(0.013)*	(0.004)**	(0.015)
CMS 30-Day Mortality Rate	0.002	0.009	0.017	0.010
	(0.003)	(0.010)	(0.003)**	(0.013)
Non-Def 365-Day Mort. Rate (FE)	0.008	0.026	0.042	0.071
	(0.002)**	(0.013)*	(0.003)**	(0.017)**
Outcome Mean	0.156	0.156	0.330	0.330
Sample Size	331,793			

Source: 2002-2010 Medicare Claims
Each estimate is from a separate regression with full controls.
Quality Measures have been standardized by 2 standard deviations so they can be interpreted like binary (low-to-high) indicators
FE = Empirical Bayes Fixed Effect
* $p<0.05$; ** $p<0.01$

Table 4: Quality Measures: AMI and Pneumonia Samples

		OLS 30-Day Readmission (1)	2SLS 30-Day Readmission (2)	OLS 365-Day Mortality (3)	2SLS 365-Day Mortality (4)
Panel A. AMI Sample	CMS 2005 Composite Process Score	0.003	-0.025	-0.048	-0.050
		(0.006)	(0.029)	(0.009)**	(0.036)
	CMS 2010 Composite Process Score	0.005	0.039	-0.063	-0.075
		(0.006)	(0.021)	(0.008)**	(0.026)**
	CMS 30-Day Readmission Rate	0.016	0.016	0.030	0.017
		(0.006)**	(0.020)	(0.008)**	(0.024)
	Non-Def 30-Day Readmission Rate (FE)	0.030	0.006	0.045	0.039
		(0.008)**	(0.029)	(0.010)**	(0.036)
	CMS 30-Day Mortality Rate	0.004	0.002	0.051	0.067
		(0.005)	(0.017)	(0.008)**	(0.025)**
	Non-Def 365-Day Mort. Rate (FE)	0.013	0.006	0.085	0.087
		(0.006)*	(0.023)	(0.008)**	(0.030)**
	Outcome Mean	0.162	0.162	0.340	0.340
	Sample Size	37,684			
Panel B. Pneumonia Sample	CMS 2010 Composite Process Score	-0.001	-0.011	-0.013	0.006
		(0.004)	(0.014)	(0.007)*	(0.017)
	CMS 2010 Composite Process Score	-0.006	0.000	-0.007	-0.015
		(0.005)	(0.015)	(0.006)	(0.020)
	CMS 30-Day Readmission Rate	0.028	0.034	0.012	0.013
		(0.005)**	(0.016)*	(0.007)	(0.020)
	Non-Def 30-Day Readmission Rate (FE)	0.033	0.055	0.017	0.016
		(0.006)**	(0.019)**	(0.008)*	(0.024)
	CMS 30-Day Mortality Rate	0.006	0.025	0.036	0.042
		(0.005)	(0.015)	(0.006)**	(0.020)*
	Non-Def 365-Day Mort. Rate (FE)	0.003	0.027	0.057	0.089
		(0.005)	(0.019)	(0.006)**	(0.024)**
	Outcome Mean	0.149	0.149	0.359	0.359
	Sample Size	51,297			

Source: 2002-2010 Medicare Claims
Quality measures reflect those measured for the specific condition (AMI or Pneumonia).
Each estimate is from a separate regression with full controls. and Health Service Area (HSA) fixed effects.
Quality Measures have been standardized by 2 standard deviations so they
can be interpreted like binary (low-to-high) indicators
FE = Empirical Bayes Fixed Effect
* p<0.05; ** p<0.01

Table 5: Inpatient Spending

Outcome: 365-Day Mortality		(1)	(2)	(3)	(4)
Panel A. OLS	Total Index Inpatient Spending	-0.065	-0.043	-0.039	-0.039
		(0.008)**	(0.008)**	(0.008)**	(0.008)**
	Non-Def. 365-Day Mortality Rate (Fixed Effect)	0.042	0.040	0.038	0.043
		(0.003)**	(0.003)**	(0.003)**	(0.003)**
Panel B. 2SLS	Total Index Inpatient Spending	-0.186	-0.179	-0.179	-0.179
		(0.071)**	(0.069)**	(0.065)**	(0.062)**
	Non-Def. 365-Day Mortality Rate (Fixed Effect)	0.085	0.082	0.088	0.090
		(0.022)**	(0.021)**	(0.020)**	(0.018)**
Outcome Mean		0.330			
ZIP Fixed Effects		Yes	Yes	Yes	Yes
Diagnosis Controls		Yes	Yes	Yes	Yes
Demographic Controls		No	Yes	Yes	Yes
Ambulance Controls		No	No	Yes	Yes
Comorbidity Controls		No	No	No	Yes

