NBER WORKING PAPER SERIES

IDENTIFYING THE HEALTH PRODUCTION FUNCTION:
THE CASE OF HOSPITALS

John A. Romley
Neeraj Sood

Working Paper 19490
http://www.nber.org/papers/w19490

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2013

We are grateful to seminar participants at ASHE, Cal State Long Beach, NBER's Health Economics Program Meeting, RAND, Rice University, UC Riverside, and the University of Houston, and to Shin-Yi Chou and Jonathan Skinner for helpful comments. Raj Mehta and Zachary Wagner-Rubin provided excellent research assistance. We are indebted to the Agency for Healthcare Research and Quality (5R01-HS01854103) for financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w19490.ack

Identifying the Health Production Function: The Case of Hospitals
John A. Romley and Neeraj Sood
NBER Working Paper No. 19490
October 2013
JEL No. D24,I1,I12

ABSTRACT

Estimates of the returns to medical care may reflect not only the efficacy of more intensive care, but also unmeasured differences in patient severity or the productivity of health-care providers. We use a variety of instruments that are plausibly orthogonal to heterogeneity among providers as well as patients to analyze the intensity of care and 30-day survival among Medicare patients hospitalized for heart attack, congestive heart failure and pneumonia. We find that the intensity of care is endogenous for two out of three conditions. The elasticity of 30-day mortality with respect to care intensity increases in magnitude from -0.27 to -0.71 for pneumonia and from -0.16 to -0.33 for congestive heart failure, when we address the identification problem. This finding is consistent with the hypotheses that care intensity at hospitals tends to decrease with hospital productivity, or increase with unmeasured patient severity.

John A. Romley
Schaeffer Center for Health Policy and Economics
University of Southern California
3335 S. Figueroa St, Unit A
Los Angeles, CA 90089-7273
romley@healthpolicy.usc.edu

Neeraj Sood
Schaeffer Center for Health Policy and Economics
University of Southern California
3335 S. Figueroa Street, Unit A
Los Angeles, CA 90089-7273
and NBER
nsood@healthpolicy.usc.edu

1. Introduction

Understanding the marginal product of medical care is critical to ongoing debates about the value of U.S. health care. An array of studies, based on the Dartmouth Atlas of Health Care, find that U.S. regions that spend more on medical care have similar or poorer patient outcomes than areas that spend less (Skinner 2011). This evidence has suggested to some observers that medical spending can be reduced without adversely affecting patient outcomes (Fisher, Wennberg et al. 2003).

In contrast, recent studies of hospital care in health economics and other literatures have found that greater spending and intensity of care are associated with lower risk-adjusted mortality (Doyle 2005; Chandra and Staiger 2007; Card, Dobkin et al. 2009; Ong, Mangione et al. 2009; Almond, Doyle et al. 2010; Barnato, Chang et al. 2010; Kaestner and Silber 2010; Silber, Kaestner et al. 2010; Doyle 2011; Romley, Jena et al. 2011; Romley, Jena et al. 2013).

These studies of the returns to hospital spending have concerned themselves with the problem of unobserved patient heterogeneity. Health or severity of illness is notoriously difficult for researchers to measure precisely, and patients or providers almost certainly observe more about health or severity of illness than a researcher does. The resulting identification problem is that the intensity of medical care is confounded with unobserved (to the researcher) health status or severity of illness. Typically, the concern is that sicker patients receive more intensive care, and so the returns to medical care are biased downward. Thus, accounting for patient selection can increase the estimated returns to medical care.

One approach to dealing with this patient selection problem is to exploit discontinuities in the intensity of care (Doyle 2005; Card, Dobkin et al. 2009; Almond, Doyle et al. 2010). While this approach is compelling, its generality may be limited. Another approach is to focus on a group of patients for whom the selection problem is plausibly mild, for example, hospitalizations of out-of-state visitors with cardiac emergencies (Doyle 2011). Still another approach exploits the fact that consumers typically prefer to receive health care close to their homes, and uses distance to providers as an exogenous source of variation in intensity (McClellan, McNeil et al.

1994; Gowrisankaran and Town 1999; Geweke, Gowrisankaran et al. 2003; Chandra and Staiger 2007).

However, patient selection is not the only threat to identification. There may also be unobserved heterogeneity in terms of the producers of health care. To see this more clearly, consider a world where we could randomly assign patients to high-intensity versus low-intensity hospitals. The random assignment of patients to hospitals implies that patient heterogeneity, whether observed or unobserved, is not a concern. However, estimates of the returns to medical care might still be biased if hospitals with high intensity of care differ from low-intensity hospitals in other respects. For example, high-intensity hospitals might have more or less effective management practices, or differential expertise in using cost-saving or quality-enhancing technologies. Such hospital characteristics are hard to measure, influence hospital productivity, and could be positively or negatively correlated with intensity, resulting in biased estimates of the returns to medical care.

(Chandra 2013) provides a cogent discussion of this identification problem, noting that even with perfect risk adjustment, comparisons across regions of medical spending and outcomes are not conclusive for answering the question of whether spending can be reduced without affecting patient outcomes. He states that "*Different delivery systems and hospital referral regions vary in expertise, and consequently, the ability to transform measured inputs like hospital days, hospital spending, imaging and specialists' visits into outputs. The variations literature has overstated the benefits to medical spending if measured inputs and expertise are positively correlated (as they will be if they are complements), and understated the value of spending if spending more (i.e., using more hospitals days or physician visits) is a substitute for low expertise*."

This threat to the identification of the production function has also long been a matter of serious concern in applied microeconomics (Marschak and Andrews Jr 1944; Griliches and Mairesse 1998). It might be particularly salient in health care where past research has documented significant variation in expertise or productivity (Skinner, Staiger et al. 2006;

Chandra and Staiger 2007; Doyle, Ewer et al. 2010; Romley and Goldman 2011; Chandra 2013; Chandra, Finkelstein et al. 2013). What is needed to identify returns to medical care is variation in intensity of care that is orthogonal to unmeasured patient *and* production heterogeneity.

In this paper, we estimate the returns to medical care in the context of hospitals. Consistent with the literature on geographic variation in care, we measure intensity at the area level and estimate its effect on thirty-day survival among Medicare beneficiaries admitted to a hospital on an emergency basis for heart attack, congestive heart failure, and pneumonia. Heart-attack care has been widely studied (McClellan, McNeil et al. 1994; Picone, Sloan et al. 2003; Skinner and Staiger 2009; Doyle 2011), while the other conditions are important causes of hospital admissions and elderly mortality. Moreover, validated and transparent measures of inpatient mortality risk and hospital processes of care are available for each of these conditions.

To deal with unobserved patient and production heterogeneity, we appeal to a standard theory of hospital behavior that suggests instruments for care intensity. These instruments include insurer concentration, state corporate tax rates, wages, and market size. As we argue below, the intensity of care among hospitals should decrease with insurer concentration, corporate tax rates, and wages, but increase with market size.

We distinguish between in-state and out-of-state patients, because the problem of patient selection should be mild for out-of-state visitors (Doyle 2011). In ordinary least squares regressions, we find that the higher intensity of hospital care results in greater survival for in-state as well as out-of-state patients. Consistent with (Doyle 2011)'s finding, the estimated effect size is significantly larger for out-of-state patients with heart attacks.

