HOSPITAL OWNERSHIP AND QUALITY OF CARE: WHAT EXPLAINS THE DIFFERENT RESULTS IN THE LITERATURE?

KAREN EGGLESTON\textsuperscript{a,}\textsuperscript{*}, YU-CHU SHEN\textsuperscript{b,y}, JOSEPH LAU\textsuperscript{c,z}, CHRISTOPHER H. SCHMID\textsuperscript{c,\textdagger} and JIA CHAN\textsuperscript{d}

\textsuperscript{a}Walter H. Shorenstein Asia-Pacific Research Center, Stanford University, Stanford, CA, USA
\textsuperscript{b}National Bureau of Economic Research and Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA, USA
\textsuperscript{c}The Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center, Boston, MA, USA
\textsuperscript{d}FSI, CHP/PCOR Research Staff, Stanford University, Stanford, CA, USA

SUMMARY

This systematic review examines what factors explain the diversity of findings regarding hospital ownership and quality. We identified 31 observational studies written in English since 1990 that used multivariate analysis to examine quality of care at nonfederal general acute, short-stay US hospitals. We find that pooled estimates of ownership effects are sensitive to the subset of studies included and the extent of overlap among hospitals analyzed in the underlying studies. Ownership does appear to be systematically related to differences in quality among hospitals in several contexts. Whether studies find for-profit and government-controlled hospitals to have higher mortality rates or rates of adverse events than their nonprofit counterparts depends on data sources, time period, and region covered. Policymakers should be aware of the underlying reasons for conflicting evidence in this literature, and the strengths and weaknesses of meta-analytic synthesis. The ‘true’ effect of ownership appears to depend on institutional context, including differences across regions, markets, and over time. Copyright © 2008 John Wiley & Sons, Ltd.

Received 9 February 2007; Revised 1 November 2007; Accepted 2 November 2007

KEY WORDS: ownership; quality; hospitals; meta-analysis; nonprofit

INTRODUCTION

Does quality of care systematically differ among government (GOV)-owned, private not-for-profit, and for-profit (FP) hospitals? This question is of considerable policy importance in the US and many other countries. A large empirical literature unfortunately does not provide clear guidance for policy, since conflicting results abound. This mixed evidence about ownership and quality suggests that any attempt to synthesize the literature must first understand why results differ. Do the estimated effects of ownership depend on the time period studied and the characteristics of the regions and markets analyzed? Are some studies more credible than others, because they apply more convincing methods for isolating the impact of ownership from confounding factors? Unlike the clinical literature, systematic reviews and meta-analysis remain rare for questions of health service delivery,\textsuperscript{1} and none to date focuses on explaining the heterogeneity in study results for ownership and quality.

\textsuperscript{*}Correspondence to: Walter H. Shorenstein Asia-Pacific Research Center, Stanford University, 616 Serra, Encina Hall E301, Stanford, CA 94305-6055, USA. E-mail: karene@stanford.edu
\textsuperscript{y}Assistant Professor.
\textsuperscript{z}Professor of Medicine and of Clinical Research.
\textsuperscript{\textdagger}Director of the Biostatistics Research Center.
\textsuperscript{1}Examples of quantitative reviews from related areas include Aletras \textit{et al.} (1997), Sowden \textit{et al.} (1997), Halm \textit{et al.} (2002) and Devereaux \textit{et al.} (2004).

Copyright © 2008 John Wiley & Sons, Ltd.
The objective of this systematic review is to examine what factors explain the diversity of findings regarding hospital ownership and quality of care. We employ meta-regressions to quantify heterogeneity explained by study-level variables hypothesized to affect results – such as each study’s methodology to account for confounding factors, time period of the data or regions covered. Our review is part of a larger project applying meta-analytic techniques to the literature since 1990 on hospital ownership and performance; Shen et al. (2007) report the results on financial performance and ownership.

Our systematic review of hospital ownership and quality differs from and complements the only previous one we are aware of by Devereaux et al. (2002). First, we focus on explaining variation, rather than suggesting ‘one overall answer’ about ownership and quality. Second, we include studies analyzing patient outcomes other than mortality (e.g. adverse events), as well as many studies published since Devereaux et al. (2002) was completed. Third, we include GOV ownership in the review. Fourth, we do not exclude studies analyzing ownership conversions, but rather test the hypothesis that conversion studies differ systematically from others. Finally, we recognize that many studies analyze overlapping samples of hospitals; hence, conventional meta-analytic techniques that provide pooled estimates of ownership effect can give misleading results.

We find that whether FP- and GOV-controlled hospitals have higher mortality rates or rates of adverse events than their nonprofit counterparts depends on a study’s data source as well as time period and region covered. Pooled estimates of ownership effects from this literature are sensitive to the subset of studies included and the extent of overlap among hospitals analyzed in separate studies. In short, the extant literature on ownership and quality does not lend itself to a definitive quantitative synthesis giving one overall ‘answer’ about whether ownership matters. Rather, data and context matter. The paper proceeds as follows. The next section provides a brief overview of theory on ownership and quality. We then describe previous literature and why a meta-analysis complements narrative reviews. The third section presents the methodology of our systematic review and meta-analysis, including the literature search, data extraction, and analytic methods. The final two sections present the results and discuss policy implications.

CONCEPTUAL BACKGROUND ON OWNERSHIP AND QUALITY

This section provides a brief conceptual background on ownership and quality of care. More comprehensive overviews of the theoretical literature on hospital ownership and performance include Weisbrod (1988), Sloan (2000), Needleman (2001), Malani et al. (2003), and Shen et al. (2007). We draw on those frameworks to inform our review of the empirical literature.

Several prominent theories of ownership are relevant for healthcare delivery. These include nonprofits as quality–quantity maximizers (Newhouse, 1970); altruism or ‘profit deviators’ (e.g. Lakdawalla and Philipson, 1998); nonprofits as physician cooperatives (Pauly and Redisch, 1973) or as a signal of trustworthiness (Arrow, 1963; Hansmann, 1980; Glaeser and Shleifer, 2000); the property rights theory of ownership (Hart et al., 1997); soft budget constraints that differentially plague GOV firms (Kornai, 1986; Kornai et al., 2003); and GOV providers as benevolent embodiment of social objectives.

