Supplemental Material

PAYER MIX AND FINANCIAL HEALTH DRIVE HOSPITAL QUALITY: IMPLICATIONS FOR VALUE-BASED REIMBURSEMENT POLICIES

Matthew Manary, Richard Staelin, William Boulding, & Seth W. Glickman

Manary, Staelin, Boulding, and Glickman, Duke University Fuqua School of Business; Glickman, University of North Carolina School of Medicine. Corresponding author's e-mail: swglick@med.unc.edu

Methods and Analysis

Methods

Formation of Variables and Standardizing Measures

Many of our measures at the population level exhibited trends over time. For example, from 2005 to 2010, adherence to clinical guidelines and patient experiences increased across the population of hospitals. In addition, measures we wished to consolidate had different intercept values even if no trending over time occurred. To net out time trends and scale-associated differences, we standardized measures by subtracting the average for all hospitals within a year and then dividing by the standard deviation for all hospitals in that same year:

$$z_{h,t} = \frac{x_{h,t} - \overline{X_t}}{\hat{\sigma}_t},$$

where x is a measure, h designates the hospital, t is the year, and $z_{h,t}$ is the standardized score. This process was repeated for each year (2005–2010). Thus, when standardized, scores for each measure within a year averaged 0 with a standard deviation of 1. Therefore, our analysis of standardized measures is a relative comparison, not an absolute measure.

Financial Health

Our analysis period is by calendar year, yet only approximately half of the financial statements were aligned to calendar years. Consequently, we converted deviating financial statements to a calendar year basis by looking at the two financial statements that covered the period of interest and then weighting the underlying financial data by the appropriate number of months associated with the calendar year of interest. For instance, for a hospital with a financial year reporting from July to June, a given year's calendar-adjusted value was weighted 1/2 from the subsequent fiscal year and 1/2 from the current fiscal year.

To construct the single measure of financial health, referred to as *DuPont* in our study, we first calculated each of the three financial measures: current ratio, return on assets, and operating margin. To put them each on a common scale, we standardized each of the three measures within each year, leaving each hospital with three different standardized measures per year. We then added all three together within each year for each hospital to make a single measure of financial performance. For ease of interpretation, our last step was to standardize the combined measure within each year. Therefore, DuPont scores for the population of hospitals for each year have a mean of 0 and a standard deviation of 1.

Clinical Adherence

From 2005 through 2010, the number of process care performance measures Hospital Compare tracked varied slightly per year and per clinical area studied, but typically there were 18 measures available from the Centers for Medicare & Medicaid Services (CMS) in these three clinical areas (7 for acute myocardial infarction, 4 for heart failure, and 7 for pneumonia). Average clinical adherence composite scores, which are the percentage of time recommended procedures are followed, exhibited time trends in each clinical area. These time-related shifts were eliminated by standardizing scores within each year for each clinical area. Therefore, each hospital for each year had three standardized quality measures of clinical care. We note these measures correlated highly with one another and loaded on a similar factor (factor loadings of 0.71, 0.86, and 0.73). We thus created a single aggregate measure of the hospital's generalized clinical adherence by summing the three standardized measures and once again standardized by year so that each hospital has a single measure of overall clinical care per year. Each year's distribution of scores across the population of hospitals has a mean of 0 and standard deviation of 1.

Patient Experience

To generate a single annual hospital value for overall patient satisfaction, we combined responses for the following two hospital specific questions: "How do you rate the hospital overall?" and "Would you recommend the hospital to friends and family?" The Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) database reported the total number of patients surveyed and the percentage of patients who responded to the different levels of the question rather than reporting on the actual average scores. That is, the two overall satisfaction questions the Hospital Compare provides has three levels (not 10), that is, a satisfaction rating of "Low" (scores of 1–6), "Medium" (scores of 7–8), and "High" (scores of 9–10). To transform those groups into a numeric overall value, we multiplied the percentage of patients who responded to the given level by the numeric values of 0, 0.5, and 1 for Low, Medium, and High, respectively. This process generates scores between 0 and 1, where 1 indicated everyone gave a "High" response and 0 indicated everyone gave a "Low" response. After adding these two measures for each hospital and each year, for ease of interpretation, the yearly patient experience scores across the population of hospitals are again standardized with a mean of 0 and a standard deviation of 1.

