DOI: 10.1111/j.1475-6773.2007.00744.x
RESEARCH BRIEF

Using Performance Data to Identify Preferred Hospitals

Meredith B. Rosenthal, Mary Beth Landrum, Ellen Meara, Haiden A. Huskamp, Rena M. Conti, and Nancy L. Keating

Objective. To explore the implications of current approaches used by health plans and purchasers to identify preferred hospitals for tiered networks using cost and quality information.

Data Sources/Study Setting. 2002 secondary data from WebMD Quality Services on hospital quality and costs in five markets (Boston, Miami, Phoenix, Seattle, and Syracuse).

Study Design. We compared four alternative tiering strategies that combine information on quality and cost to designate "preferred" (defined as ranking in the top quartile) hospitals. Within each market we identified the sets of hospitals designated preferred according to each strategy and examined the overlap in these sets across strategies.

Principal Findings. Compared with identifying preferred hospitals based on quality scores only, we found little overlap with the sets of hospitals that would be preferred based on cost scores only, cost scores after applying minimal quality standards, and an equally weighted quality and cost measure. The last two approaches, commonly used and intuitively appealing strategies to identify high-value hospitals, led to substantially different results.

Conclusions. The lack of agreement among alternative strategies to combine cost and quality data for ranking hospitals suggests the need for clear prioritization by payers and the application of more rigorous methods to identify high-value hospitals.

Key Words. Quality improvement, report cards, hospitals, referral networks

Employers and health plans are experimenting with a variety of market-based approaches to improve the value of health care spending. Many have adopted or are contemplating benefit designs with tiered provider networks that differentiate point-of-service copayments based on the cost and quality of the patient's choice of provider (Gabel, Lo Sasso, and Rice 2002; Mays, Claxton, and Strunk 2003; Robinson 2003, 2004; Rosenthal and Milstein 2004; Steinbrook 2004). Approximately 13 percent of individuals with employer-sponsored insurance—more than 20 million people—were enrolled in plans

with tiered provider networks in 2005 (Kaiser Family Foundation and Health Research and Education Trust 2005). To construct tiered benefit designs, payers must choose and reconcile multiple measures of quality and/or cost to determine which providers patients should be encouraged to select. The implications of empirical strategies to combine cost and quality information are unclear, as these two parameters are not consistently and systematically related (Wennberg, Fisher, and Skinner 2002; Fisher et al. 2003).

In practice, payers have used a variety of strategies to designate "preferred" providers (typically, but not always, hospitals). Some have done so based only on cost information, others based only on quality information, and still others have attempted to reconcile both cost and quality information into a single index of performance (Robinson 2003). One key question is whether these alternative strategies differ with regard to the set of providers deemed preferred. For example, if the best quality providers are also the least costly as some have asserted,[1] then there should be substantial overlap between the highest quality and lowest cost providers, and ratings that combine cost and quality information should not differ much from ratings on either scale.

In this research brief, we examine the sensitivity of hospital tiering outcomes (i.e., the set of hospitals identified as preferred) to whether and how quality and cost information is incorporated into the rating system. Assuming that the purpose of tiering hospitals is to increase the value, or benefits net of costs, from health spending, our goal is to illustrate the challenges a payer seeking to steer enrollees to high-value hospitals might face. We simulate several alternative rating schemes using the same information currently relied on by many health plans to designate groups of hospitals as "preferred" or "top tier." We examine and compare the implications of two extreme strategies—using quality data or cost data alone for hospital tiering—as well as two alternatives that integrate quality and cost ratings. All four approaches that we simulate have commercial analogs (Table 1). For example, Tufts Health Plan in Massachusetts classifies hospitals into tiers by applying separate thresholds for performance on cost and quality indices (e.g., "best" hospitals are those

---

Address correspondence to Meredith B. Rosenthal, Ph.D., Department of Health Policy and Management, Harvard School of Public Health, 677 Huntington Avenue, Room 405, Boston, MA 02115. Mary Beth Landrum, Ph.D., Ellen Meara, Ph.D., and Haiden A. Huskamp, Ph.D., are with the Department of Health Care Policy, Harvard Medical School, Boston, MA. Rena M. Conti, Ph.D., is with the Department of Pediatrics, Institute for Molecular Pediatric Sciences, CheSS, University of Chicago, Chicago, IL. Nancy L. Keating, M.D., M.P.H., is with the Division of General Internal Medicine, Brigham and Women's Hospital and the Department of Health Care Policy, Harvard Medical School, Boston, MA.