Economic theory does not give a clear-cut prediction about the relationship between ownership and quality of care. Any single empirical prediction about ownership and quality of care may be consistent with several different theories of ownership. For example, an empirical finding of higher quality in private nonprofits compared with FPs, which in turn have higher quality than GOV-owned hospitals, is consistent with almost every theory of ownership. Moreover, a single theory may be consistent with...
More than one empirical prediction about ownership and quality. One reason for this theoretical ambiguity is that most models include only two of the three ownership forms. Newhouse (1970) and many others focus on private not-for-profit versus (private) FP forms; Hart et al. (1997), Kornai (1986) and others focus on private FPs compared with GOV ownership. Few model the interaction of all three ownership forms, a mixture of ownership uncommon in other sectors of the economy.

Economic theory is ambiguous about whether quality would be closer to socially optimal with different ownership forms. For example, Lakdawalla and Philipson (1998) posit that not-for-profits effectively have lower marginal costs than FPs because of their preference for output or quality. Altruism and direct preference for quality both lead to predictions of higher quality among nonprofit firms. By contrast, the property rights theory of ownership (Hart et al., 1997) predicts that quality at private FPs may be higher or lower than that of GOV-owned providers, depending on the relative importance of quality innovations and quality-damaging cost reduction. Ownership differences also depend on the institutional and competitive context, including the mix of ownership forms in the market (although the literature to date has generally not focused on examining such spill-over effects).

Finally, healthcare quality encompasses multiple dimensions, not all of which are positively correlated. Frank et al. (2000) show how a profit-maximizing supplier would wish to distort the quality of different services – or manipulate the accessibility of services, through location, network, or capacity – to attract a profitable mix of patients. The fact that organizational forms differ systematically in the services that they provide, according to profitability (Horwitz, 2005a,b), further substantiates this prediction. Empirical studies differ in how they take account of differences in service offerings, and in what measures of quality they are able to include.

Clearly then, economic theory provides a valuable framework for thinking about when and how ownership might matter. But theoretical predictions are mixed, and the impact of ownership on quality remains an empirical question.

**OVERVIEW OF PREVIOUS LITERATURE: WHAT CAN META-ANALYSIS CONTRIBUTE?**

Qualitative reviews on ownership of health service providers, such as Sloan (2000) or Needleman (2001), provide valuable insights, with previous results placed in context through expert interpretation. Yet the qualitative approach has limitations as a general method for synthesizing evidence. For example, there is no straightforward way for a less-informed reader to detect selective emphasis on studies that most conform to the author’s prior beliefs or to the target audience’s expectations.

Some authors seek to provide a quantitative foundation by cataloguing the results of published papers in the literature and discussing what number (or percentage) of studies find certain results. Recent contributions such as Rosenau and Linder (2003) and Schlesinger and Gray (2006) exemplify this approach. Importantly, both provide evidence across a wide spectrum of US healthcare suppliers, including nursing homes and health plans. However, such a ‘vote counting’ approach also has multiple limitations (Hedges and Olkin, 1985). Moreover, many studies in this literature include the same set of hospitals – and potentially the same patient outcomes – in overlapping years. Giving equal weight to each study is analogous to physicians counting a 50-patient study of a new drug or procedure as one vote and a 10,000-patient study as a second vote – even if the 50 patients were a subset of the 10,000.

---

4 Policy debate and the empirical literature tend to equate increases in quality with improvements in social welfare. However, conceptually, resource allocation in health care is like any other area of society: in the face of competing social needs, investment in healthcare quality could be excessive compared with what would maximize social welfare (see discussion in Gaynor, 2006). Examples of excessive quality might include ‘gold-plating,’ providing amenities (such as hotel services) in excess of their value, or investing in technologies that improve quality/safety by a small amount compared with their cost and compared with alternative, more cost-effective ways of improving population health and patient outcomes.

Systematic reviews and meta-analyses are important complementary methodologies for synthesis. A series of tools have been developed to address differences in statistical power, heterogeneous study designs, and nonindependent observations (Rosenthal, 1991; Sutton et al., 2000). Policymakers designing healthcare delivery policy deserve a rigorous evidence base to inform their decisions, and these methods hold considerable promise. However, policymakers should also be aware of the shortcomings of meta-analytic synthesis, which is not a ‘magic bullet.’ For example, pooling results across studies yields more powerful information, yet it can also obscure differences among subgroups (such as secular and religious nonprofits), and frequently forces the effects of differing study features on the outcome variable to be homogenous across studies, despite underlying heterogeneity. Moreover, researchers are limited by the institutional and contextual detail supplied in each published study, and by the total number of studies available. Well-done qualitative and quantitative syntheses are best-considered complements rather than substitutes.

The only previous systematic review of hospital ownership and quality of which we are aware (Devereaux et al., 2002) identified 15 studies published between 1988 and 2001 comparing mortality rates among private not-for-profit and FP hospitals. Our study complements and contrasts with theirs. Of the 31 studies we identified, 10 were also included in their review. Devereaux et al. (2002) aimed to provide a single definitive answer to the question of whether quality differs in FP and private not-for-profit hospitals. Yet the nature of observational studies in this area severely limits the ability of meta-analysis to achieve that goal. Although Devereaux and colleagues describe their review as including ‘fifteen observational studies, involving more than 26 000 [US] hospitals,’ there are actually fewer than 5000 general acute inpatient hospitals in the entire US. The authors do not acknowledge that the studies are not independent samples, although they do note that almost all studies draw from the same database. Another reason for interpreting the single answer from Devereaux et al. (2002) with caution is that they use patient sample size to define confidence intervals, whereas ownership is a hospital-level characteristic. Since there are many fewer hospitals than patients, using patient sample size yields a misleadingly precise estimate of hospital-level mortality differences and tends to give the largest weight to studies with the fewest controls for confounding factors (because authors often restrict sample size specifically to remove residual confounding from case mix or market characteristics).