Outcomes

The hospital-level 30-day risk-standardized readmission and mortality rates for acute myocardial infarction, heart failure, and pneumonia were obtained directly from Yale University's Center for Outcomes Research & Evaluation and provide hospital rates on an individual calendar year basis. CMS provides these same risk-adjusted measures on the Hospital Compare website; however, they are bundled in three-year increments, making longitudinal analysis difficult. The risk-adjustment process controls a particular hospital's outcome rates for patient demographics (gender and age), cardiovascular condition (such as chronic atherosclerosis and arrhythmia), and comorbidity (such as dementia and senility and iron deficiency and other/unspecified anemia and blood disease). Unlike clinical adherence and patient experiences, risk-adjusted outcomes were largely unchanged between 2005 and 2010, although mortality and readmission rates for acute myocardial infarction showed a significant decrease (p = .001 and p = .03, respectively). Pneumonia mortality and heart failure readmission and mortality rates showed no improvement, although all the signs were increasing (p = .41, p = .16, and p = .14, respectively), while pneumonia readmission rates actually significantly increased over the time horizon (p = .02). Because clinical area mean values differed and some clinical areas had changes in mean values over time, we standardized each value within each year, resulting in six fallible measures of outcome per hospital per year all on the same scale.

After conducting exploratory factor analysis, it appeared that readmission and mortality may represent two different dimensions of quality. However, a comparison of our base model (see Table S7) using the two different measures were not meaningfully different (Fisher's Z = 1.13, p = .26). Consequently, we decided to use a formative model of quality (versus a reflective model) by combining all six standardized measures into a single measure of outcome for each hospital and each year. Again, we standardized these values within each year for ease of interpretation. As further justification for combining these two measures, we note that combining measures that potentially represent two different factors into a single measure would increase the combined measure's noise. This would make it harder to produce significant results in our regressions. We did not find such an occurrence.

Analysis Approaches

Modeling Unobserved Variables

We acknowledge that it is possible for the hospital's performance, structure, and processes to be affected by factors not explicitly contained in our model formulation. For example, it is well known that an organization's culture and managerial skills can affect the firm's performance, structure, and processes. However, most analyses, including ours, do not contain such organizational measures but instead relegate them to the error term. We are not willing to assume that the effects of such unobserved factors are random over time, but instead we believe they are "sticky," that is, they tend to persist over time, especially within an organization. Ignoring such persistent factors results in autocorrelated errors, a condition observed in our data if we do not control for such unobserved firm specific factors.

The implication of autocorrelated errors in a linear models setting is that the estimates of the standard errors are biased, which, in turn, affects any hypothesis testing coming from the regression analysis. Perhaps more seriously, if the unobserved factors are not only correlated with the dependent variable (a condition that holds by assumption) but also one or more of the independent variables, then not only are the standard errors biased but the bias also occurs for any of the estimated coefficients.

As seen in Figure 1, many of the estimating equations contain process and structure measures as independent variables. Given our belief that unobserved factors such as culture and managerial expertise will affect many if not all of these independent variables as well as the dependent variable, we need to remove these unobserved factors from our analyses or control for them. We do this by postulating that (and then testing for) unobserved factors could be fixed or gradually change over time. We start with the latter assumption by postulating the following model:

$$y_{i,t} = e_{i,t} \,,$$

where

$$e_{i,t} = \rho * e_{i,t-1} + \beta * x_{i,t} + n_{i,t}.$$

In words, the dependent variable $Y_{i,t}$ is a function of an error term that is composed of last period's observed state, that is, $e_{i,t\text{-}1}$, as well as this period's observed factors and an unobserved random error, $n_{i,t}$. For example, the hospital's observed performance ($y_{i,t}$) is a function of last period's observed performance (which contains both last period's unobserved and observed factors) as well as a new unobserved error term. In this way, unobserved factors such as culture persist over time, albeit with a declining influence.

Mathematically, this translates into the following estimation model:

$$y_{i,t} = \rho * y_{i,t-1} + \beta * x_{i,t} + n_{i,t},$$

where ρ captures the persistence of all unobserved and observed factors.