Source: 2002-2010 Medicare Claims
Each estimate is from a separate regression with full controls.
Quality Measures have been standardized by 2 standard deviations so they can be interpreted like binary (low-to-high) indicators
All reimbursement measures are hospital averages and are in logs.
* $p<0.05$; ** $p<0.01$

Table 6: Total 90-Day Spending

Outcome:	365-Day Mortality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. OLS	90-Day Inpatient Spending	-0.031		-0.029			-0.034		
		(0.007)**		(0.007)**			(0.007)**		
	90-Day Non-Inpatient Spending		0.016	0.014			0.019		
			(0.006)*	(0.006)*			(0.006)**		
	90-Day Total Spending				-0.004			-0.006	
					(0.002)			(0.002)**	
	End of Life Spending (Dartmouth Atlas)					-0.004			-0.003
						(0.001)**			(0.001)*
	Non-Def. 365-Day Mort. Rate (FE)						0.043	0.041	0.042
							(0.003)**	(0.003)**	(0.003)**
Panel B. 2SLS	90-Day Inpatient Spending	-0.306		-0.239			-0.251		
		(0.057)**		(0.064)**			(0.065)**		
	90-Day Non-Inpatient Spending		0.407	0.371			0.395		
			(0.047)**	(0.049)**			(0.051)**		
	90-Day Total Spending				0.003			0.004	
					(0.004)			(0.004)	
	End of Life Spending (Dartmouth)					-0.001			0.001
						(0.001)			(0.001)
	Non-Def. 365-Day Mort. Rate (FE)						0.123	0.085	0.087
							(0.020)**	(0.019)**	(0.020)**
Outcome Mean		0.330							
Sample Size		331,793							

Source: 2002-2010 Medicare Claims
Each estimate is from a separate regression with full controls.
Quality Measures have been standardized by 2 standard deviations so they can be interpreted like binary (low-to-high) indicators
All reimbursement measures are hospital averages and are in logs.
FE=Empirical Bayes Fixed Effect
* p<0.05; ** p<0.01

Table A1: Principal Diagnoses in Main Analysis

3-digit Principal Diagnosis	ICD-9 Category (1)	Weekend Rate of Admission (2)	Observations (3)
038 Septicemia	All Other	0.265	30,826
162 Malignant neoplasm of trachea, bronchus, and lung	All Other	0.269	3,451
197 Secondary malignant neoplasm of respiratory and digestive systems	All Other	0.269	2,448
410 Acute myocardial infarction	Circulatory	0.270	31,981
431 Intracerebral hemorrhage	Circulatory	0.282	6,050
433 Occlusion and stenosis of precerebral arteries	Circulatory	0.264	3,721
434 Occlusion of cerebral arteries	Circulatory	0.274	30,662
435 Transient cerebral ischemia	Circulatory	0.274	11,798
482 Other bacterial pneumonia	Respiratory	0.269	3,770
486 Pneumonia, organism unspecified	Respiratory	0.272	38,317
507 Pneumonitis due to solids and liquids	Respiratory	0.278	11,088
518 Other diseases of lung	Respiratory	0.272	17,053
530 Diseases of esophagus	Digestive	0.268	3,927
531 Gastric ulcer	Digestive	0.265	4,325
532 Duodenal ulcer	Digestive	0.280	2,997
557 Vascular insufficiency of intestine	Digestive	0.279	2,473
558 Other and unspecified noninfectious gastroenteritis and colitis	Digestive	0.282	2,879
560 Intestinal obstruction without mention of hernia	Digestive	0.277	7,736
599 Other disorders of urethra and urinary tract	All Other	0.265	20,939
728 Disorders of muscle, ligament, and fascia	All Other	0.257	2,431
780 General symptoms	All Other	0.286	34,424
807 Fracture of rib(s), sternum, larynx, and trachea	Injury	0.276	2,140
808 Fracture of pelvis	Injury	0.264	5,782
820 Fracture of neck of femur	Injury	0.267	48,556
823 Fracture of tibia and fibula	Injury	0.264	1,962
824 Fracture of ankle	Injury	0.266	4,027
959 Injury, other and unspecified	Injury	0.257	702
965 Poisoning by analgesics, antipyretics, and antirheumatics	Injury	0.265	632
969 Poisoning by psychotropic agents	Injury	0.283	588