When we instrument for hospital intensity, we can reject the exogeneity of hospital intensity for out-of-state patients admitted with congestive heart failure or pneumonia. The elasticity of thirty-day mortality with respect to hospital intensity increases in magnitude from -0.27 (OLS) to -0.71 (IV) for pneumonia, and from -0.16 (OLS) to -0.33 (IV) for congestive heart failure. We show that our instruments are strongly related to the intensity of hospital care and are plausibly exogenous. The results are also robust to a number of specification checks.

Altogether, our findings indicate that unobserved heterogeneity is substantial enough to confound the intensity-survival relationship for pneumonia and congestive heart failure, even among populations for whom the patient selection problem is relatively mild. The downward bias in the returns to hospital care under OLS is consistent with the hypotheses that care intensity decreases with hospital productivity, or increases with unobserved patient severity.

Our finding that the returns to hospital care are positive implies that broad-based reductions in hospital spending would likely have adverse effects on patient outcomes. However, our findings do not imply that there is no "wasteful" hospital spending in the U.S. It is certainly possible that there are large inefficiencies across or within areas, or even within particular hospitals (Chandra 2013). As a corollary, the impact of reductions in medical spending is likely to be context-specific. The impact might be influenced by where and how spending is reduced. For example, policies that aim to reduce spending by penalizing hospital readmissions might produce different outcomes than policies that aim to reduce spending through across-the-board cuts in hospital reimbursement.

The remainder of this paper is organized as follows: In Section 2, we review the threats to identification of the health production function and present our identification strategy for the case of hospitals. In section 3, we implement and assess our empirical approach with an analysis of intensity of care and thirty-day survival for Medicare beneficiaries admitted to the hospital. Section 4 concludes.

2. Identification Problems and Identification Strategy

2.1 Identification Problem

We consider the following production function for a health outcome:

$$\bar{H} \equiv E\left(H \middle| I, \mathbf{S}, s, A, a\right) = f(I, \mathbf{S}, s, A, a), \qquad (1)$$

in which H is health (the output), and $\bar{H}$ is its conditional expectation. Health is determined by the intensity of health inputs, measured by the index *I*. We abstract from issues of input aggregation, following the relevant literature (Fisher, Wennberg et al. 2003; Doyle 2011; Romley, Jena et al. 2011; Chandra, Finkelstein et al. 2013). Health is further determined by the initial severity of illness $\boldsymbol{S}$, a vector observed by the researcher. Unobserved health status is measured by *s*, while *A* and *a* measure observed and unobserved provider heterogeneity.

Our goal is to consistently estimate the health production function *f*, which is assumed known up to a parameter vector. The marginal product of intensity f_I is of particular interest, due to its importance to debates about the social value of health care.

There are two fundamental threats to the identification of the health production function. These threats arise from the sources of unobserved heterogeneity in the production function of equation (1). Loosely speaking, there are distinct threats from the consumers of health and from the producers of health. The consumer threat is that severity of illness drives the intensity of care. For example, sicker people often require, and are likely to receive, more intensive care. Insofar as severity is unobserved, the marginal product of intensity will tend to be understated.

The top panel of Figure 1 illustrates the problem of patient heterogeneity. Consider an increase in intensity from point A to point B. If the researcher observed severity perfectly (and there were no productivity differences), she could normalize each hospital's production function so that variation in intensity traced out variation in health outcomes along a common production frontier. In reality, a researcher almost certainly cannot observe severity perfectly. Absent an effective strategy for dealing with patient heterogeneity, a researcher is effectively comparing intensity and health levels across multiple production functions differentiated by patient severity. The lower production function in the figure corresponds to a hospital with sicker patients. If this hospital delivered more intensive care, a naïve comparison of points A and B would understate the marginal product of care intensity.

This patient selection problem has been a central concern of prior studies of the returns to health care. One group of studies employs regression-discontinuity designs that exploit clear breaks in intensity that are plausibly unrelated to severity. For example, (Almond, Doyle et al.

2010) analyzed treatment intensity and mortality among newborns just above and just below the threshold for very low birth weight. (Card, Dobkin et al. 2009) compared hospital intensity and mortality among patients just under the age of 65 to similar patients who are just over 65, and thus categorically eligible for Medicare coverage. Both studies found that higher intensity was associated with significantly lower mortality. While this evidence is compelling, regression-discontinuity designs avail themselves in a limited range of settings, and thus their results may not generalize to other contexts.[1]

Other studies focus on groups of patients who are relatively homogeneous. For example, (Doyle 2005) analyzed patients involved in automobile accidents, because medical care for these serious events is typically not discretionary. Patients with health insurance received more intensive treatment, and were substantially more likely to survive their accidents. (Doyle 2011) investigated hospital patients admitted for cardiac emergencies, with a particular focus on out-of-state visitors. The rationale is that heart-related emergency admissions represent serious health shocks for which patients have limited discretion in selecting hospitals [(Card, Dobkin et al. 2009) similarly focused on emergency hospitalizations]. Furthermore, out-of-state visitors comprise a small share of patients, and are therefore unlikely to affect the intensity decisions of hospitals. Another example of this kind of strategy is (Fisher, Wennberg et al. 2003), who measured spending among decedents on the grounds that illness severity is similar among those who die, while intensity for decedents is correlated with intensity among all patients. Still other studies instrument for the intensity of hospital care. For example, (McClellan, McNeil et al. 1994) used plausibly exogenous variation in distance between patients and hospitals with or without intensive cardiac care facilities to estimate the causal effect of intensity on mortality among heart-attack patients.

[1]Regression discontinuity designs can in principle account for provider heterogeneity by focusing on discontinuities in intensity of care within providers.

The producer threat to identification is that heterogeneity among providers may also determine the intensity of care. For example, hospitals with poor management or inferior doctors may have high costs per patient for the health outcomes that are achieved (Doyle, Ewer et al. 2010). In such situations, hard-to-measure differences among providers would lead to a downward bias in estimates of returns to medical care. Alternatively, hospitals that can achieve good outcomes for a given intensity of care may tend to supply more intensive care (Marschak and Andrews Jr 1944; Griliches and Mairesse 1998). If so, the returns to medical care can be overstated. In point of fact, evidence of productivity differences within the health-care sector is compelling (Skinner, Staiger et al. 2006; Chandra and Staiger 2007; Skinner and Staiger 2009; Weinstein and Skinner 2010; Romley and Goldman 2011; Chandra 2013; Chandra, Finkelstein et al. 2013).

The bottom panel of Figure 1 illustrates the problem of production heterogeneity for estimates of the returns to hospital care. Absent an effective strategy for dealing with production heterogeneity, a researcher necessarily compares multiple production functions differentiated by productivity. The higher production function corresponds to a hospital with higher productivity. If this hospital delivered less intensive care, a naïve comparison of points A and B would understate the marginal product of care. The threat to identification from productivity differences, or other aspects of provider heterogeneity, has received less attention than patient selection in the empirical literature on the returns to health care.

Strategies which are effective in dealing with patient heterogeneity may not be effective in dealing with production heterogeneity. For example, while out-of-state patients admitted to hospitals for heart-related emergencies are plausibly homogeneous in terms of severity, their hospitals may nevertheless differ in their productivity, resulting in systematic and confounded differences in care intensity.