METHODS

Literature search, selection, and extraction

Data sources. Our literature search employed the keywords ‘hospitals,’ ‘ownership,’ ‘for-profit,’ ‘not-for-profit,’ ‘nonprofit,’ and their combinations. Databases included Medline, EconLit (Economics Literature), and Proquest/ABI (for dissertations), as well as hand search and work cited in previous qualitative reviews. We included all published and unpublished articles or book chapters written in English between January 1990 and July 2004, initially resulting in 1357 potentially relevant studies.

---

6Our review includes 16 study results on mortality, and six on other patient outcomes, published in the period covered in Devereaux et al.’s review. Five studies that were included in Devereaux et al. (2002) did not fit our inclusion criteria either because they were published before 1990 or did not employ robust multivariate regression to control for confounding effects. Admittedly, any cut-off date for publication is somewhat arbitrary; we decided on 1990 because the competitive context and payment methods of hospitals in the US in the early 1980s were so different from today that those studies would be of questionable comparability. This assumption is validated to some extent by our empirical finding that the time period of data impacts study results. For a discussion of earlier studies, see Sloan (2000).

7To illustrate, consider Kuhn et al. (1994) and McClellan and Staiger (2000). In Devereaux et al.’s meta-analysis (see Figure 2, p. 1403), Kuhn et al. (1994) receives over eight times the weight (12.34 versus 1.48) as McClellan and Staiger (2000) in estimating the pooled effect of ownership, yet McClellan and Staiger’s study includes more hospitals than that of Kuhn. Using the hospital sample size instead, Kuhn’s estimate in fact is not statistically different from zero, implying no ownership difference (see Figure 1), whereas in Devereaux et al. (2002) Kuhn’s estimate is one of the largest and most statistically significant suggesting nonprofits have higher quality.
We identified 77 additional studies by asking corresponding authors for any new or unpublished work comparing hospitals of different ownership forms.

**Study selection.** We selected studies of general, acute, short-stay hospitals in the US that used multivariate analysis to study hospital performance. The hospital sample had to include hospitals from at least two of our three included ownership forms (GOV owned, private not for profit and private FP). We excluded studies that only compared sub-categories of ownership (such as religious versus secular not for profits) or focused on hospitals of the Department of Veterans Affairs. From the 340 studies fitting these selection criteria, we further narrowed down the field to studies examining one or more of the following outcomes: financial performance; patient outcomes (mortality, complications, and other adverse outcomes); uncompensated care or other measures of community benefits\(^8\); or staffing. Two authors (Eggleston and Shen) independently applied these selection criteria, achieving 7% discrepancy in coding, with consensus resolution of all differences. From the 169 citations that fell within this outcome scope, we excluded duplicates (such as dissertations and subsequently published papers), studies with very small sample sizes (fewer than 50 hospitals), and studies that did not report needed data (e.g. ownership coefficients in a regression) and for which authors were unable to supply that data.

Through this search and selection process, we identified 141 studies. Readers interested in a graphical presentation of our selection process can refer to Figure 1 in our companion paper (Shen *et al.*, 2007). In this paper, we report results from analysis of the 31 studies of patient outcomes.

**Data extraction.** We extracted data needed to construct measures of treatment effect and to test our hypotheses about underlying sources of heterogeneity in studies of hospital ownership. The data fields included each study’s data sources; sampling frame (covered years, covered regions, number of hospitals, and patients); detailed outcome definition(s); and the ownership coefficients and their associated statistics (such as standard errors, degrees of freedom of the empirical model, etc.).\(^9\)

We coded each study’s analytic methods using binary variables for whether the study included the following information: patient demographic characteristics; co-morbidities; hospital-level characteristics (such as bed size, share of Medicare, and/or Medicaid admissions); market-level characteristics (such as per capita income, population size, concentration, or competitiveness measures); and cross-sectional or panel data estimation methods. We also coded whether ownership variables were only included as control variables and not the main research question of the study, and whether the analyses explicitly address ownership conversion.

Since extracting and analyzing all sensitivity analyses for each study would neither be practical nor yield additional independent information, we attempted to identify the author(s)’ ‘preferred’ specification for each study. However, we did extract data for multiple estimates when a single study reported estimates that appeared to be roughly equal in their importance to the author’s conclusions, such as separate estimates for different states or different years.\(^10\)

---

\(^8\)We included studies that measure provision of services to specific groups – such as the uninsured or Medicaid beneficiaries – as part of a measure of community benefits. We did not include studies examining whether GOV, FP, and nonprofit hospitals differ in service offerings. See Horwitz (2005a,b) for excellent discussion of this closely related topic.

\(^9\)Not all studies report the necessary statistics. For example, many policy journals report only the significance level of the coefficient. We contacted the corresponding authors to get the exact *t*-statistics where possible. In cases where we could not get the exact number, we estimate the *t*-statistic by generating a random number within the reported significance range.

\(^10\)Undoubtedly even this method leaves some room for bias associated with our identifying which estimate represented the ‘preferred’ one from each study, especially when the abstract and conclusion did not rely heavily on a specific estimate in summarizing results. Unfortunately, avoiding this problem was infeasible within the constraints of our grant, as it would have meant coding from 5 to 20 or more variations for each study and then trying to weight the results in an aggregate estimate, which would re-introduce the problem of subjectively deciding weights. Note that putting equal weight on each estimate is only one example of such a weighting scheme, and lacks face validity given authors’ often openly stated preference for certain specifications.
estimate meta-regressions with each year’s estimate appearing separately, and compare this with a meta-regression with only a single, average result from each study. The average estimate is constructed from the point estimates and sample sizes of each independent estimate in a given publication. The figures graph the single, average estimate for each study.