29 principal diagnoses were chosen as diagnoses that had a weekend admission rate that was as close or closer to 2/7 as hip fracture in the full inpatient Medicare data. Weekend admission rates reported here are those in the main analysis sample, which is limited to ambulance transfers. Results are nearly identical when the broader category of "General symptoms" are excluded from the analysis. 2002-2008 Medicare Part A Claims Data

Table A2: Sample Means

Measure	Mean (1)	Standard Deviation (2)
Panel A. Quality Measures		
CMS 2005 Composite Process Score	0.819	0.067
CMS 2010 Composite Process Score	0.949	0.036
CMS 30-Day Readmission Rate	0.209	0.018
Non-Deferrable Condition Composite 30-Day Readmission Rate (Fixed Effect)	0.125	0.017
CMS 30-Day Mortality Rate	0.116	0.012
Non-Deferrable Condition Composite 30-Day Mortality Rate (Fixed Effect)	0.124	0.020
Non-Deferrable Condition Composite 365-Day Mortality Rate (Fixed Effect)	0.313	0.032
Panel B. Medicare Reimbursement Measures		
Hospital Avg. Log (Total Index Inpatient)	8.90	0.211
Hospital Avg. Log (Total 90 Day)	9.74	0.178
Hospital Avg. Log (90 Day Inpatient)	9.37	0.217
Hospital Avg. Log (90 Day Non-Inpatient)	7.49	0.319
Hospital Avg. Log (End of Life)	11.01	0.207

Table A3: OLS, First-Stage and 2SLS Estimates

		(1)	(2)	(3)	(4)
Panel A. First Stage	Amb. CMS 2010 Composite Process Score	0.542	0.541	0.542	0.542
		(0.026)**	(0.026)**	(0.026)**	(0.026)**
	Amb. CMS 2005 Composite Process Score	0.560	0.559	0.559	0.559
		(0.014)**	(0.014)**	(0.014)**	(0.014)**
	Amb. CMS 30-Day Readmission Rate	0.520	0.519	0.518	0.518
		(0.015)**	(0.015)**	(0.015)**	(0.015)**
	Amb. Non-Def. 30-Day Readm. Rate (FE)	0.420	0.419	0.420	0.420
		(0.012)**	(0.012)**	(0.012)**	(0.012)**
	Amb. CMS 30-Day Mortality Rate	0.510	0.510	0.509	0.509
		(0.014)**	(0.014)**	(0.014)**	(0.014)**
	Amb. Non-Def. 365-Day Mort. Rate (FE)	0.457	0.457	0.459	0.459
		(0.014)**	(0.014)**	(0.014)**	(0.014)**
Outcome: 365 Day Mortality					
Panel B. OLS	CMS 2005 Composite Process Score	-0.013	-0.011	-0.009	-0.010
		(0.002)**	(0.002)**	(0.002)**	(0.002)**
	CMS 2010 Composite Process Score	-0.020	-0.015	-0.013	-0.015
		(0.003)**	(0.003)**	(0.003)**	(0.003)**
	CMS 30-Day Readmission Rate	0.003	0.004	0.003	0.004
		(0.004)	(0.003)	(0.003)	(0.003)
	Non-Def. 30-Day Readm. Rate (FE)	0.020	0.019	0.017	0.018
		(0.004)**	(0.004)**	(0.004)**	(0.004)**
	CMS 30-Day Mortality Rate	0.014	0.015	0.015	0.017
		(0.003)**	(0.003)**	(0.003)**	(0.003)**
	Non-Def. 365-Day Mort. Rate (FE)	0.043	0.040	0.038	0.042
		(0.003)**	(0.003)**	(0.003)**	(0.003)**
Panel C. 2SLS	CMS 2005 Composite Process Score	-0.006	-0.004	-0.005	-0.007
		(0.016)	(0.015)	(0.015)	(0.014)
	CMS 2010 Composite Process Score	-0.032	-0.026	-0.029	-0.028
		(0.015)*	(0.014)	(0.013)*	(0.012)*
	CMS 30-Day Readmission Rate	-0.015	-0.011	-0.006	0.003
		(0.018)	(0.016)	(0.015)	(0.014)
	Non-Def. 30-Day Readm. Rate (FE)	0.022	0.016	0.021	0.023
		(0.019)	(0.017)	(0.016)	(0.015)
	CMS 30-Day Mortality Rate	0.003	0.005	0.008	0.009
		(0.016)	(0.015)	(0.014)	(0.013)
	Non-Def. 365-Day Mort. Rate (FE)	0.068	0.061	0.068	0.071
		(0.022)**	(0.020)**	(0.019)**	(0.017)**
Sample Size	331,793				
Outcome Mean	0.330				
Demographic Controls		No	Yes	Yes	Yes
Ambulance Controls		No	No	Yes	Yes
Comorbidity Controls		No	No	No	Yes