2.2 Identification strategy

Our strategy is to appeal to a standard theory of hospital behavior to find instruments for the intensity of hospital care *I*. In a model developed by (Hodgkin and McGuire 1994), hospitals maximize utility based on profits and intensity of care. Intensity of care can directly affect utility because altruistic hospitals might care about the quality of care or health benefits enjoyed by patients. At the margin, hospitals balance the direct utility from intensity with its indirect utility through profits. The marginal effect of intensity on profits can be decomposed into two effects. First, increases in intensity reduce variable profits as the cost of care increases, given that prices are fixed in the model (as is the case for Medicare). Second, increases in intensity increase variable profits as higher intensity attracts more patients to the provider. Intensity essentially serves as an indicator of hospital quality and thus increases demand for the hospital.

Given this model, we posit several instruments for intensity of care.[2] First, higher input prices raise the cost of increased intensity. Second, greater market power among insurers will reduce hospital prices (Ho and Lee 2013), and thus reduce the return to attracting new patients through increased intensity. In prior work, (Picone, Sloan et al. 2003) used similar instruments for intensity; however, market power was measured on the basis of concentration among hospitals, which is plausibly correlated with unmeasured determinants of patient health outcomes (Kessler and McClellan 2000). In the spirit of (Hodgkin and McGuire 1994), we also use state corporate tax rates. Higher corporate tax rates reduce the financial returns to for-profit hospitals from attracting new patients, potentially reducing intensity. Moreover, not-for-profit hospitals might reduce intensity of care in response to higher corporate tax rates as they compete with for-profit hospitals in the same market.

Motivated by the literature on pharmaceutical research and development, we posit the market size of a hospital as another instrument for the intensity of its care. Drug developers have

[2] Largely outside of health economics, a number of alternative strategies for dealing the identification threat from unobserved productivity have been pursued, including dynamic panel models and control functions (Arellano and Bond 1991; Olley and Pakes 1996).

greater incentive to invest in R&D for new drugs or quality improvement in therapeutic markets with greater demand (Finkelstein 2003; Acemoglu and Linn 2004; Blume-Kohout and Sood 2013). In our context, intensity may entail costs that are fixed with respect to the number of patients treated, for example, the capital cost of diagnostic imaging technologies. If so, variable profits will increase with a hospital market's size, all else equal and assuming that price-cost margins are positive.[3] Effectively, the fixed cost of intensity can be spread across a larger number of patients.

To be valid, our instruments must satisfy two properties. First, they must be correlated with intensity of care. Second, they must be uncorrelated with unobserved determinants of patient survival. We present details of the empirical assessments of the validity of our instruments in the next section. To summarize, we show that the instruments are highly correlated with intensity of care. We also perform overidentification tests to check the validity of our multiple instruments. In addition, we focus on urban areas and use state-of-the-art (though necessarily imperfect) controls for patient health status and hospital productivity, to limit the possibility that our instruments might be correlated with unmeasured determinants of patient survival. Finally, we show that even if our instruments were correlated with unmeasured determinants of survival, the relationship between our instruments and unmeasured factors would have to be much stronger than the relationship with measured factors, in order to negate our findings.

3. Empirical Analysis

In this section we first present our empirical framework. We then describe our analyses of hospital intensity and survival and their results. Initially we deal with unobserved patient

[3]In a simple model, the marginal cost of a hospital stay is constant with respect to quantity of stays but increasing with respect to intensity. Doubling the size of the market then leaves variable profits, and the marginal utility of intensity through higher profits, unchanged on a per patient basis. Across all patients, marginal utility from profits doubles.

heterogeneity by focusing on patients for whom the selection problem should be relatively mild. We then address any remaining patient and production heterogeneity by instrumenting for hospital intensity.

3.1 Empirical framework

We analyze intensity of care and thirty-day survival among elderly Medicare fee-for-service beneficiaries admitted to the hospital on an emergency basis with a principal diagnosis of heart attack (acute myocardial infarction), congestive heart failure, or pneumonia. Each of these medical conditions is a common cause of hospitalization, and risk-adjusted mortality for each condition corresponds to an Inpatient Quality Indicator (IQI) developed and approved by the Agency for Healthcare Research and Quality (AHRQ) for the purpose of assessing hospital quality (Agency for Healthcare Quality and Research 2008).

Thirty-day survival / mortality has been a common outcome measure for health care researchers and stakeholders. For example, the Centers for Medicare and Medicaid Services (CMS) reports on thirty-day mortality for a number of conditions on its Hospital Compare website. Compared to inpatient mortality, thirty-day mortality is believed to be less susceptible to manipulation by hospitals, who might, for example, seek to discharge patients who are likely to die.[4] Medicare Denominator files from CMS report date of death, validated against Social Security Administration records. We measure thirty-day survival by linking the Denominator files to Medicare Provider Analysis and Review Files on hospitalizations in the 50 U.S. states and the District of Columbia over the period 2003-2007.

We measure intensity of hospital care by the logarithm of the costs of a hospital stay. To do so, we apply cost-to-charge ratios from Medicare Impact Files to total charges, including physician fees for services provided in the hospital. To make intensity comparable across areas,

[4]The AHRQ IQIs focus on inpatient mortality, because inpatient mortality is typically available in administrative data on hospital discharges (Agency for Healthcare Quality and Research 2008).

we adjust costs according to the Medicare Hospital Wage Index. We convert costs to 2011 dollars using the market basket for inpatient services from CMS. Following the literature on the returns to health care (Fisher, Wennberg et al. 2003), we aggregate adjusted costs to the area level. Specifically, we average costs within Hospital Service Areas (HSAs), as (Doyle 2011) did. HSAs are defined by the Dartmouth Atlas of Health Care based on cities or towns, and the zip codes that supply Medicare patients to the hospitals located there.(Wennberg and Cooper 1996) We restrict our analysis to urban areas, defined by an average population density of at least 1,000 persons per square mile residing within 2.5 miles of hospitals in an HSA in the 2000 Census, as some covariates and instruments are measured at the level of metro areas. Our analysis includes 1,517 HSAs. Intensity is measured at the condition level based on decedent cases, following (Fisher, Wennberg et al. 2003) and (Doyle 2011).

We estimate linear regressions of the following form:

$$S_i = \beta_0 + \beta_I I_{HSA_i} + \beta_S R_i + \beta_S R_i \bar{R}_{HSA_i} + \beta_A A_{HSA_i} + \mathbf{W}_{\mathbf{h_i}} \boldsymbol{\beta}_{\mathbf{W}} + \mathbf{X_i} \boldsymbol{\beta}_{\mathbf{X}} + \boldsymbol{\theta}_{\mathbf{s_i}} + \boldsymbol{\delta}_{\mathbf{t_i}} + \varepsilon_i, \quad (2)$$
$$\varepsilon_{ih} = \beta_s s_i + \beta_a a_{HSA_i} + \omega_i$$

in which S_i equals one if patient i treated survived thirty days and zero otherwise, and I_{HSA_i} is the logged intensity of care (just described) at hospitals in the patient's area. In sensitivity analysis, we explore a more flexible specification of intensity.

We account for a variety of other factors in equation (2). A key strength of our analysis is the availability of a patient-specific measure of inpatient mortality risk R_i from the AHRQ IQIs. These measures are based on validated and published risk adjustment models, which incorporate not only age and sex, but also All Patient Refined Diagnosis Related Groups (APR-DRGs) and APR-DRG risk-of-mortality subclasses defined by patient diagnoses and medical procedures from the discharge records (Agency for Healthcare Quality and Research 2008). We interact patient-level risk with average risk within areas ($\bar{R}_{HSA_i}$), to mitigate bias arising from geographic variation in the diagnosis of co-morbid conditions (Song, Skinner et al. 2010; Welch, Sharp et al. 2011).