Table 1: Description of Alternative Strategies of Identifying Preferred Hospitals

Strategy	*Description*	*Description of Simulated Strategy*	*Analogous Real-World Example*
1	Overall medical/surgical* quality	Ranks weighted† average of summary quality‡ scores	Boeing waived copayments for employees using hospitals that meet criteria for quality set by the Leapfrog Group
2	Overall medical/surgical cost	Ranks weighted† average of standardized cost measures§	Blue Shield of California initially tiered hospitals based on a cost per episode measure before including quality measures in formula
3	Equal weighting of overall medical/surgical quality and cost	Ranks difference in standardized overall quality‡ score from overall cost score	An anonymous regional PPO rates physicians based on an overall performance measure computed as the sum of a standardized cost score and two times a standardized quality score
4	Overall medical/surgical cost with quality threshold	Ranks top 75% of hospitals according to overall quality based on overall cost scores	Tufts Health Plan in Massachusetts rates hospitals on standard measures of quality already being reported for regulatory/accreditation purposes and a measure of health plan cost; hospitals are categorized as good/better/best based on whether they exceed a cost or quality threshold, or both

*Adult medical and surgical admissions include: angioplasty, CABG, cardiac catheterization, colon surgery, COPD, craniotomy, disc surgery, gall bladder removal, hip replacement, heart attack, congestive heart failure, heart valve replacement, irregular heartbeat, knee replacement, pneumonia, prostatectomy (radical and transurethral), spinal fusion, stroke, total abdominal hysterectomy, total mastectomy for cancer, and urinary tract infection.

†Each category of admission is weighted by its share of all admissions of that type in the market.

‡The summary quality score is a weighted average of the standardized measures of mortality, complications, failure to rescue, volume, ICU staffing, and computerized physician order entry. The weighting scheme we rely on was developed over time by WebMD Quality through developing hospital ranking methods tailored for their health plan and employer clients. The weights reflect both values about the importance of each measure and perceptions of the strength of the measures themselves. This weighting scheme is essentially used as the default or starting point for summarizing hospital quality, although individual clients may choose to tailor the weights.

§Costs are computed using charge data collected by state agencies that maintain inpatient discharge databases. Hospitals report charges by department for each discharge and separately file cost reports to the Centers for Medicare and Medicaid Services (CMS) that include cost-to-charge ratios for each hospital department. These ratios were then used by WebMD to compute department-level costs. Total costs are then computed by summing the calculated department-level costs.

PPO, preferred provider organization; COPD, chronic obstructive pulmonary diseases; CABG, coronary artery bypass graft; ICU, intensive care unit.

that are both above the threshold for quality and below the threshold for cost) for three categories of care: adult medical/surgical, pediatric, and obstetrical admissions. We were particularly interested in examining the implications of two strategies that combine cost and quality information to identify high-value hospitals: the application of a separate threshold for cost and/or quality and the use of a single index that is the (weighted or unweighted) sum of standardized quality and cost scores.

METHODS

Data

We obtained 2002 data on hospital quality and costs from WebMD Quality Services ("WebMD," formerly HealthShare Technology Inc.), a private vendor of hospital quality information and decision support. At present, health plan clients representing over 30 million members are using WebMD's tools for hospital tiering, pay-for-performance, and center of excellence initiatives. WebMD's consumer tools are available to employers and health plans representing another 90 million members. Moreover, numerous other vendors, health plans, and employers are using similar measures to rate hospitals based on the same all-payer data. The data supplied by WebMD include quality and cost measures constructed from state hospital all-payer discharge data, which include information on all admissions, and other publicly available patient safety indicators collected by the Leapfrog Group. All-payer hospital discharge data were available for 24 states from which we purposively selected five metropolitan areas to illustrate how the tensions inherent in network tiering might vary by market: Boston, Miami, Phoenix, Seattle, and Syracuse. These markets were selected because they had sufficient numbers of hospitals to make tiering a reasonable prospect (which we deemed to be at least 10), included some markets with larger numbers of hospitals, and were geographically diverse.