We next postulate that some unobserved factors are fixed, that is, do not change over time. We test this assumption within our state-dependent model. Specifically, we do this by first ρ-differencing to remove the fixed effect from our model. We then use the Hausman test (Hausman, 1978) to compare our estimates with and without controlling for the fixed effects. We find no indication that these unobserved fixed effects render any of our coefficients to be inconsistent. Thus we only control for persistent but not fixed unobserved variables by always including a lagged dependent variable in our estimation models.

Further Tests for Alternative Explanations of the Results

Although we tested for unobserved fixed effects and controlled for some aspects of unobserved variables that cut across equations (e.g., management expertise) via the state-dependent structure within each equation, other unobserved factors such as macroeconomic shocks could correlate across several of our equations. For example, a significant drop in macroeconomic conditions may simultaneously reduce the number of privately insured patients (that is, lower employment levels) and/or hurt hospitals' financial positions through lower returns on endowments or fewer gifts, and patients may experience poorer outcomes as a result of being admitted with more serious complications as a result of postponing care expenditures. Although our outcome and patient experience measures are already risk-adjusted on the basis of diagnosed comorbidity at admission and through self-reported health levels, our financial performance and privately insured patient levels are not controlled for economic shocks. Therefore, our equation relating financial health and payer coverage could be subject to error. To test for contemporaneously correlated alternative explanations, we instrumented payer coverage on the basis of our equation relating payer coverage to demographics and retested the financial health and payer coverage relationship through a two-stage least squares approach. We note patient age and ethnicity are not changed by economic shocks, but those demographics are highly correlated with payer coverage. We found, after instrumenting payer coverage, that there is still a significant relationship between financial health and payer mix, even while controlling for autocorrelated errors ($\beta = 0.213$, $p < .0001$). Therefore, by testing for unobserved fixed effects (Hausman specification; Boulding, 1990), serially correlated errors (Jacobson, 1990), state dependence (Jacobson, 1990), and areas of greatest potential simultaneous contemporaneous shocks, we believe we have addressed most alternative hypotheses regarding our results.

Estimation

The models presented in Tables S2–S7 were estimated using iteratively reweighted least squares (IRLS). This approach allows us to treat outliers in our data without excluding the hospital. More generally, IRLS is a robust regression technique that downweights large residual values (Holland & Welsch, 1977). For our specification, we followed the weighting function recommended by Beaton and Tukey (1974) with Myers's (1990) recommended tuning constant and measure of error computed as 1.5 * Median (|least square residuals|). The results from this technique were very similar to running ordinary least squares but using an outlier removal process recommended by the National Institute of Standards and Technology, as well as running weighted least squares (WLS) where the weights were based on bed count.

The reported Granger-causality statistics were based on the aforementioned WLS models rather than IRLS as (a) the model results are very similar (the only differences being that the underrepresented minority coefficient to payer mix is significant at the $p = .061$ level and the relationship between financial health with equipment investment was insignificant for WLS) and (b) the weighted observations for WLS are the same for the main and nested models allowing the *F* test comparison, whereas the weighted observations with IRLS can differ between the main and nested models.

Structural Equations and Estimation Equations

Let Y_1 = Patient Insurance Coverage

Y_2 = Financial Health of Hospital

Y_3 = Capital Investments

Y_4 = Clinical Adherence

Y_5 = Patient Experience

Y_6 = Quality Index (outcomes)

URM = Percent of Underrepresented Minority Patients

Age = Percentage of Patients 60 years or older

Then our system of equations is as follows:

$$Y_1 = a_1 + b_1*URM + c_1*Age + d_1*\text{Controls} + \varepsilon_1$$

$$Y_2 = a_2 + b_2 * Y_1 + c_2 * \text{Controls} + \varepsilon_2$$

$$Y_3 = a_3 + b_3 * Y_2 + c_3 * \text{Controls} + \varepsilon_3$$

$$Y_4 = a_4 + b_4 * Y_2 + c_4 * \text{Controls} + \varepsilon_4$$

$$Y_5 = a_5 + b_5 * Y_2 + c_5 * \text{Controls} + \varepsilon_5$$

$$Y_6 = a_6 + b_6 * Y_3 + c_6 * Y_4 + d_6 Y_5 + c_6 * \text{Controls} + \varepsilon_6$$

where the ε_is capture all unobserved effects including fixed effects, serially correlated factors, and contemporaneous shocks, and Controls represents the vector of the following control variables: licensed beds, teaching hospital status, ownership (e.g., investor, government, nonprofit), and presence of 24-hour emergency services.