**Statistical methodology**

Analytical models among studies vary greatly in this literature; some use linear models and some use nonlinear models, for example. Our analysis requires defining a standardized statistical measure, an ‘effect size’, for each study that renders diverse studies comparable. We use this effect size as the dependent variable in our meta-regressions. Constructing the effect size measure for each study is critical in the success of this meta-analysis, since the 31 studies in our sample analyze diverse dependent variables $Y$, including mortality rates and rates of other adverse events, such as surgical complications or medical errors. Even within a narrowly defined category such as mortality, some dependent variables are measured as continuous variables at the hospital level; others are dichotomous outcomes at the patient level.\(^{11}\)

To have a standardized measure of ownership effect based on statistics that are commonly reported in published studies, we use the partial correlation coefficient, $r^{\text{own}}$. The partial correlation coefficient $r^{\text{own}}$ describes the relationship between the dependent variable $Y$ and the ownership variable, holding other variables constant. It measures the correlation between a given ownership form and quality, controlling for the effect of covariates.\(^{12}\) The effect size $r^{\text{own}}$ has the advantage of being unit free and independent of a study’s sample size. Thus, an important advantage of using test statistics to define the partial correlation coefficient is that, as Rosenthal (1991, p. 19) points out, ‘it makes no difference whether the data are in dichotomous or continuous form, or whether they are ranked’ when all we are doing is capturing the product moment correlation coefficients (see also Hedges and Olkin, 1985, p. 101).

Each ownership comparison is represented by a different partial correlation. A study that investigates the relative performance of FP and nonprofit hospitals, for example, would report statistics that can be used to define $r^{\text{FP}}$ for that study. The result of a study comparing GOV and nonprofit hospitals would be captured by $r^{\text{GOV}}$. If the $i$th study includes all three ownership forms, its results are coded in our analysis as two separate effect sizes comparing FPs and GOV hospitals to private not-for-profits, $r^{\text{FP}}_i$ and $r^{\text{GOV}}_i$, respectively. These partial correlation coefficients are defined as follows:

$$r^{\text{own}} = \sqrt{\frac{(r^{\text{own}})^2}{(r^{\text{own}})^2 + \text{degrees of freedom}}} \quad \text{own} \in \{\text{FP, GOV}\}$$  \hspace{1cm} (1)

where the $t$-statistic for each ownership category and degrees of freedom in (1) come from the following regression in a given underlying study:

$$Y = \alpha + \beta_1 \text{FP} + \beta_2 \text{GOV} + \gamma X + \xi$$  \hspace{1cm} (2)

where $Y$ is a measure of hospital quality (in the case of nonlinear models with dichotomous variables – representing almost half our sample – the left-hand side variable could be replaced with $\text{prob}(Y = 1)$); FP indicates for-profit ownership, GOV signifies GOV ownership (the omitted group is not-for-profit ownership); $X$ is a vector of hospital and market characteristics; and $\xi$ is the error term. If estimated

\(^{11}\)For example, only two of the studies used plain ordinary least squares; other methods included logistic regressions, ANOVA with covariates, hospital fixed effects (14 out of the 31 studies used panel data methods), GLS matched with propensity scores, two- and three-stage estimations, Cox-proportional hazard model, and Bayesian estimation with a selection model (assuming mortality and choice of hospital are endogenous).

\(^{12}\)Of the 31 studies, 71% include controls for patient case mix; 84% include patient demographics; 100% include hospital-level control variables; and 71% include market-level control variables. For 19 (61%) of the studies, ownership variables are included only as control variables, since the primary research question is not ownership.
properly, the coefficients $\beta_1$ and $\beta_2$ capture the average effect on $Y$ of FP and GOV ownership, respectively, relative to not-for-profit ownership. To define the effect sizes for a given study, we use the degrees of freedom of the empirical model (2) used in that study and the $t$-statistic associated with its regression coefficient to define $r_{\text{own}}$ according to (1). For example, the $t$-statistic associated with $\beta_2$, $r_{\text{GOV}}$, defines the effect size

$$r_{\text{GOV}} = \sqrt{\frac{(r_{\text{GOV}})^2}{(r_{\text{GOV}})^2 + \text{degrees of freedom}}}$$

The definition of effect size $r_{\text{own}}$ in (1) will always give rise to a positive number. We then attribute to $r_{\text{own}}$ the sign of the original study statistic, so that a positive effect size indicates a higher rate of adverse events (lower quality) relative to nonprofit hospitals.

Since the distribution of $r_{\text{own}}$ becomes more skewed as the population value of $r_{\text{own}}$ gets further away from zero, we standardize using the Fisher transformation and compute the confidence interval around the standardized effect size (see Hedges and Olkin, 1985; Greene, 1997; Shen et al., 2007). Specifically, the dependent variable in our meta-regressions is defined by the following formula:

$$z_{ri}^{\text{own}} = \frac{1}{2} \frac{1 + r_{ri}^{\text{own}}}{1 - r_{ri}^{\text{own}}} \quad \text{own} \in \{\text{FP, GOV}\}$$

Our conclusions are not sensitive to whether we standardize the effect size or not (i.e. the results are qualitatively similar using $r_{ri}^{\text{own}}$ or $z_{ri}^{\text{own}}$ as the dependent variable in our meta-regressions).

Explanation variation in study results using meta-regression analysis. We employed random-effects meta-regression analysis to quantify to what extent various study characteristics account for heterogeneity of findings.\(^1\) Below we provide a brief summary of the methods and their application in our context.

Meta-analyses may assume either that the treatment effects – in this case, the effects of ownership – are fixed or random. A fixed-effect model assumes that there is one single true effect, whereas a random-effect model assumes that there is a distribution of true effects. To illustrate, let $\theta_{ri}^{\text{own}}$ be the true effect size of ownership (either FP or GOV ownership, relative to private not-for-profit ownership) for the $i$th study among a total of $k$ studies. Let $z_{ri}^{\text{own}}$ be the estimated effect size of ownership reported in the $i$th study (we use $z_{ri}^{\text{own}}$ because it is more likely to follow the assumed normal distribution). Then we can relate the estimated and true effect sizes in the following way:

$$z_{ri}^{\text{own}} = \theta_{ri}^{\text{own}} + e_i \quad \text{own} \in \{\text{FP, GOV}\}$$

where $e_i$ is assumed to be independently and normally distributed with a mean zero and variance $v_i$ due to sampling error in the $i$th study. We assume that this variance known and given by the sample variance of the observed effect size in the $i$th study. We can express the variance of $z_{ri}^{\text{own}}$ as $\text{var}(r_{ri}) = \tau_0^2 + v_i$, where $\tau_0^2$ is the between-study, or random-effects variance for studies comparing those ownership categories. This between-study variance $\tau_0^2$ is unknown and must be estimated from the data.