Quality Measures

We used four claims-based measures of inpatient quality and patient safety (mortality, complications, failure-to-rescue, and volume) and two structural quality measures obtained from the Leapfrog Group (intensive care unit [ICU] intensivist staffing and computerized physician order entry [CPOE]). A complete description of the quality measures and their specification is provided in Appendix A (available online). Because individual complication measures apply differentially to specific admission types (e.g., some refer specifically to

surgical complications, which would be irrelevant for medical admissions), the specific mix of complications included in the complication score varies by condition. The claims-based measures were calculated separately for each of 22 adult medical or surgical admission types (Appendix B, available online). To facilitate comparisons among hospitals, WebMD adjusts claims-based performance measures to account for differences in case mix across hospitals using the APR™-DRG mortality score (3M Health Information Systems 2005).

We created a quality summary score for each of the 22 admission types by first standardizing each risk-adjusted measure so that the within-market mean equaled zero and the standard deviation equaled one. Next, we computed the quality summary score for each admission type as the weighted sum of the standardized measures for mortality (0.2), complications (0.3), failure to rescue (0.1), volume (0.3), ICU (0.05), and CPOE (0.05). This weighting scheme, supplied by WebMD, reflects a typical approach of their clients who are using quality summary scores for tiering or pay-for-performance purposes.[2]

We then computed summary measures of the overall quality of adult medical/surgical care as the weighted average of quality summary scores across all 22 adult medical/surgical admissions. The weight on each admission equaled the sum of all admissions of that type across hospitals in a market divided by the sum of all admissions in a market (its market-level share). These weights reflect the presumption that purchasers will prioritize clinical areas in assessing hospital performance based on the distribution of admissions across all hospitals in the market. When hospitals had no admissions of a particular type, the weights for the remaining categories were recalibrated to sum to one.

Cost Measures

Because the data included charges rather than actual payments to hospitals, WebMD computed a hospital cost measure by first averaging hospital-reported charges by department for each admission type and then applying department-level Medicare cost-to-charge ratios to the charge data before summing charges to the admission level. Cost measures were risk adjusted using the APR™-DRG severity score and standardized as above. Overall medical/surgical costs were computed as weighted averages across admission types using the same approach as for quality.

Alternative Tiering Strategies

We examined alternative approaches to ranking hospitals based on quality and cost measures to understand potential implications of various tiered

network designs (Table 1). Strategy 1 ranked all hospitals based on the overall quality of adult medical/surgical care. Strategy 2 ranked hospitals based on the average cost per admission for all adult medical/surgical care. Strategy 3 ranked hospitals after equally weighting overall quality and the negative of cost scores for adult medical/surgical admissions. Strategy 4 identified a minimal threshold for adult medical/surgical quality (top 75 percent) and then ranked hospitals meeting this minimal standard based on lowest costs.

Analyses

Within each market, we compared sets of preferred hospitals, defined as hospitals in the top quartile, among the alternative tiering strategies. First, we examined the degree to which a simple tiering mechanism based on overall medical/surgical quality (Strategy 1) would yield the same set of preferred hospitals as alternative strategies that incorporated payer preferences for lower cost hospitals (Strategies 2–4). We calculated the percentage of hospitals identified as preferred based on overall medical/surgical quality (Strategy 1) that were also preferred under Strategies 2–4. We also compared the overlap in preferred hospitals identified by Strategy 3 compared with Strategy 4 to examine the agreement in the two principal approaches used to combine cost and quality in a single index.