In addition, we attempt to measure productivity within a hospital's HSA, A_h. In the spirit of (Skinner and Staiger 2009), we consider technologies and processes of care that represent best practices, but do not contribute meaningfully to costs / intensity, such as the provision of aspirin to heart-attack patients upon arrival at a hospital.[5] Table 1 shows process-of-care measures from the CMS Hospital Compare website. These measures are specific to care for heart attacks, congestive heart failure, and pneumonia, and endorsed by the Hospital Quality Alliance. Percutaneous intervention (PCI) for heart attack is an intensive treatment (McClellan, McNeil et al. 1994), and is therefore excluded from productivity[6]; in a sensitivity analysis, we exclude measures which do not correspond to medical treatments delivered during the hospital stay (for example, the provision of discharge instructions.) For each condition studied, we perform a factor analysis of the selected Hospital Compare measures, and estimate an index of hospital productivity using a single-factor model, which is then aggregated to the HSA. The May, 2005 reporting period is used to limit behavioral responses by hospitals, and to ensure adequate variation in the measures across hospitals.

To control for confounders of the intensity-survival relationship, equation (2) includes patient age and its square, gender, and race /ethnicity (white, black, Hispanic, and other.) Additional patient- and hospital-level covariates motivated by a seminal study of health spending and mortality are also used (Fisher, Wennberg et al. 2003). These covariates include the volume of patients with the condition at a hospital in quartiles (Luft, Hunt et al. 1987), a hospital's teaching status (membership in the Council of Teaching Hospitals) from the American Hospital Association Annual Surveys (American Hospital Association 2006), the number of Charlson-Deyo co-morbidities that a patient had (Deyo, Cherkin et al. 1992), and socio-demographic

[5](Skinner and Staiger 2009) study the Cooperative Cardiovascular Project (CCP), which involved chart reviews of 160,000 eldery heart-attack patients in 1994/1995. Medical records are more reliable than the administrative data used here. In terms of technologies, Skinner and Staiger consider aspirin and beta blocker use, as well as reperfusion therapy. At the time of the CCP, these technologies were not commonly viewed as quality measures. For heart-attack patients, we consider quality measures based on aspirin and beta blocker use.

[6]Fibrinolytic medication within 30 minutes of arrival was not used because this measure was reported for less than twenty percent of patients in our primary analysis, as reported in Table 1.

characteristics of the patient's zip code from the 2000 Census. Zip code-level Census variables include median annual household income; average annual Social Security income; the percentage of the population that was living below the poverty level; was employed; had less than a high school education; were white, black, Hispanic or other; was single; was elderly and institutionalized; and was elderly and non-institutionalized, but with various disabilities. To deal with differences in cost of living across areas, household income in 1999 is deflated with the ACCRA Cost of Living Index by linking zip codes and cost indices at the level of the Combined Statistical Area. Income is categorized by approximate quartiles (<$30,000; $30,000 - $34,999; $35,000 - $44,999; $45,000+), with an additional category for missing values.

Table 2 presents the summary statistics for our analysis sample. The analysis also includes fixed effects for the states in which patients resided.(Doyle 2011) Year fixed effects allow for secular trends in survival.

3.2 Initial analysis

The first specification of Table 3 reports estimation results for the full samples of patients admitted to the hospital on an emergency basis. Initially, we measure intensity by the log cost of a hospital stay at the *patient* level. A one log-point increase in intensity is associated with a 2.5 percentage point *decrease* in 30-day survival for patients with congestive heart failure, and a 5.2 percentage point decrease in survival for pneumonia patients (specification 1).

When patient mortality risk, the productivity index of effective low-cost processes of care, and other covariates are included in the analysis (specification 2), these parameter estimates become positive, but are of negligible magnitude (0.007 and 0.004, respectively). These findings are consistent with the body of evidence from Dartmouth researchers and others that greater medical spending is frequently not associated with better quality of care [e.g., (Fisher, Wennberg et al. 2003)].

In the case of heart attack, a one log-point increase in intensity is estimated to increase 30-day survival by 7.8 percentage points in the regression that includes only intensity. When patient mortality risk, hospital productivity, and other factors are incorporated, the magnitude of the effect becomes more positive (9.5 percentage points.)

In specification 3 and all further analyses, intensity is measured by costs among decedents at each hospital, but averaged to the *HSA* level. For each condition, the intensity parameter estimate is positive, and statistically and economically significant. The estimates for congestive heart failure and pneumonia increase to 0.0190 and 0.028, respectively. Heart attack is again different, with its intensity parameter estimate decreasing in magnitude (from 0.095 to 0.041.)

3.3 Dealing with unobserved patient heterogeneity: Ordinary least squares analysis of in-state versus out-of-state patients

Following (Doyle 2011), we distinguish between patients treated at hospitals in their state of residence, and out-of-state patients. (In a sensitivity analysis, we deal with urban areas that cross state lines by redefining out-of-state patients admitted to hospitals within 25 miles of home as in-state.) As Table 2 showed, roughly 5% of patients were treated out-of-state, regardless of condition; our intensity measure includes all patients.

Table 4 compares the results between in-state and out-of-state patients from regressions that include all covariates (specification 3 of Table 3.) For heart attack, the intensity parameter estimate is larger for out-of-state patients than for in-state patients, 0.053 versus 0.039, and this difference is statistically significant at the 10% level. For pneumonia, the out-of-state estimate is again larger (0.033 versus 0.027), though not significantly different. Altogether, these finding lend support to (Doyle 2011)'s strategy of focusing on patients for whom the selection problem is relatively mild.

3.4 Dealing with residual production and patient heterogeneity: Instrumental variables

Nevertheless, there may be residual heterogeneity. We are particularly concerned about production heterogeneity, because a focus on patients whose health status is relatively homogeneous does not deal with productivity differences across hospitals.

3.4.1 Instruments

Appealing to the theory of hospital behavior discussed in Section 2, we use a variety of instruments for the intensity of hospital care. First, we measure insurer concentration using a Herfindahl-Hirschman index. Insurer concentration has been publicly reported based on HMO and PPO enrollment as of January, 2005, using data from HealthLeaders and InterStudy.(American Medical Association 2007) We link concentration for 313 metropolitan standard areas in 44 states to hospitals in our sample, and take averages by HSA. Second, we measure corporate tax rates by averaging the maximum marginal rate across the analysis period for each state, linking to hospitals within states, and averaging across hospitals within HSAs; these data are available from the Tax Foundation. Third, we measure area wages using the Medicare Hospital Wage Index, averaging across years and hospitals within HSAs.

Finally, we measure market size using the 2000 Census to quantify the population residing in zip codes within 5 miles of each hospital, and take the average across hospitals within each HSA. Hospital geo-coordinates are reported in AHA Annual Surveys (where unavailable, we use a hospital's zip code and its corresponding geo-coordinates from the 2000 Census). Individuals aged 65 or older are excluded from the population counts, because elderly individuals may be more likely to decide where to live based on their own health and area health-care resources.