A random-effects meta-analysis model like that of (4) can be estimated using maximum likelihood approximation (Sutton et al., 2000, p. 74).

Now suppose the true effect of ownership also depends on $p$ known study features, $X_1, \ldots, X_p$, such as data source or region covered. We can explain the variation in the true effect size $\theta_{ri}^{\text{own}}$ with a linear model:

$$\theta_{ri}^{\text{own}} = \delta_0^{\text{own}} + \delta_1^{\text{own}} X_{i1} + \delta_2^{\text{own}} X_{i2} + \cdots + \delta_p^{\text{own}} X_{ip} + u_i \quad \text{own} \in \{\text{FP, GOV}\}$$

\(^{13}\) Stanley (2001) and Sutton et al. (2000) provide an accessible introduction to this approach. The concept of random versus fixed-effects regression models in meta-analysis is quite different from how these terms are applied in longitudinal analysis (where the data would contain multiple observations from one study unit).
where the random effects \( u_i \) are assumed to be independent, with a mean of zero and variance \( \tau^2_0 \); and the \( \delta \)'s are regression coefficients estimating how study features \( X \) impact a study’s estimated ownership effect \( \theta^\text{own}_i \). The random effect \( u_i \) is the deviation of the study’s true effect size from that predicted by the model \( \theta^\text{own}_i + \delta_1^\text{own} X_{i1} + \delta_2^\text{own} X_{i2} + \cdots + \delta_p^\text{own} X_{ip} \). For example, the true impact of GOV ownership on a hospital’s performance may vary across institutional contexts (such as time period, region, hardness of budget constraint, etc.). Even after accounting for observable differences \( X_{1i}, \ldots, X_{pi} \) reported in each study, there remains unexplained variation in the true effect of GOV ownership. This unexplained variation – attributable to unobserved factors such as managerial initiative or softness of the budget constraint (factors rarely included in hospital ownership studies but known to impact performance) – is described by a distribution of true effects with variance \( \tau^2_0 \).

Substituting the model of the true effect size (5) into the observation equation (4), we obtain

\[
z^\text{own}_{i} = \delta_0^\text{own} + \delta_1^\text{own} X_{i1} + \delta_2^\text{own} X_{i2} + \cdots + \delta_p^\text{own} X_{ip} + u_i + e_i \quad \forall \, i \in \{\text{FP, GOV}\}
\]

The variance of \( z^\text{own}_{i} \), controlling for the \( X \)'s, is var(\( r_i \)) = \( \tau^2_0 + v_i \). Thus, the estimated effect of ownership in the \( i \)th study varies both because of within-study sampling error, \( e_i \), and because of between-study variation in the true effect size that is not explained by study features, \( u_i \). These give rise to two components of variance, the known within-study variance \( v_i \) and the unknown between-study variance \( \tau^2_0 \). Both are assumed to have a normal distribution.

Model (6) is the most appropriate method for analyses like ours that attempt to explain variation among studies with widely differing research questions (e.g. the effects of integration or competition as well as those of ownership) and policy contexts (such as the mid-1980s versus late 1990s). Because of the relatively limited number of studies compared with the large number of factors that shape hospital performance, only a limited array of study features \( X \) can be included in our meta-regressions. Using a random-effects model – that is, estimating the full equation (6) – acknowledges that there remains unexplained variation across studies not captured by our coded set of study features, both because not all reported features could be coded and because not all factors that shape hospital performance are explicitly reported in the underlying studies.

As Sutton et al. (2000, p. 98) note, estimation of the random-effects meta-regression model in (6) generally requires iterative or simultaneous estimation of the regression coefficients (the \( \delta \)'s) and the random-effects variance \( \tau^2_0 \), because they are mutually dependent in the sense that estimation of one requires an estimation of the other. We use the command ‘metareg’ in Stata 9.0 to carry out the estimation. The metareg command uses a restricted maximum likelihood algorithm to estimate the between-study component of variance. Since a random-effects model allows for random error across studies as well as sampling error within each study, the confidence intervals constructed under random-effects meta-regression models are always bigger than those under fixed-effects models. Estimation also assumes that the residual or unexplained variation is random and not systematically correlated with the coded study features \( X \). Although we hypothesize that this assumption holds for our analysis, interpretation of the meta-regression results should nevertheless be cautious given the inability to test this assumption.

In our meta-regression analysis, the dependent variable is the standardized effect size from each study as defined in (3). Empirical features of each study serve as explanatory variables \( X \). The regression model, following (6), is necessarily parsimonious, given the limited number of studies and the collinearity of many study features (such as data source and region covered; correlation matrices of study features are included in the appendix of Eggleston et al., 2006).

We hypothesized that several factors (\( X \)) might account for a substantial fraction of between-study variation in findings. Results might be driven by the underlying data sources: studies using Medicare claims (serving elderly patients under administered prices) might produce different results from those using all-payer data (including nonelderly patients and market-based prices). Variation in results might also be explained by the covered periods of the study or the regions studied. Another factor that could
plausibly explain different results is whether a study explicitly accounts for hospital market competition (for example, by implementing a selection model) or employs panel data estimation methods to minimize the unobserved heterogeneity across markets.

Our effect size measures, defined by partial correlation coefficients, should remain valid when observations are correlated (Rosenthal, 1991). However, the studies do not represent independent samples, and statistical power will not increase proportionally with each additional study’s sample size. We therefore do not emphasize the pooled estimates of ownership effects.