RESULTS

Table 2 compares the preferred hospitals across alternatives varying the relative importance of cost and quality information. Because high-quality hospitals did not have lower costs in most markets, there is little overlap between sets of preferred hospitals across alternatives. Few hospitals with the lowest overall costs for adult medical/surgical admissions were preferred hospitals based on overall quality, and comparisons of rankings based on cost alone to those based on quality alone revealed little agreement. The extent of overlap between the top quartile of hospitals according to quality versus cost, however, varied across markets with greater concordance in Miami, Boston, and Phoenix than in the other two markets. The combined index of cost and quality demonstrated fair to moderate agreement with rankings based on quality information alone, with overlap in preferred hospitals ranging from 50 to 64 percent. The strategy ranking hospitals on costs with a quality threshold (ranking only hospitals with quality scores above the 25th percentile) produced a set of preferred hospitals with little overlap with the group of hospitals

Table 2: Comparison of Strategies to Combine Cost and Quality Relative to Overall Quality

	Boston (N= 47)	*Miami (N= 30)*	*Phoenix (N= 27)*	*Seattle (N= 20)*	*Syracuse (N= 17)*
Among preferred hospitals for overall medical/surgical quality (Strategy 1), % preferred under					
Strategy 2: medical/surgical cost	18	29	17	0	0
Strategy 3: equal weighting of medical/surgical cost and quality	64	57	50	60	50
Strategy 4: medical/surgical cost with quality threshold	33	29	33	20	0
Among preferred hospitals for equal weighting of medical/surgical cost and quality (Strategy 3), % preferred under					
Strategy 4: medical/surgical cost with quality threshold	67	57	71	40	75

Source: Performance scores for each strategy based on authors' calculations using WebMD Quality Services data; top tier is defined as ranking in the top quartile.

that were preferred in terms of overall medical/surgical quality alone (0–33 percent). Comparing the two methods of combining cost and quality that have been used to date, we found more substantial overlap, although the extent of agreement regarding preferred status varied by market. In Seattle, for example, only 40 percent of hospitals designated preferred according to an equal weighting of cost and quality (Strategy 3) were also preferred when applying separate thresholds for cost and quality (Strategy 4).

DISCUSSION

Our simulations suggest that conceptually appealing and commonly used methods of identifying "preferred" hospitals often yield very different results. In particular, we found little overlap between the set of high-quality and low-cost hospitals, a finding that is consistent with previous work on geographic area variation (Wennberg, Fisher, and Skinner 2002; Fisher et al. 2003). Lack of agreement between quality and cost measures suggests that purchasers will have to make explicit tradeoffs between these two objectives. Moreover, we found that two common methods of reconciling cost and quality to identify high value can also lead to substantially different conclusions about which hospitals should be preferred.

Conceptually, neither the notion of setting separate thresholds for performance on measures of quality and cost, as in our Strategy 4, nor summing

separate cost and quality indexes, as in our Strategy 3 can guarantee that the highest value hospitals are identified. In particular, these approaches do not permit appropriate weighing of the magnitudes of incremental cost and quality differences. To illustrate, a hospital just below the quality threshold with 10 more complications than the hospital just above the threshold, but costs that are $1,000,000 lower, may be preferred by payers. The relative ranking of these two hospitals should depend upon whether the payer is willing to pay more or <$1,000,000 to avoid 10 complications. A systematic approach to making tradeoffs between cost and the individual components of quality is needed. Conceptually, the problem resembles a cost-effectiveness analysis in which the hospitals become the health care interventions to be ranked. Preferred status would then be a function of the payer's budget and the incremental cost effectiveness of a particular hospital relative to the closest alternative.