The results from first-stage regressions of intensity on the instruments (and all other covariates) are shown in Table 5. As predicted by our theory of hospital behavior, intensity generally decreases with insurer concentration, corporate tax rates, and area wages, while increasing with market size. Because the instruments (and intensity) are logged, these results represent elasticities. Thus, the elasticity of care intensity for congestive heart failure with respect to the corporate tax rate is -0.087, while the elasticity of pneumonia intensity with respect to the insurer concentration index is -0.092.

3.4.2 Main results of instrumental-variables analysis

Table 6 shows the instrumental variables results for the out-of-state patient samples. The instruments have sizable first-stage F statistics, which range from 15.07 for pneumonia to 38.98 for congestive heart failure. The IV intensity parameters are positive and significant for all three conditions. For heart attack, we cannot reject the hypothesis that intensity is exogenous. By contrast, there is substantial evidence for the endogeneity of the intensity of care intensity for congestive heart failure and pneumonia.

To interpret these findings, we report elasticities of 30-day mortality with respect to area-level intensity in Table 7. For out-of-state patients with heart attacks, we could not reject the exogeneity of intensity, and the elasticity based on the OLS results (Table 3) is -0.32. In the case of pneumonia, the elasticity increases in magnitude from -0.27 to -0.71 when we instrument for endogenous intensity. Similarly, for congestive heart failure, the elasticity increases from -0.16 to -0.33 when we instrument for intensity.

The results are qualitatively unchanged when we redefine out-of-state patients admitted to hospitals within 25 miles of home as in-state, and exclude measures which are not related to medical care during the hospital stay from our productivity index, as shown in Appendix Tables 1 and 2.

3.4.3 Instrument validity

We assess the validity of our instrumental variables analysis in a number of ways. First, note that our empirical model is overidentified. As Table 6 shows, standard overidentification tests for model validity could not be rejected.

Next, we explore whether the exogenous determinants of survival vary systematically with our instrumental variables.(McClellan, McNeil et al. 1994; Altonji, Elder et al. 2005) To do so, we first create an index of our multiple IVs by using variation in our instruments from a regression of whether an area had above or below median intensity on the IVs and all other exogenous factors (that, all other covariates except intensity). Areas with an above-median value of the resulting index are characterized as "high" in terms of our IVs (and vice versa). Next, we want to understand whether areas with high versus low values of the IV index value differ systematically with the observed exogenous factors. If we find a strong correlation between observed determinants of survival and the IV index, that could suggest that our instrument is also correlated with unobserved determinants of survival, consequently biasing our instrumental variable results. To conduct this indirect test, we regress patient survival on the exogenous determinants included in the empirical model, and compare predicted survival rates across areas with low versus high values of the IV index.

The results are shown in Table 8. For congenital heart failure, the exogenous survival rate is 89.9% in areas with a low value of the IV index versus 90.5% in areas with high instrument values, and this difference is statistically significant. Because high values of the instrument index are associated with greater intensity of care, the higher survival rate based on exogenous factors in areas with high instrument values raises some concern that unobserved factors may also be more favorable to survival in these areas. It is somewhat reassuring that the difference in exogenous survival (0.6 percentage points) is only about one quarter of the magnitude of the estimated effect of greater intensity based on a high value of the instrument index (2.4 percentage points).[7] For pneumonia, exogenous survival is again higher in areas with high instrument values, but this difference is not statistically significant.

---

[7]Here we performed IV analyses of survival on the exogenous factors and an indicator variable for above-median spending, with an indicator for an above-median value of the instrument index as the instrument for intensity. These analyses had substantial power, with first-stage *F* statistics in excess of 1,000.

3.4.4 Heterogeneity in returns to hospital intensity

Next, we explore heterogeneity in the returns to intensity. To begin with, we stratify the analysis of equation (2) by the health status of patients, as measured by the risk of inpatient mortality based on the AHRQ IQI models.[8] The results are shown in Table 9. As one might expect, for both heart failure and pneumonia, “sick” patients — with mortality risk above the median value — experience a significantly higher return to the intensity of care. For pneumonia, the return to intensity is more than three times greater for sick patients than for health patients. In the case of heart failure, we cannot reject the hypothesis that healthy patients do not benefit from greater intensity. We also stratify the analysis by the teaching status and ownership of hospitals, and find that the returns to intensity could not be distinguished by hospital type, as shown in Tables 10 and 11. Finally, we allow for a more flexible relationship between intensity and survival, by including the square of logged intensity as an additional (endogenous) covariate; there is no evidence of such a relationship (results available from authors upon request.)

4. Conclusions

Our concern here is with the identification of production functions for health. In this context, there are two distinct threats to identification, unobserved patient heterogeneity and unobserved production heterogeneity. To assess and deal with these threats to identification, we analyzed the intensity of area-level hospital care and thirty-day survival among Medicare beneficiaries admitted on an emergency basis for heart attack, congestive heart failure and pneumonia.

We first distinguished between patients treated at hospitals in their state of residence, and out-of-state patients. The problem of patient selection should be relatively mild for visitors (Doyle 2011). We found in ordinary least squares regressions that the marginal product of the

[8]In the analyses of heterogeneity by patient health, area-level intensity and its instruments are measured based on the full sample of patients. For the analyses by hospital type, area-level intensity and its instruments are measured among hospitals of the relevant type.

intensity of medical care was positive, leading to increased survival for both in-state and out-of-state patients. The estimated effect was significantly larger for out-of-state patients with heart attacks.

We then addressed any remaining confounding of the returns to hospital care among out-of-state patients. To do so, we appealed to a theory of hospital behavior to identify potentially exogenous determinants of intensity. Our instruments were strongly related to the intensity of hospital care, and we were able to reject the exogeneity of intensity for two of the conditions studied. In this analysis, the elasticities of 30-day mortality with respect to care intensity more than doubled in magnitude for patients with pneumonia and congestive heart failure. Our assessments of the validity of our identification strategy did not raise significant concerns.

Our findings indicate that there is substantial unmeasured heterogeneity in the production of health by hospitals. In particular, the estimated returns to intensity of hospital care showed evidence of downward bias in a number of the analyses. Such bias is consistent with a negative correlation between intensity and hospital productivity. For example, (Doyle, Ewer et al. 2010) found that residents affiliated with a lower-ranked medical school substitute diagnostic tests for clinical judgment, and obtain similar health outcomes at higher cost. The direction of bias that we found is also consistent with a positive correlation between intensity of care and unobserved patient severity.

This work has significant implications for understanding the production of health. It shows that effective empirical strategies for identifying returns to medical care might vary with the clinical circumstances of patients. For example, we found that focusing on emergency hospitalizations of out-of-state visitors was sufficient to deal with endogeneity of intensity for heart attacks, but not for pneumonia or heart failure. These latter conditions have relatively low acuity, and as a consequence it seems likely there is greater scope for selection based on health status. It is noteworthy that the direction of bias was inconsistent with the admission of relatively healthy patients in high-intensity areas. The endogeneity of intensity for these conditions could also be a result of provider heterogeneity. For example, the use of effective, low-cost treatments and care processes could vary more across providers and areas for patients with heart failure, than for patients with heart attack.