Publication bias. We looked for evidence of publication bias by examining the relationship between the absolute value of $t$-statistics and the square root of degrees of freedom, following Card and Krueger (1995). Similar to studies of ownership and other aspects of performance, a plot of sample size and $t$-statistics does not reveal a pattern of $t$-statistics hovering around 2 regardless of sample size, as would be the case if studies are only published when they find statistically significant effects (not shown). Thus, we find no evidence of publication bias. This finding is consistent with the theoretical ambiguity in this field as well as the empirical results of Devereaux et al. (2002), who report that they did not find evidence of publication bias among the 15 studies on mortality that they reviewed.

RESULTS

Summary of studies' ownership estimates

Table I summarizes our 31 included studies. Outcomes differ: 25 analyze mortality rates, and 13 analyze other adverse patient outcomes, such as surgical complications or medical errors. Most studies are representative of the entire US. Hospital sample sizes are often quite large, especially for studies using panel data, suggesting that many studies analyze the same hospitals in similar time periods. Many hospital-level analyses do not report patient sample sizes. For readers interested to know the analytical models of each study, an extended Table I with a methodology summary is available from the authors upon request.

The effect sizes and their confidence intervals from random-effects meta-analysis models (as described by Equation (4)) appear in the figures. Figure 1 shows effect sizes for studies comparing FPs to private not-for-profits; Figure 2 shows effect sizes for studies comparing GOV hospitals to private not-for-profits. A positive effect size indicates that FPs (or GOV hospitals) are associated with a higher rate of adverse events compared with private not-for-profit hospitals.

The majority of studies find no statistically significant difference between not-for-profit and FP hospitals in mortality or other adverse events. Figure 2 illustrates that studies using nationally representative data find GOV hospitals to have either higher or similar rates of mortality and other adverse events as private not-for-profits do.

In all cases the chi-square test for heterogeneity clearly rejects the null hypothesis that our combined studies are a homogeneous set (data not shown), thus validating our approach of using random-effects regression models in our meta-regressions. Below we present our meta-regression results; we revisit the issue of overlap in the sensitivity analysis at the end.

Explaining variation in studies examining patient outcomes

Table II presents our meta-regression results. For studies that report multiple estimates of the ownership effect, such as separately reporting estimates for different years, the top panel includes all

---

14See Shen et al. (2007) for a fuller description of this method for identifying publication bias.
### Table I. Summary of included studies

<table>
<thead>
<tr>
<th>Article ID</th>
<th>Mortality</th>
<th>Other outcomes</th>
<th>Covered region</th>
<th>Covered years</th>
<th>Number of hospitals</th>
<th>Number of patients</th>
<th>Ownership forms studied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Haider and Wan (1991)</td>
<td>X</td>
<td>US</td>
<td>1984</td>
<td>243</td>
<td>N/A</td>
<td>N, F</td>
<td></td>
</tr>
<tr>
<td>Bond et al. (1999)</td>
<td>X</td>
<td>US</td>
<td>1992</td>
<td>3763</td>
<td>N/A</td>
<td>N, F, G</td>
<td></td>
</tr>
<tr>
<td>Brennan et al. (1991)</td>
<td>X</td>
<td>NY</td>
<td>1984</td>
<td>51</td>
<td>31 429</td>
<td>N, F, G</td>
<td></td>
</tr>
<tr>
<td>Brook et al. (1990)</td>
<td>X</td>
<td>13 Geographic areas</td>
<td>1981</td>
<td>1171</td>
<td>N, F, G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ettner and Hermann (2001)</td>
<td>X</td>
<td>US</td>
<td>1990</td>
<td>2349</td>
<td>N/A</td>
<td>N, F</td>
<td></td>
</tr>
<tr>
<td>Kuhn et al. (1994)</td>
<td>X</td>
<td>US</td>
<td>1988</td>
<td>3782</td>
<td>N/A</td>
<td>N, F, G</td>
<td></td>
</tr>
<tr>
<td>Lee and Wan (2002)</td>
<td>X</td>
<td>X</td>
<td>US</td>
<td>1997</td>
<td>358</td>
<td>N/A</td>
<td>N, F, G</td>
</tr>
<tr>
<td>Manheim et al. (1992)</td>
<td>X</td>
<td>US</td>
<td>1987</td>
<td>3796</td>
<td>N/A</td>
<td>N, F, G</td>
<td></td>
</tr>
<tr>
<td>Mukamel et al. (2001)</td>
<td>X</td>
<td>134 MSAs</td>
<td>1990</td>
<td>1927</td>
<td>N/A</td>
<td>N, F, G</td>
<td></td>
</tr>
<tr>
<td>Pitterle et al. (1994)</td>
<td>X</td>
<td>US</td>
<td>1988</td>
<td>4864</td>
<td>N/A</td>
<td>N, F, G</td>
<td></td>
</tr>
<tr>
<td>Wan (1992)</td>
<td>X</td>
<td>VA</td>
<td>1987</td>
<td>85</td>
<td>N/A</td>
<td>N, F</td>
<td></td>
</tr>
<tr>
<td>Yuan et al. (2000)</td>
<td>X</td>
<td>US</td>
<td>1984–1993</td>
<td>5127</td>
<td>N/A</td>
<td>N, F, G</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>25</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Patient outcomes other than mortality include rates of complications or rates of adverse events. MSA, metropolitan statistical areas; N/A, not available.*

*Studies of ownership conversion.*
HOSPITAL OWNERSHIP AND QUALITY OF CARE

Figure 1. Summary of effect sizes for patient outcomes: studies comparing not-for-profit and for-profit hospitals

NOTE: Weights are from random effects analysis

FP has lower rate of adverse events  Effect size  FP has higher rates of adverse events

estimates separately and the bottom panel includes only a single, averaged estimate for each study. The coefficients are quite similar across both specifications, although of course the standard errors are all bigger in the regression with a single observation per study.