As discussed, to control for unobserved variables, we found it necessary to account for serially correlated factors associated with state dependence. We do this by introducing a lagged dependent variable into each of our structural equations (Jacobsen, 1990). The coefficients on these lagged variables should be interpreted as an estimate of the degree of state dependence. Note, however, that the coefficients on the other variables are still those found in our structural equations. It is these variables that we are most concerned with.

Improvement Program Effect on Low-Performing Hospitals

Some have argued disproportionate share reimbursement programs and the Improvement scoring category in CMS's Hospital Value-Based Performance Program (HVBP) levels the field for poorer performing hospitals that take on a large case load of government-program-insured patients. However, none of the 13 Fiscal Year 13 HVBP improvement scores were correlated with net income (taken from the Medicare Cost Report), and with the exception of doctor communication improvement from HCAHPS, the remaining 12 improvement scores were either uncorrelated (9) or positively correlated (3) with operating margin (that is, more profitable hospitals actually had greater improvement scores). It is important to note that improvement scoring is not based on hospitals' placement in measure quartiles but on relative achievement and improvement over time.

Data, Analyses, and Results

Preliminary Results and Overview of Analysis Approach

Our primary objective is to identify the links between a hospital's patient population and its quality of care and specifically whether these relationships are mediated through the financial health of the hospital. We begin our analysis by first testing if the percentage of underrepresented minorities has a direct association with the three performance measures that CMS uses in its pay-for-performance programs. (Note that for outcomes and patient experiences, CMS controls for age.) We do this by running three regressions where the dependent variables are clinical adherence, patient experience, and hospital outcomes. We find that the percentage of underrepresented minorities has a significantly negative relationship with hospital patient experience, clinical adherence, and outcomes (all ps < .0001). Although we do not prescribe any direct causal link between racial composition and these downstream quality measures, the finding is in concert with the multiple studies reported in the main article. We explore why we see such relationships by sequentially estimating the relationships of Figure 1.

To test our patient-demographic-to-quality-of-care framework, we estimate a series of separate linear models that take into consideration the longitudinal aspects of our data, thereby allowing us to assess the veracity of the causal linkages shown in Figure 1. Our linear models are discussed in the following order: (1) a hospital's patient insurance coverage mix as a function of its patient demographics (e.g., insurance coverage mix is the dependent variable and the patient demographics are the independent variables); (2) a hospital's financial health as a function of its patient insurance coverage mix; (3)–(5) patient experiences, clinical adherence, and investment in equipment each separately as a function of the hospital's financial health; and (6) hospital outcomes as a simultaneous function of its patient experiences, clinical adherence, and investment in equipment. The unit of analysis is the hospital for a calendar year, and all regressions include hospital controls. Each of these six equations can be thought of as our hypothesized relationships, that is, our belief of the true relationship between the dependent variable and the relevant independent variables. Similar to Bazzoli et al. (2008), we found our dependent measures were influenced by an autocorrelated unobservable variable, which we controlled for by using the lagged dependent measure as another independent variable in all our models (Jacobson, 1990). Note, however, that this lagged variable is not part of our hypothesized relationship but only there as a control. We used JMP Version Pro9 (SAS Institute, Cary, North Carolina) and SAS Version 9.2 to carry out all analyses; two-tailed tests with $\alpha = .05$ were established as the level of significance.

Testing Figure 1 Relationships and Discussion of Results

In Table S1, we provide a comparison of our California hospital sample with the national population of hospitals. Statistical tests show the hospitals in our sample were larger and had lower clinical adherence for pneumonia, higher mortality rates for pneumonia, and lower patient satisfaction, while all other measures of ownership makeup, clinical adherence, and hospital outcomes were not found to be significantly different. Perhaps more important, our sample has a wide dispersion on all the variables allowing us to estimate relationships.