In terms of public policy, our finding that the returns to hospital care are positive suggest that broad-based reductions in hospital spending could have adverse effects on patient outcomes. However, it is important to note some caveats for interpreting our findings. First, our findings do not indicate that there are no avenues for curtailing spending without hurting patient outcomes. In other words, it is possible that some spending is "wasteful," and reducing such spending will not hurt patient outcomes. The impacts of reductions in medical spending are likely to be context-specific, and to be influenced by where and how spending is reduced. Second, our study is motivated by productivity differences; however, it does not shed light on the root causes of such differences. Understanding whether such differences arise from failures of management or governance, from differences in specialization or skills, or from inadequate public policies can shed light on the extent to which productivity differences can be eliminated. How difficult or easy is it to improve the productivity of low-performing hospitals or regions, and how this can be done, are largely unknown. Finally, our study followed the literature in focusing on variation in intensity across areas, and it is certainly possible that there are large inefficiencies within regions or even within hospitals. Identifying such inefficiencies and effective policy solutions are important directions for future research.

References

Acemoglu, D. and J. Linn (2004). "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." Quarterly Journal of Economics **119**(3): 1049-1090.

Agency for Healthcare Quality and Research (2008). "Guide to Inpatient Quality Indicators."

Almond, D., J. J. Doyle, et al. (2010). "Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns." The Quarterly Journal of Economics **125**(2): 591-634.

Altonji, J. G., T. E. Elder, et al. (2005). "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." The Journal of Human Resources **40**(4): 791-821.

American Hospital Association (2006). 2005 Annual Survey. Chicago, IL, Health Forum, LLC.

American Medical Association (2007). Competition in Health Insurance: A Comprehensive Study of U.S Markets.

Arellano, M. and S. Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." The Review of Economic Studies **58**(2): 277-297.

Barnato, A. E., C. C. Chang, et al. (2010). "Is survival better at hospitals with higher "end-of-life" treatment intensity?" Med Care **48**(2): 125-132.

Blume-Kohout, M. E. and N. Sood (2013). "Market Size and Innovation: Effects of Medicare Part D on Pharmaceutical Research and Development." J Public Econ **97**: 327-336.

Card, D., C. Dobkin, et al. (2009). "Does Medicare Save Lives?" The Quarterly Journal of Economics **124**(2): 597-636.

Chandra, A. (2013). Economics Meets the Geography of Medicine.

Chandra, A., A. Finkelstein, et al. (2013). Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector. National Bureau of Economic Research Working Paper. Cambridge, MA, National Bureau of Economic Research.

Chandra, A. and D. O. Staiger (2007). "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." Journal of Political Economy **115**(1): 103-140.

Deyo, R. A., D. C. Cherkin, et al. (1992). "Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases." Journal of Clinical Epidemiology **45**(6): 613-619.

Doyle, J. J. (2005). "Health Insurance, Treatment and Outcomes: Using Auto Accidents as Health Shocks." Review of Economics and Statistics **87**(2): 256-270.

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

Doyle, J. J., Jr., S. M. Ewer, et al. (2010). "Returns to physician human capital: evidence from patients randomized to physician teams." Journal of health economics. **29**(6): 866-882.

Finkelstein, A. (2003). Health Policy and Technological Change: Evidence from the Vaccine Industry. Cambridge, MA, National Bureau of Economic Research.

Fisher, E. S., D. E. Wennberg, et al. (2003). "The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care." Annals of Internal Medicine **138**(4): 288-298.

Geweke, J., G. Gowrisankaran, et al. (2003). "Bayesian Inference for Hospital Quality in a Selection Model." Econometrica **71**(4): 1215-1238.

Gowrisankaran, G. and R. J. Town (1999). "Estimating the quality of care in hospitals using instrumental variables." Journal of Health Economics **18**(6): 747-767.

Griliches, Z. and J. Mairesse (1998). Production functions: the search for identification, Cambridge University Press.

Ho, K. and R. Lee (2013). Insurer Competition and Negotiated Hospital Prices. National Bureau of Economic Research Working Paper. Cambridge, MA, National Bureau of Economic Research.

Hodgkin, D. and T. G. McGuire (1994). "Payment levels and hospital response to prospective payment." Journal of Health Economics **13**(1): 1-29.

Kaestner, R. and J. H. Silber (2010). "Evidence on the Efficacy of Inpatient Spending on Medicare Patients." Milbank Q **88**(4): 560-594.

Kessler, D. P. and M. B. McClellan (2000). "Is Hospital Competition Socially Wasteful?" Quarterly Journal of Economics **115**(2): 577-615.

Luft, H. S., S. S. Hunt, et al. (1987). "The volume-outcome relationship: practice-makes-perfect or selective-referral patterns?" Health Services Research **22**(2): 157-182.

Marschak, J. and W. H. Andrews Jr (1944). "Random simultaneous equations and the theory of production." Econometrica, Journal of the Econometric Society: 143-205.

McClellan, M., B. J. McNeil, et al. (1994). "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables." JAMA **272**(11): 859-866.

Olley, G. S. and A. Pakes (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry." Econometrica **64**(6): 1263-1297.

Ong, M. K., C. M. Mangione, et al. (2009). "Looking Forward, Looking Back: Assessing Variations in Hospital Resource Use and Outcomes for Elderly Patients With Heart Failure." Circulation: Cardiovascular Quality and Outcomes.

Picone, G. A., F. A. Sloan, et al. (2003). "Does Higher Hospital Cost Imply Higher Quality of Care?" The Review of Economics and Statistics **85**(1): 51-62.

Romley, J. and D. P. Goldman (2011). "How Costly Is Hospital Quality? A Revealed-Preference Approach." Journal of Industrial Economics.

Romley, J. A., A. B. Jena, et al. (2011). "Hospital Spending and Inpatient Mortality: Evidence from California: An Observational Study." Annals of Internal Medicine **154**(3): 160-167.

Romley, J. A., A. B. Jena, et al. (2013). "Spending and Mortality in US Acute Care Hospitals." Am J Manag Care **19**(2): e46-54.

Silber, J. H., R. Kaestner, et al. (2010). "Aggressive Treatment Style and Surgical Outcomes." Health Services Research **45**(6 Pt 2): 1872-1892.

Skinner, J. (2011). Causes and Consequences of Regional Variations in Health Care. Handbook of Health Economics. M. V. Pauly, T. G. McGuire and P. P. Barros. Amsterdam, Elsevier. **2:** 45-93.

Skinner, J. and D. Staiger (2009). Technology Diffusion and Productivity Growth in Health Care. Cambridge, MA, National Bureau of Economic Research.

Skinner, J. S., D. O. Staiger, et al. (2006). "Is technological change in medicine always worth it? The case of acute myocardial infarction." Health Affairs **25**(2): w34-47.

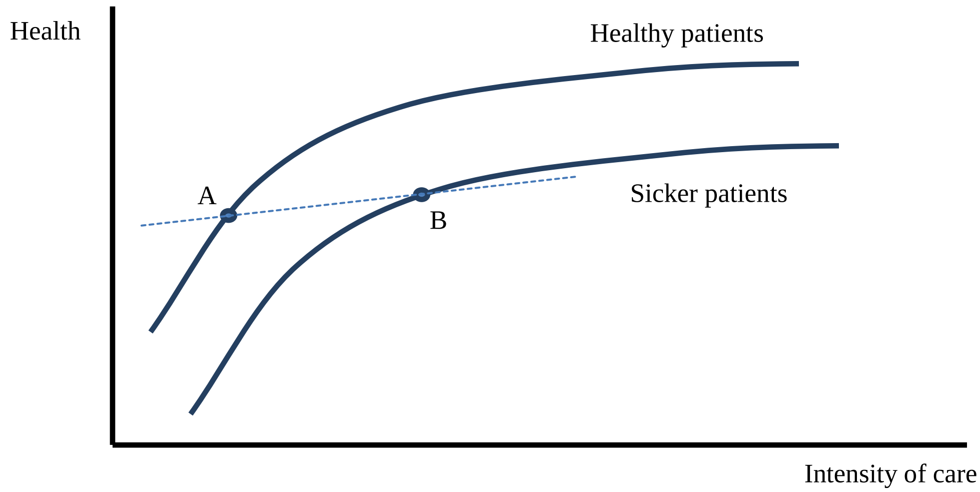
Song, Y., J. Skinner, et al. (2010). "Regional variations in diagnostic practices." N Engl J Med **363**(1): 45-53.