Figure 2. Summary of effect sizes for patient outcomes: studies comparing not-for-profit and government hospitals

NOTE: Weights are from random effects analysis.
Table II. Meta regression results examining variation in studies of patient outcomes (mortality and adverse events)

Using multiple estimates per study

<table>
<thead>
<tr>
<th>Study features</th>
<th>Share of studies with given study feature (%)</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>Share of studies with given study feature (%)</th>
<th>Coefficient</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicare claims</td>
<td>60</td>
<td>-0.005</td>
<td>0.011</td>
<td>72</td>
<td>0.046**</td>
<td>0.021</td>
</tr>
<tr>
<td>National Long-term Care Survey</td>
<td>14</td>
<td>-0.023**</td>
<td>0.005</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data from 1980s only</td>
<td>22</td>
<td>0.020*</td>
<td>0.010</td>
<td>24</td>
<td>-0.028*</td>
<td>0.016</td>
</tr>
<tr>
<td>Explicit modeling for market competition</td>
<td>76</td>
<td>-0.003</td>
<td>0.006</td>
<td>78</td>
<td>-0.009</td>
<td>0.010</td>
</tr>
<tr>
<td>Representative of the US</td>
<td>67</td>
<td>0.017*</td>
<td>0.009</td>
<td>61</td>
<td>0.019</td>
<td>0.014</td>
</tr>
<tr>
<td>Constant (reference group)</td>
<td></td>
<td>-0.009</td>
<td>0.011</td>
<td></td>
<td></td>
<td>0.004</td>
</tr>
<tr>
<td>Proportion of variation due to heterogeneity (I-square)</td>
<td></td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td>0.26</td>
</tr>
<tr>
<td>Number of observations</td>
<td></td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
</tbody>
</table>

Using a single, averaged estimate per study

<table>
<thead>
<tr>
<th>Study features</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>Share of studies with given study feature (%)</th>
<th>Coefficient</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicare claims</td>
<td>-0.007</td>
<td>0.011</td>
<td>0.051*</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td>National Long-term Care Survey</td>
<td>-0.023**</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data from 1980s only</td>
<td>0.022**</td>
<td>0.010</td>
<td>-0.030</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>Explicit modeling for market competition</td>
<td>-0.003</td>
<td>0.007</td>
<td>0.014</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>Representative of the US</td>
<td>0.017*</td>
<td>0.009</td>
<td>0.017</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>Constant (reference group)</td>
<td>-0.009</td>
<td>0.011</td>
<td>0.002</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td>Proportion of variation due to heterogeneity (I-square)</td>
<td>0.02</td>
<td></td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>42</td>
<td></td>
<td>18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.10; **p < 0.05. Note: A positive coefficient on the constant term indicates that for-profit (or government) hospitals are associated with higher mortality or a higher rate of adverse events than private not-for-profit hospitals for the reference group. A positive coefficient on a study feature indicates that for-profit (or government) hospitals are associated with higher mortality or a higher rate of adverse events than private not-for-profit hospitals for studies with a given study feature compared with the reference group. The $I^2$ for the null model with no covariates is 0.244 for studies comparing nonprofits and for-profits, and 0.405 for studies comparing nonprofits and government hospitals.
We test our hypotheses that data sources, regions and years covered, and methodology might explain a significant proportion of study differences. Studies in the reference group have the following features: (1) they do not use two popular data sources – Medicare claims or National Long-term Care Survey data (such as studies using state administrative databases); (2) they include data from 1990 or later; (3) they do not include explicit modeling of market competition; and (4) they use hospital samples that are not US representative (such as single-state analysis). The estimated constant term, the ownership effect for this reference group, is vanishingly small and not statistically different from zero for all four meta-regressions in Table II. This result indicates that studies with these features do not find any difference in measured patient outcomes among hospitals of different ownership forms.

To see how specific study features modify this ‘overall prediction,’ consider first the meta-regression results for studies comparing FPs and nonprofits. The two-thirds of studies that use hospital data representative of the US are associated with a finding of higher mortality rates or adverse event rates in FP hospitals relative to the reference group, consistent in sign and relative magnitude with the findings of the previous Devereaux et al. (2002) meta-analysis. Ownership effects based on 1980s data suggest a wider quality gap between private hospitals. In contrast, the few (14% of) studies using the National Long-term Care Survey find a narrower quality difference between FPs and nonprofits. The negative coefficient associated with this data source in our meta-regression – which would indicate that FPs have lower mortality rates than nonprofits – stems from the fact that in the original articles the coefficients on FP ownership are mostly negative although statistically insignificant. Since all National Long-term Care Survey studies are also representative of the US (in the sense that the survey includes over 35 000 Medicare beneficiaries from all over the US), the negative coefficient should be interpreted as offsetting the positive coefficient on the latter study feature (−0.023 + 0.017 = −0.006). None of the studies based on the National Long-term Care Survey (all by the same group of authors) conclude that FPs and nonprofits statistically differ in measured outcomes. This result is captured famously by the title of one of these studies: ‘Hospital ownership and cost and quality of care: Is there a dime’s worth of difference?’ (Sloan et al., 2001).

Interestingly, whether or not a study explicitly models market competition does not appear to be a significant predictor of different results. This may stem in part from the fact that the overwhelming majority of studies in this literature (76%) do use relatively sophisticated methods to account for competition, and the other studies usually include at least a few control variables at the market level. Differences in analytic methods, including measures of study quality based on control for competition, case mix and other confounding factors, provide less traction in explaining divergent results for this literature than among studies examining hospital financial outcomes, where many studies use less sophisticated methods (Shen et al., 2007).

For private not-for-profit and GOV hospitals, the 72% of studies using Medicare claims are associated with findings of significantly higher mortality and adverse event rates in GOV-owned hospitals. No other study features consistently explain variation in study results. Although the negative coefficient on 1980s data suggests that studies using earlier data find a smaller quality gap among nonprofit and GOV hospitals, this study feature is not statistically significant when each study contributes only a single observation to the meta-regression. (The correlation coefficient between having data from the 1980s and using Medicare claims is −0.02 for this group of studies.)

Overall, these study features explain a substantial portion of variation in findings across studies of patient outcomes. The $I^2$-squared is a statistical measure of the proportion of variation due to heterogeneity between studies. The $I^2$-squared for the null model with no covariates is 0.244 for studies.