Table S1: Characteristics of the Study Hospitals: Median (2.5%, 25%, 75%, 97.5%). Year = 2007

	Study Hospitals (*n* = 265)	National Hospitals (*n* = 3,451)
Hospital characteristics		
Number of beds, mean***	197 (31, 116, 356, 717)	168 (14, 42, 230, 651)
Hospital type, proportion		

Teaching	0.05	0.07
Investor Governed Hospital	0.24	0.21
Nonprofit Governed Hospital	0.55	0.61
Government/University Hospital	0.21	0.18
On-site 24-hour Emergency Services	0.95	0.94
Patient mix (by percentage)		
Over 60 years old	0.39 (0.12, 0.30, 0.47, 0.74)	
Black, Hispanic, & Native American	0.33 (0.04, 0.17, 0.52, 0.90)	
Payer Private Coverage	0.28 (0.03, 0.19, 0.39, 0.63)	
Technical Care Scores		
Clinical Adherence: Heart Attack	0.94 (0.64, 0.89, 0.97, 1.00)	0.94 (0.50, 0.88, 0.97, 1.00)
Clinical Adherence: Heart Failure	0.87 (0.40, 0.76, 0.93, 0.99)	0.85 (0.23, 0.74, 0.92, 0.99)
Clinical Adherence: Pneumonia**	0.89 (0.56, 0.82, 0.93, 0.98)	0.91 (0.69, 0.86, 0.94, 0.98)
Interpersonal Care Scores		
Overall Satisfaction**	0.77 (0.60, 0.73, 0.81, 0.88)	0.79 (0.62, 0.75, 0.83, 0.92)
Outcomes		
Heart Attack (Mortality)	0.16 (0.14, 0.16, 0.17, 0.18)	0.16 (0.14, 0.16, 0.17, 0.19)
Heart Attack (Readmission)	0.20 (0.19, 0.20, 0.20, 0.21)	0.20 (0.19, 0.20, 0.20, 0.22)
Heart Failure (Mortality)	0.11 (0.09, 0.10, 0.12, 0.14)	0.11 (0.09, 0.10, 0.12, 0.14)
Heart Failure (Readmission)	0.25 (0.22, 0.24, 0.25, 0.28)	0.25 (0.22, 0.24, 0.26, 0.28)
Pneumonia (Mortality)**	0.12 (0.09, 0.11, 0.13, 0.16)	0.11 (0.09, 0.10, 0.12, 0.15)
Pneumonia (Readmission)	0.18 (0.16, 0.18, 0.19, 0.21)	0.18 (0.16, 0.18, 0.19, 0.21)
Financial Measures		

Annual Revenue ($ millions)	133 (11, 57, 277, 999)	
Current Ratio	1.57 (0.38, 0.99, 2.30, 4.88)	
Operating Margin†	0.01 (-0.40, -0.06, 0.06, 0.16)	0.00 (-0.30, -0.06, 0.05, 0.23)
Return on Assets	0.04 (-0.44, -0.02, 0.09, 0.30)	
Percentage Change in Equipment	0.07 (-0.36, 0.03, 0.14, 0.61)	

** $p = .01$. *** $p < .0001$.

†The national sample is based on the Medicare Cost Report and is non-GAAP , meaning it does not conform to a category of Generally Accepted Accounting Principles.

Tables S2, S3, S6, and S7 present the results of the regressions associated with Figure 1 (in the main article) after controlling for hospital characteristics and autocorrelated errors. Note that in most of the tables, we do not show the coefficients for the hospital control variables, although they are included in the estimation. The reader should interpret the coefficient on the lagged dependent variable as a measure of the size of the unobserved autocorrelated error term.

The results of our first equation, which regress the percentage of privately insured patients against the percentage of patients that are (a) underrepresented minorities and (b) over 60 years old on the percentage of privately insured patients, are found in Table S2. Both measures have significant and negative coefficients ($p < .0001$ and $p < .0001$, respectively), indicating that hospitals with high levels of underrepresented minorities populations have fewer privately insured patients (i.e., our measure of patient insurance coverage).