It is important to note that designers of tiered networks face another tension in designating preferred hospitals: which admission types to combine for the purposes of tiering. Previous research has shown that quality of care within a hospital varies widely by department (Jha et al. 2005). Thus, high-quality medical/surgical care may not translate to high-quality obstetrics care, and even within departments, hospitals that excel at hip replacement surgery may not excel at coronary artery bypass grafting. To accommodate this variation, payers may need to identify multiple sets of preferred hospitals that vary by condition or groups of conditions. For example, Tufts Health Plan's Navigator product tiers hospitals separately for pediatric, obstetrical, and other adult admissions and in theory separate networks could be identified at the department or procedure level. The natural limitation to extending this approach is the confusion such a benefit design might cause enrollees, diluting any hoped-for behavioral response to copayment differentials (Hibbard et al. 2002; Hibbard and Peters 2003).

Our analysis is intended to be exploratory and has several limitations. Most importantly, the quality and cost measures are based on administrative data, which may confound results for two reasons. First, high-quality hospitals may have better reporting of complications and thus incorrectly appear lower quality. Second, while WebMD Quality uses standard risk-adjustment methods, claims data are limited as sources of clinical information for risk adjustment (Krumholz et al. 2002). Better risk adjustment could result in very different patterns of cost and quality within and between hospitals. Purchasers seeking to improve the value of care through the use of performance data might be well-served by investments that would generate better data for risk adjustment, such as medical record abstraction (Keating et al. 2003) or im-

proving the richness of information recorded on medical claims (Dudley, Rittenhouse, and Bae 2002). Nonetheless, our analysis used data that are most readily available to payers and the questions we highlight will be relevant to any set of quality and cost metrics for any set of providers.

The cost data present another limitation. Because data on actual payments to hospitals were unavailable, we used a cost measure based on adjusting charges using the Medicare cost-to-charge ratios as a proxy for resource use. To understand the true cost of a particular hospital selection, payers should use their own payment algorithms to compute cost indices representing the full cost implications to the plan of steering a patient to the hospital in question, including the cost of the admission, subsequent readmissions, and postacute care. In addition, the risk-adjustment approach used by WebMD (APR™-DRGs) may not be ideal for examining cost–quality tradeoffs as we do here because it adjusts for complications that may be preventable. Thus some of the costs associated with poor quality care are not accounted for in the final scoring because they are attributed to patient severity of illness. Ideally, the risk-adjustment method would adjust only for conditions present upon admission.

Third, because we were unable to obtain detailed information on payers' specific formulas for tiering, our alternative strategies are somewhat stylized. For example, we adopted the weighting scheme used by many of WebMD's clients, which places substantial emphasis on volume, but many other weighting systems are possible and might alter the results. In exploratory work we found that rankings were often sensitive to whether and how individual quality measures were included, consistent with other work on composite quality measures (Jacobs, Goddard, and Smith 2005). Many health plans do not reveal to the public the details of their ranking methodology. In our view, this lack of transparency is counterproductive. Patients selecting a health plan with a tiered network need to understand the implications for their care choices and, once enrolled, should have granular information on quality measures by condition so they can choose whether to pay the difference in copayment to go to a nonpreferred hospital.

With no "silver bullet" in sight, health plans and purchasers of health benefits will continue to experiment with tiered networks as one strategy for improved quality and cost efficiency. The first generation of tiered networks has evolved and already many payers are taking a more sophisticated approach to measuring cost efficiency and quality. Without parallel development of methods to translate these data into benefit designs that steer patients to the highest value providers, however, tiered networks will fail to achieve their critical objectives of improving the affordability and quality of care.

ACKNOWLEDGMENTS

We are grateful to Rick Siegrist, M.B.A., M.S., Senior Vice President and General Manager of WebMD Quality Services and Bob Tavares of United HealthCare (formerly of WebMD Quality) for providing access to the data as well as advice and technical support. Thanks also to Arnold Epstein, M.D., M.A., Tom Lee, M.D., and David Blumenthal, M.D., M.P.P. for helpful comments on an earlier draft. Financial support for the project from the Harvard University Program on Health Systems Improvement is gratefully acknowledged.

NOTES

1. For example, the Bridges to Excellence (an employer-organized nonprofit entity that has licenses a pay-for-performance program for physicians) website states "Better Quality Costs Less." Available at http://www.bridgestoexcellence.org/about_us/home.htm [accessed March 8, 2007].
2. We also conducted several sensitivity analyses related to the quality summary score; these are reported in online Appendix C.