15Since we have too few studies and too much correlation among some explanatory factors to include them all in one meta-regression, we used partial regressions to identify factors that appear to play a primary rather than a secondary or no significant role in explaining inter-study differences. See Eggleston et al. (2006) for the series of ‘univariate’ or partial regressions that helped to guide the specification of our overall meta-regression, as well as correlation tables for study features.
comparing nonprofits and FPs, and 0.405 for studies comparing nonprofits and GOV hospitals. With the study features as covariates, the $I^2$-squared declines to 0.018 for private hospitals and 0.328 for the meta-regression of studies comparing GOV and private nonprofit hospitals, indicating a fuller explanation of heterogeneity among studies, especially for private hospitals. However, not all discrepancies among studies of ownership can be explained with this analysis, suggesting again caution in drawing overall conclusions about ownership and quality.

Further exploration of overlapping samples

Probably the best way to proceed in light of overlap among studies is to identify a systematic way of categorizing the degree of overlap between individual studies, and then estimate meta-analyses that take account of the complex correlation matrix between study estimates of ownership effects. Although developing such a methodology falls outside the scope of the present study, it could prove useful to develop for future work in quantitative synthesis for health service delivery topics, where the units of analysis (hospitals and other organizations or systems) are less numerous than individual patients and therefore may appear in many studies.

To illustrate the sensitivity of pooled ownership effect estimates to this overlapping issue, we instead developed a simple categorization of overlap based on each study’s dependent variable, ownership forms compared, regions covered and years of data. Of course, there may be a strong correlation in quality measures for any given hospital over time, so that it is an exaggeration to claim that studies with non-overlapping years represent independent results. But categorizing overlap at least along one dimension does give a more concrete way to understand the difficulty with pooling study results.

Consider the following pooled estimates of the difference between private nonprofit and FP hospital quality, pooling across studies of mortality and other adverse outcomes, based on random-effects meta-analysis. For the 10 studies included both in our review and the previous meta-analysis (Devereaux et al., 2002), we find a pooled effect size of about 0.01, not statistically different from zero at conventional levels (0.01, 95% confidence interval −0.003 to 0.024). Updating this estimate by including all studies published since the Devereaux et al. review was published (a total of 36 ownership estimates), we again find an estimate of about 0.01, this time statistically significant (0.011, 95% confidence interval 0.003–0.019). This positive pooled effect indicates that FPs have higher mortality rates or rates of adverse events than nonprofits do.

However, many of these studies significantly overlap in years studied and thus may be using data on the same patients from the same hospitals. The statistical precision reported for the above pooled estimates is therefore misleading. Using a sample of studies with non-overlapping years of data, each of which examines a sample representative of the US as a whole, we find a pooled estimate of 0.017 ($p > 0.024$) for one group but an effect size of zero ($−0.0084, p > 0.443$) for a similar group of studies with non-overlapping years. Studies from single states as well as from multiple states also yield pooled effect sizes statistically equivalent to zero, suggesting no difference between FPs and nonprofits in quality of care.

The problem is relatively clear: a combination of studies with enough statistical power to identify a non-zero ownership effect is also almost inevitably a combination of studies that analyze overlapping samples of hospitals. If synthesis is confined to a group of studies that do not overlap, then that group usually consists of too few studies to have power to find any significant effect (given that the ownership differences appear to be ‘small’). Thus, we should be cautious in attempting to provide ‘one answer’ in

\[ \text{HOSPITAL OWNERSHIP AND QUALITY OF CARE} \]  

Copyright © 2008 John Wiley & Sons, Ltd.  

\[ \text{Health Econ. 17: 1345–1362 (2008)} \]  

DOI: 10.1002/hec

\[ 16 \text{Al-Haider and Wan (1991), Bond et al. (1999), Ettner and Hermann (2001), Kuhn et al. (1994), Lanska and Kryscio (1998), Lee and Wan (2002), Manheim et al. (1992), and McClellan and Staiger (2000).} \]  

\[ 17 \text{Slonim et al. (2003), Manheim et al. (1992), Lanska and Kryscio (1998), Bond et al. (1999), and Al-Haider and Wan (1991).} \]  

\[ 18 \text{Six estimates from Brennan et al. (1991), Deily and McKay (2006), Unruh (2000), Wan (1992), and Farsi and Wan (2004).} \]  

\[ 19 \text{Four estimates from Sloan (2002), Keeler et al. (1992), and Brook et al. (1990).} \]
synthesizing this literature; conventional meta-analysis of all previous studies does not yield a statistically reliable estimate of the overall impact of ownership on measures of quality. Our meta-regression approach focuses on an important complementary task: that of attempting to identify what study features help to explain the differences between studies.

CONCLUSIONS

We undertook a systematic review of the empirical literature since 1990 to examine what factors explain the diversity of findings regarding hospital ownership and quality of care. Pooled estimates of ownership effects from this literature are sensitive to the subset of studies included and the extent of overlap among hospitals analyzed in the underlying studies. Meta-regression reveals that estimates of the relationship between hospital ownership and adverse patient outcomes differ systematically according to a study’s data source, time period examined, and region covered. Studies representative of the US as a whole tend to find lower quality among FPs than private nonprofits. More research on ownership, such as in-depth understanding of organizational decision making and market-level dynamics across a range of economies, can contribute to a better understanding of the institutional contexts in which ownership matters for provider performance.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge funding for this research from grant #050953 under the Robert Wood Johnson Foundation’s Changes in Health Care Financing and Organization (HCFO) Initiative. We thank Mantian Hu, Ania Aksan, and Pam Kum for excellent research assistance, and two anonymous referees for constructive criticism. We also thank Drs Dena Bravata, Ingram Olkin, and our expert panel members for providing valuable comments: Drs Kathleen Carey, Shin-Yi Chou, Leemore Dafny, Guy David, Mark Duggan, Darrel Gaskin, Martin Gaynor, Michael McCue, Jack Needleman, Michael Rosko, Joanne Spetz, Robert Town, and Kevin Volpp. The authors have no competing interests.

REFERENCES