Table S2: Patient Demographic Impact on a Hospital's Privately Insured Mix

Independent Measures	Estimate	Standard Error	*t* Value	*p* Value
Over 60 years old	-0.016	0.003	-5.20	<.0001
Black, Hispanic, & Native American (%)	-0.009	0.002	-4.26	<.0001
Payer Private Coverage (-1)	0.966	0.003	340.89	<.0001

Table S3 displays the relationship between the percentage of privately insured patients and our measure of higher financial performance ($p < .0001$).

Table S3: Percentage Covered as a Determinant of Financial Health

Independent Measures	Estimate	Standard Error	*t* Value	*p* Value
Payer Private Coverage	0.224	0.044	5.07	<.0001
DuPont (-1)	0.815	0.0084	96.45	<.0001

Tables S4 and S5 present the standard mediation test results. We find payer private coverage fully mediates the association between underrepresented minority patient composition and age with the financial health of the hospital (Sobel $z = -3.15$, $p = .002$). In addition, the percentage of privately insured patients Granger-causes financial performance in the subsequent period ($F_{1,1194} = 14.7$, $p = .0001$). This latter point highlights the potentially complex and long-lasting effect payer coverage has on a hospital's financial health and indirectly its ability to provide quality care both contemporaneously and into future periods.

Table S4: Patient Demographic as a Determinant of Financial Health

Independent Measures	Estimate	Standard Error	*t* Value	*p* Value
Over 60 years old	-0.097	0.064	-1.53	.127
Black, Hispanic, & Native American (%)	-0.113	0.040	-2.85	.005
DuPont (-1)	0.845	0.010	87.14	<.0001

Table S5: Payer Mix Mediating Patient Demographic as a Determinant of Financial Health

Independent Measures	Estimate	Standard Error	*t* Value	*p* Value
Over 60 years old	0.005	0.065	0.08	.934192
Black, Hispanic, & Native American (%)	0.0003	0.045	0.01	.993893
Payer Private Coverage	0.252	0.057	4.41	.006
DuPont (-1)	0.843	0.010	88.31	<.0001

Table S6 presents the results where financial health is the independent variable and the three quality measures of care as shown in Figure 1 of the main article are the dependent measures. We find significant results for clinical performance and the change in equipment investment, although patient experiences are not significantly correlated with contemporaneous financial health (p = .19; we note that when a second lag for HCAHPS scores is included, which is statistically significant and the same number of lags used in the Granger causality test, financial health is significant at the p = .05 level). As before, we test for Granger causality and find lagged financial health has a Granger-causal relationship with patient experience as measured by HCAHPS scores, investment in equipment, and clinical performance ($F_{1,\,412}$ = 16.7, p < .0001; $F_{1,\,949}$ = 4.9, p = .027; and $F_{1,\,1061}$ = 5.8, p = .016, respectively). Thus, it appears that a hospital's financial health has widespread impact on firm behavior and structure.

Table S6: Coefficient Estimates for Relation Between DuPont With Three Dependent Measures: Adherence to Clinical Guidelines, Patient Experience, and Infrastructure Equipment Investment

Dependent Measure	Estimate of DuPont Coeff.	Standard Error	*t* Value	*p* Value
Clinical Adherence	0.063	0.017	3.71	.0002
Patient Experience	0.012	0.009	1.306	.195
Equipment Investment	0.0101	0.003	3.76	.0002

Finally, we find clinical adherence and patient experiences are, in turn, significantly correlated with better hospital outcomes (where lower outcome measures are better; see Table S7). That is, hospital-level increases in adherence to clinical guidelines (p = .003) and positive patient experiences (p < .0001) are associated with better hospital-wide outcomes, even after controlling for the effect of the other factors (including investment in equipment) and autocorrelated errors.