Weinstein, M. and J. Skinner (2010). "Comparative Effectiveness and Health Care Spending--Implications for Reform." New England Journal of Medicine.

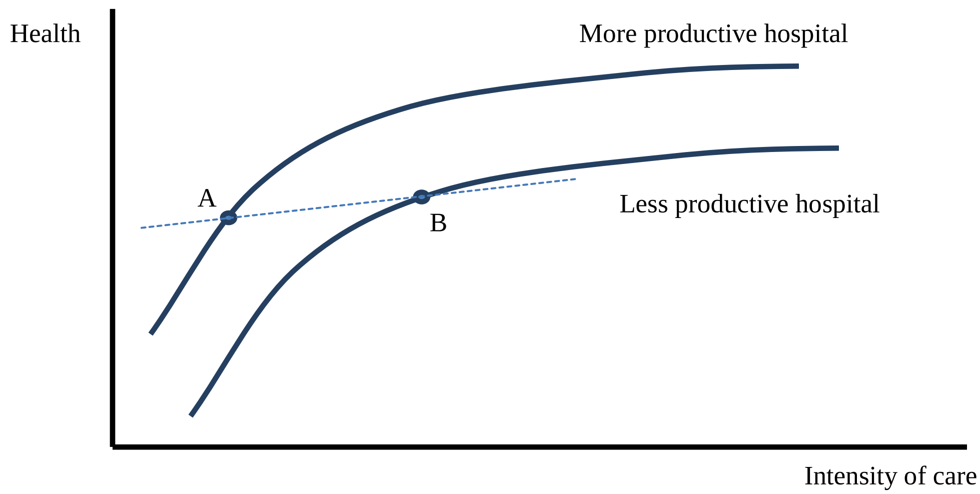
Welch, H. G., S. M. Sharp, et al. (2011). "Geographic variation in diagnosis frequency and risk of death among Medicare beneficiaries." JAMA **305**(11): 1113-1118.

Wennberg, D. E. and M. M. Cooper (1996). Dartmouth Atlas of Health Care. Chicago, IL, American Hospital Publishing, Inc.

Figure 1: Threats to Identification of a Health Production Function



Marginal Product of Intensity of Hospital Care Understated,
If Patients with High Unobserved (to the Researcher) Severity Receive More Intensive Care



Marginal Product of Intensity of Hospital Care Again Understated,
If Hospitals with High Unobserved (to the Researcher) Productivity Provide Less Intensive Care

Table 1:
Process of Care Measures in May, 2005 Release of Hospital Compare

Measure	Mean	Availability	Used in productivity Index
Heart Attack			
ACE Inhibitor or ARB for LVS Dysfunction	78.7%	99.7%	Yes
Aspirin at Arrival	94.7%	99.9%	Yes
Aspirin at Discharge	93.4%	99.9%	Yes
Beta Blocker at Arrival	89.8%	99.9%	Yes
Beta Blocker at Discharge	91.3%	99.9%	Yes
Fibrinolytic Medication Within 30 Minutes Of Arrival	33.7%	14.1%	
PCI Within 90 Minutes Of Arrival	37.0%	23.4%	
Smoking Cessation Advice/Counseling	82.5%	79.0%	Yes
Congestive Heart Failure			
Discharge Instructions	50.3%	87.6%	Yes
Smoking Cessation Advice/Counseling	69.9%	88.0%	Yes
Evaluation of LVS Function	88.3%	99.8%	Yes
ACE Inhibitor or ARB for LVS Dysfunction	75.7%	99.8%	Yes
Pneumonia			
Assessed and Given Pneumococcal Vaccination	43.3%	99.7%	Yes
Initial Antibiotic(s) within 6 Hours After Arrival	68.7%	99.8%	Yes
Oxygenation Assessment	98.9%	99.8%	Yes
Smoking Cessation Advice/Counseling	64.0%	88.9%	Yes
Initial Emergency Room Blood Culture Performed Prior To Administration Of First Hospital Dose Of Antibiotics	82.3%	88.5%	Yes

Notes: Availability is the percentage of patients in analysis samples (summarized in Table 3) for whom the measure is reported. Mean is conditional on applicability of measure to patient, and availability of measure among applicable patients. "Used in productivity index" refers to primary analysis. Acronyms: ACE = angiotensin-converting-enzyme, ARB = angiotensin receptor blocker, LVS = left ventriculay systolic, and PCI = percutaneous intervention.

Table 2:
Summary Statistics for Patient Samples

Variable	AMI	CHF	Pneumonia
Patients, #	654,767	1,425,296	1,144,745
30-day survival, %	83.20 (37.39)	89.64 (30.47)	87.90 (32.61)
Cost per stay (2011 dollars)	18002 (5554)	15476 (4989)	14785 (4169)
Inpatient mortality risk, %	11.13 (10.10)	3.68 (4.08)	4.79 (5.14)
Age, years	79.3 (8.5)	80.3 (8.5)	80.4 (8.4)
Male, %	48.0 (50.0)	41.2 (49.2)	44.4 (49.7)
White, %	86.4 (34.2)	80.1 (39.9)	86.3 (34.3)
Black, %	8.9 (28.5)	15.0 (35.7)	8.7 (28.2)
Hispanic, %	1.8 (13.4)	2.3 (15.1)	2.1 (14.4)
Charlson-Deyo co-morbidities, #	2.4 (1.1)	2.4 (1.0)	1.6 (1.0)
Teaching hospital, %	24.1 (42.7)	20.9 (40.7)	17.5 (38.0)
Hospital volume, annual cases	194.1 (154.4)	363.4 (255.7)	254.9 (181.2)
In-state hospital, %	94.2 (23.4)	95.7 (20.4)	95.2 (21.4)
Year	2004.9 (1.4)	2005.0 (1.4)	2004.9 (1.4)
ZIP code demographics			
Median household income, $	40472 (14700)	39361 (14723)	40453 (14917)
Below poverty line, %	11.2 (8.2)	12.3 (9.1)	11.5 (8.4)
Social Security income, mean, $	11515 (1477)	11325 (1508)	11450 (1478)
White, %	73.0 (25.6)	69.1 (28.4)	72.8 (25.9)
Black, %	11.9 (19.4)	15.1 (22.9)	11.6 (18.9)
Hispanic, %	9.8 (16.0)	10.5 (17.3)	10.2 (16.6)
Single, %	43.3 (9.2)	44.7 (9.8)	43.5 (9.3)
Less than high school, %	19.0 (10.9)	20.2 (11.5)	19.2 (11.2)
Employed, %	59.6 (9.2)	59.2 (8.9)	59.9 (8.9)
ZIP code health characteristics among population 65 and older			
Institutionalized, %	4.5 (5.1)	4.5 (5.1)	4.7 (5.2)
Physical disability, %	28.4 (6.8)	29.1 (6.9)	28.6 (6.9)
Mental disability, %	10.6 (4.2)	11.0 (4.3)	10.8 (4.2)
Sensory disability, %	13.8 (3.8)	14.0 (3.7)	14.0 (3.9)
Self-care disability, %	9.5 (3.7)	9.9 (3.8)	9.6 (3.7)
Home-bound disability, %	20.5 (5.7)	21.3 (5.9)	20.7 (5.8)

Notes: Standard deviations are in parentheses. AMI is acute myocardial infarction, or heart attack. CHF is congestive heart failure.