REFERENCES

3M Health Information Systems. 2005. *All-Patient Refined Diagnosis Related Groups.* Wallingford, CT: 3M Health Information Systems.

Dudley, R., D. Rittenhouse, and R. Bae. 2002. *Creating a Statewide Hospital Quality Reporting System.* Oakland, CA: California Health Care Foundation.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder. 2003. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138 (4): 288–98.

Gabel, J. R., A. T. Lo Sasso, and T. Rice. 2002. "Consumer-Driven Health Plans: Are They More Than Talk Now?" *Health Affairs (Millwood)* 22 (suppl web exclusives): W395–407.

Hibbard, J. H., and E. Peters. 2003. "Supporting Informed Consumer Health Care Decisions: Data Presentation Approaches That Facilitate the Use of Information in Choice." *Annual Review of Public Health* 24: 413–33.

Hibbard, J. H., P. Slovic, E. Peters, and M. L. Finucane. 2002. "Strategies for Reporting Health Plan Performance Information to Consumers: Evidence from Controlled Studies." *Health Services Research* 37 (2): 291–313.

Jacobs, R., M. Goddard, and P. C. Smith. 2005. "How Robust Are Hospital Ranks Based on Composite Performance Measures?" *Medical Care* 43 (12): 1177–84.

Jha, A. K., Z. Li, E. J. Orav, and A. M. Epstein. 2005. "Care in U.S. Hospitals—The Hospital Quality Alliance Program." *New England Journal of Medicine* 353 (3): 265–74.

Kaiser Family Foundation and Health Research and Education Trust. 2005. *Employer Health Benefits 2005 Survey.* Menlo Park, CA: Kaiser Family Foundation and Health Research and Education Trust.

Keating, N. L., M. B. Landrum, B. E. Landon, J. Z. Ayanian, C. Borbas, and E. Guadagnoli. 2003. "Measuring the Quality of Diabetes Care Using Administrative Data: Is There Bias?" *Health Services Research* 38 (6, part 1): 1529–45.

Krumholz, H. M., S. S. Rathore, J. Chen, Y. Wang, and M. J. Radford. 2002. "Evaluation of a Consumer-Oriented Internet Health Care Report Card: The Risk of Quality Ratings Based on Mortality Data." *Journal of the American Medical Association* 287 (10): 1277–87.

Mays, G. P., G. Claxton, and B. C. Strunk. 2003. "Tiered-Provider Networks: Patients Face Cost-Choice Trade-Offs." *Issue Brief Cent Stud Health Syst Change* (71): 1–7.

Robinson, J. C. 2003. "Hospital Tiers in Health Insurance: Balancing Consumer Choice with Financial Incentives." *Health Affairs (Millwood)* 23 (suppl web exclusives): W3-135–46.

———. 2004. "Reinvention of Health Insurance in the Consumer Era." *Journal of the American Medical Association* 291 (15): 1880–6.

Rosenthal, M., and A. Milstein. 2004. "Awakening Consumer Stewardship of Health Benefits: Prevalence and Differentiation of New Health Plan Models." *Health Services Research* 39 (4, part 2): 1055–70.

Steinbrook, R. 2004. "The Cost of Admission—Tiered Copayments for Hospital Use." *New England Journal of Medicine* 350 (25): 2539–42.

Wennberg, J. E., E. S. Fisher, and J. S. Skinner. 2002. "Geography and the Debate over Medicare Reform." *Health Affairs (Millwood)* 22 (suppl web exclusives): W96–114.

SUPPLEMENTARY MATERIAL

The following supplementary material for this article is available:

Appendix A: Quality Measure Definitions.

Appendix B: Admission Types Included in Analyses.

Appendix C: Sensitivity Analyses.

This material is available as part of the online article from http://www.blackwell-synergy.com/doi/abs/10.1111/j.1475-6773.2007.00744.x (this link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.