Table S7: Predictors of Hospital-Wide Quality Index (smaller values indicate better outcomes)

Effect	Estimate	Standard Error	*t*-Value	*p* Value
Intercept	0.0219	0.106	0.208	.8438
Outcomes (-1)	0.306	0.020	15.42	<.0001
Patient Experience	-0.169	0.024	-7.04	<.0001
Clinical Adherence	-0.045	0.0156	-2.93	.00354
Equipment Investment	-0.020	0.033	-0.60	.5507

Licensed Beds	-0.001	0.00014	-4.93	<.0001
24-Hour On-Site Emergency Services	0.1659	0.100	1.652	.09968
Government Hospital*	0.063	0.055	1.15	.251099
Investor Governed Hospital*	-0.079	0.054	-1.47	.1421
Teaching Hospital	0.273	0.093	2.938	.00354

*Nonprofit and church hospitals are the reference group.

Limitations

Our study has several possible limitations. First, using a data set collected solely from the California hospital system potentially limits the generalizability of some of our findings. Unfortunately, some needed information, such as patient demographics, payer coverage, and audited financial measures, were not available for our national data set, thereby precluding us from testing our model on a wider set of hospitals. However, our sample of California hospitals appears to be consistent with the national pool in both makeup and performance across multiple areas (see Table S1). Furthermore, we found no differences in the relationship between outcomes and patient experiences and clinical adherence between the Californian and non-Californian hospitals.

Second, because this is an observational study rather than a controlled experiment, we relied on a conceptual model of the process and statistical tests to evaluate the hypothesized relationships. And, as with most models, there may be alternative explanations of our results. However, to reduce that risk, we ensured our models were either unbiased by—or corrected for—many possible types of errors or unobserved factors.

Although Figure 1 of the main manuscript implies causality, our state-dependent models only provide correlational associations over time. Consequently, we further test for the veracity of the sequential relationships in our hierarchy through tests for mediation and causality. We use Baron and Kenny's (1986) mediation test to understand whether a given construct operates directly or indirectly on another downstream construct. We test for causality using the methodology proposed by Granger (1969). This methodology uses past observations of the dependent variable as a control and then looks to see if an independent variable provides additional information—that is, causes the dependent variable after the controls.

Additional References

Barocas, D. A., Alvarez, J., Koyama, T., Anderson, C. B., Gray, D. T., Fowke, J. H., . . . Penson, D. F. (2014). Racial variation in the quality of surgical care for bladder cancer. *Cancer, 120,* 1018–1025.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Beaton, A.E, J.W. Tukey (1974). Fitting of Power-Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data, *Technometrics,* 16(2), 147-185

Boulding, W. (1990). Unobservable Effects and Business Performance: Do Fixed Effects Matter? *Marketing Science*, 9(1); 88-91.

Hausman, J.A. (1978). Specification Tests in Econometrics, *Econometrica,* 46(6), 1251-1271

Holland, P.W., & Welsch, R.E. (1977). Robust Regression Using Iteratively Reweighted Least Squares. *Communications in Statistics Part A-Theory and Methods*, 6(9), 813-827

Jacobson, R. (1990). Unobservable effects and business performance. *Marketing Science, 9,* 74–85.

McCutcheon, B. A., Chang, D. C., Marcus, L. P., Inui, T., Noorbakhsh, A., Schalhorn, C., . . . Talamini, M. A. (2014). Long-term outcomes of patients with nonsurgically managed uncomplicated appendicitis. *Journal of the American College of Surgeons, 218,* 905–913.

Myers, R. (1990). *Classical and Modern Regression With Applications Duxbury Advanced Series in Statistics and Decision Sciences*. Boston: PWS-Kent Publishing Company.

Neighbors, C. J., Rogers, M. L., Shenassa, E. D., Sciamanna, C. N., Clark, M. A., & Novak, S. P. (2007). Ethnic/racial disparities in hospital procedure volume for lung resection for lung cancer. *Medical Care, 45,* 655–663.

Neuhausen, K., Davis, A. C., Needleman, N., Brook, R. H., Zingmond, D., & Roby, D. H. (2014). Disproportionate-share hospital payment reductions may threaten the financial stability of safety-net hospitals. *Health Affairs, 33,* 988–996.

Rathore, S. S., Masoudi, F. A., Wang, Y. F., Curtis, J. P., Foody, J. M., Havranek, E. P., & Krumholz, H. M. (2006). Socioeconomic status, treatment, and outcomes among elderly patients hospitalized with heart failure: Findings from the National Heart Failure Project. *American Heart Journal, 152,* 371–378.