Table 3:
Regressions of 30-Day Survival on Area Hospital Intensity, by Patient Sample

Specification number	1	2	3
Controls for patient, hospital and area factors	N	Y	Y
Heart attack			
Log of patient costs / intensity	0.078*** (0.001)	0.095*** (0.001)	
Log intensity, average within area (HSA) among decedents			0.041*** (0.003)
Congestive heart failure			
Log of patient costs / intensity	-0.025*** (0.001)	0.007*** (0.001)	
Log intensity, average within area (HSA) among decedents			0.019*** (0.002)
Pneumonia			
Log of patient costs / intensity	-0.052*** (0.001)	0.004*** (0.001)	
Log intensity, average within area (HSA) among decedents			0.028*** (0.003)

Notes: Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical
5% level, and *** at the 1% level.

Table 4:
Regressions of 30-Day Survival on Area Hospital Intensity, In-State Patients Versus Out-of-State Patients

Patient sample	*In state*	*Out of state*
Heart attack	0.039*** (0.003)	0.053***^ (0.008)
Congestive heart failure	0.019*** (0.002)	0.016*** (0.006)
Pneumonia	0.027*** (0.003)	0.033*** (0.006)

Notes: Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level. ^ indicates that out-of-state estimate differs from in-state estimate at the 10% level, ^^ at the 5% level, and ^^^ at the 1% level.

Table 5:
Elasticities of Area Hospital Intensity with Respect to Instruments from First-Stage Regressions for Out-of-State Patients

	Wage index	*Insurer concentration index*	*Corporate tax rate*	*Market size*
Heart attack	-0.442***	-0.064	-0.057*	0.099***
	(0.094)	(0.040)	(0.034)	(0.016)
Congestive heart failure	-0.181**	-0.029	-0.087***	0.139***
	(0.081)	(0.037)	(0.028)	(0.012)
Pneumonia	-0.204**	-0.092***	-0.100***	0.070***
	(0.087)	(0.035)	(0.028)	(0.013)

Notes: Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level. Market size is number of persons under age 65 residing within 2 miles of hospital, averaged within Hospital Service Areas.

Table 6:
Instrumental-Variables Regressions
of 30-Day Survival on Area Hospital Intensity,
Out-of-State Patients

	First stage F statistic	*Hospital intensity parameter estimate*	*Endogeneity test, p value*	*Overidentification test, p value*
Heart attack	18.07	0.055** (0.025)	0.92	0.33
Congestive heart failure	38.98	0.034*** (0.011)	0.065	0.15
Pneumonia	15.07	0.086*** (0.023)	0.016	0.33

Notes: Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 7:
Elasticity of 30-Day Mortality with Respect to Area Hospital Intensity,
Out-of-State Patients

Method	*OLS*	*IV*
Heart attack	-0.32	-0.33
Congestive heart failure	-0.16	-0.33^
Pneumonia	-0.27	-0.71^^

Note: ^ indicates the IV estimate differs from OLS estimate at the 10% level,
^^ at the 5% level, and ^^^ at the 1% level.

Table 8:
Amount of Selection on Unobservables Relative to Selection on Observables Required To Attribute the Entire Effect of Above-Median Hospital Intensity to Selection Bias

Condition	Exogenous survival rate, by IV index			$\hat{\beta}_I$
	Below median index	*Above median index*	*Difference*	
CHF	0.899	0.905	0.006*** (0.002)	0.024** (0.012)
Pneumonia	0.890	0.893	0.003 (0.002)	0.080*** (0.022)

Notes: CHF is congestive heart failure. Exogenous survival rate is predicted survival from regression with intensity excluded. Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level. $\hat{\beta}_I$ is instrumental variables estimate of above-median intensity on survival, based on above-median value of fitted instrument index from first-stage regression.

Table 9:
Instrumental-Variables Regressions by Patient Health,
Out-of-State Patients

	All	*Healthy*	*Sick*
Congestive heart failure	0.034*** (0.011)	0.002 (0.013)	0.064***^^^ (0.018)
Pneumonia	0.086*** (0.023)	0.042** (0.022)	0.126***^^ (0.039)

Notes: Healthy patients are above the median with respect to AHRQ predicted inpatient survival. Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level. ^ indicates that sick estimate differs from healthy estimate at the 10% level, ^^ at the 5% level, and ^^^ at the 1% level, based on bootstrapped test statistic of stratified analyses by patient health status.

Table 10:
Instrumental-Variables Regressions by Teaching Status,
Out-of-State Patients

	All	*Teaching*	*Non-Teaching*
Congestive heart failure	0.034*** (0.011)	0.025 (0.024)	0.041*** (0.014)
Pneumonia	0.086*** (0.023)	0.056 (0.059)	0.092*** (0.026)

Notes: Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level. ^ indicates that estimate differs from teaching estimate at the 10% level, ^^ at the 5% level, and ^^^ at the 1% level, based on bootstrapped test statistic of stratified analyses by teaching status.

Table 11:
Instrumental-Variables Regressions by Hospital Ownership,
Out-of-State Patients

	All	*For Profit*	*Not for Profit*	*Public*
Congestive heart failure	0.034*** (0.011)	-0.002 (0.026)	0.058*** (0.014)	0.017 (0.039)
Pneumonia	0.086*** (0.023)	0.076* (0.041)	0.137*** (0.033)	0.103 (0.086)

Notes: Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level. ^ indicates that estimate differs from not-for-profit estimate at the 10% level, ^^ at the 5% level, and ^^^ at the 1% level, based on bootstrapped test statistic of stratified analyses by hospital ownership.

Appendix Table 1:
IV Regressions of Out-of-State Patients,
Excluding Patients within 25 Miles of Out-of-State Hospital

	First stage F statistic	*Hospital intensity parameter estimate*	*Endogeneity test, p value*	*Overidentification test, p value*
Congestive heart failure	28.89	0.053*** (0.015)	0.014	0.59
Pneumonia	13.14	0.114*** (0.028)	0.002	0.48

Notes: Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Appendix Table 2:
IV Regressions of Out-of-State Patients,
with Narrowed Set of Measures in Productivity Index

	First stage F statistic	*Hospital intensity parameter estimate*	*Endogeneity test, p value*	*Overidentification test, p value*
Congestive heart failure	41.83	0.036*** (0.010)	0.046	0.08
Pneumonia	15.84	0.086*** (0.023)	0.019	0.31

Notes: Standard errors clustered at HSA level, and are reported in parentheses. * indicates statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level. Analyses exlcludes measures in Table 1 for discharge instructions and smoking cessation / counseling.