Source: 2002-2010 Medicare Claims
Each estimate is from a separate regression with diagnosis controls and ZIP fixed effects.
Quality Measures have been standardized by 2 standard deviations so they
can be interpreted like binary (low-to-high) indicators
* $p<0.05$; ** $p<0.01$

Table A4: Inpatient Spending

Outcome:	365-Day Mortality	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. OLS	Inpatient Spending	-0.027	-0.024	-0.030	-0.029	-0.032	-0.039	
		(0.007)**	(0.008)**	(0.008)**	(0.008)**	(0.007)**	(0.008)**	
	CMS 2010 Composite Process Score	-0.010						
		(0.003)**						
	CMS 2005 Composite Process Score		-0.015					
			(0.003)**					
	CMS 30-Day Readmission Rate			0.005				
				(0.004)				
	Non-Def. 30-Day Readm. Rate (FE)				0.017			
					(0.004)**			
	CMS 30-Day Mortality Rate					0.015		
						(0.003)**		
	Non-Def. 365-Day Mort. Rate (FE)						0.043	
							(0.003)**	
Panel B. 2SLS	Inpatient Spending	-0.174	-0.170	-0.192	-0.176	-0.173	-0.179	
		(0.059)**	(0.060)**	(0.061)**	(0.060)**	(0.059)**	(0.062)**	
	CMS 2010 Composite Process Score	-0.009						
		(0.015)						
	CMS 2010 Composite Process Score		-0.023					
			(0.013)					
	CMS 30-Day Readmission Rate			0.021				
				(0.016)				
	Non-Def. 30-Day Readm. Rate (FE)				0.027			
					(0.017)			
	CMS 30-Day Mortality Rate					0.002		
						(0.015)		
	Non-Def. 365-Day Mortality Rate (FE)						0.090	
							(0.018)**	
Outcome Mean		0.330						
Sample Size		331,793						

Source: 2002-2010 Medicare Claims
Each estimate is from a separate regression with full controls.
Quality Measures have been standardized by 2 standard deviations so they can be interpreted like binary (low-to-high) indicators
All reimbursement measures are hospital averages and are in logs.
FE=Empirical Bayes Fixed Effect
* p<0.05; ** p<0.01

Table A5: 90-Day Spending: Specific Conditions

Outcome: 365 Day Mortality		(1)	(2)	(3)
Panel A. 2SLS - AMI Sample	90-Day Inpatient Spending	-0.113		-0.112
		(0.139)		(0.149)
	90-Day Post Discharge Spending	0.307		0.334
		(0.180)		(0.191)
	90-Day Total Spending		-0.002	
			(0.014)	
	AMI 365-Day Mort. Rate (FE)			0.133
				(0.044)**
	Sample Size: 37,684			
Panel B. 2SLS - Pneumonia Sample	90-Day Inpatient Spending	-0.154		-0.171
		(0.122)		(0.135)
	90-Day Post Discharge Spending	0.006		0.027
		(0.107)		(0.116)
	90-Day Total Spending		-0.001	
			(0.012)	
	Pneumonia 365-Day Mort. Rate (FE)			0.133
				(0.040)**
	Sample Size: 51,297			

Source: 2002-2010 Medicare Claims

Within each panel, each column reports estimates from a separate regression with full controls.

All regressions include HSA fixed effects

Quality Measures have been standardized by 2 standard deviations so they can be interpreted like binary (low-to-high) indicators

All reimbursement measures are hospital averages and are in logs.

FE = Empirical Bayes Fixed Effect

* $p<0.05$; ** $p<0.01$