Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts?

Results from a study in Wisconsin suggest that making performance information public stimulates quality improvement.

by Judith H. Hibbard, Jean Stockard, and Martin Tusler

ABSTRACT: This study evaluates the impact on quality improvement of reporting hospital performance publicly versus privately back to the hospital. Making performance information public appears to stimulate quality improvement activities in areas where performance is reported to be low. The findings from this Wisconsin-based study indicate that there is added value to making this information public.

PUBLIC REPORTING OF HEALTH CARE PERFORMANCE has grown substantially in recent years, and considerable resources are spent on quality measurement and reporting. Yet it is unclear what impact, if any, these activities have had on quality improvement. Further, the relative impact of reporting for consumer choice (public reporting) versus reporting for internal consumption (private reporting) on providers' motivation to improve has not been examined.

Most proponents of the public release of health care performance information believe that making this information public will increase health care providers' motivation to improve. Motivation is thought to be driven by a desire to protect or enhance public reputation or market share, or both. Simply knowing that performance is inadequate may not be sufficiently motivating.

Existing evidence on the efficacy of publicly reporting to stimulate improvements is mixed.[1] Some studies have found that hospital mortality is reduced following the release of performance data; other studies report no effect of public releases.[2] Very few studies have looked at the impact of public reporting on subsequent quality improvement efforts.[3] The strength of the research designs and the quality of the reporting efforts may contribute to the mixed findings. Almost no evaluations of the impacts of public performance reports have used controlled experimental designs. Most studies assess performance before the public release of information and again after the release.[4]

There is also a great deal of variation in how well the reports are designed and

Judith Hibbard and Jean Stockard are professors in the Department of Planning, Public Policy, and Management at the University of Oregon in Eugene. Martin Tusler is a data analyst in the same department.

disseminated. If the efficacy of making performance information public rests on the ability of reports to increase providers' concerns about their public reputations or market share, then the reports themselves must be widely disseminated to the public. They must also be very easy to understand, and the data must be displayed so that the reader can effortlessly discern which are the high- and low-performing providers. If either of those factors is not present, then the "public" report may not be sufficient to motivate improvements.

The 'QualityCounts' Report

Our study uses an experimental design to evaluate the impact of a public hospital performance report on subsequent hospital quality improvement efforts. The report on hospital safety was produced and disseminated by the Alliance, a large employer-purchasing cooperative in the Madison, Wisconsin, area. It was carefully designed to be easy to use by consumers and was widely disseminated in the community. Other Wisconsin hospitals, not in the public report, were randomly assigned to receive either a private report on their own performance or no report. The report, titled *QualityCounts*, compared the performance of twenty-four hospitals in south central Wisconsin. Two summary indices of adverse events (deaths and complications) occurring within the broad categories of surgery and nonsurgery were included, along with indices in three individual clinical areas: hip/knee surgery, cardiac care, and maternity care (Exhibit 1). Hospitals were rated as better than expected (fewer deaths/complications), as expected, or worse than expected. The data were derived from the Wisconsin Bureau of Health Information inpatient public use data sets. Measures were adapted from the original Healthcare Cost and Utilization Project (HCUP) quality indicators and were summarized and risk adjusted by Medstat using its disease staging methodology. The same data for the other Wisconsin hospitals, risk-adjusted, were used for the creation of both private hospital reports and control variables in our analysis.

Several aspects of this public report were unique and may have helped enhance the "publicness" of the report. First, the report was designed to be highly evaluable for consumers. Previous controlled laboratory studies have found that if comparative performance data were presented in a way that made it easy to discern high and low performers on a quality report (the data were highly "evaluable"), the information was more likely to actually get weighted and used in choices.[5] For example, hospitals were ordered by performance in the *QualityCounts* report, with the top performers at the top and the poor performers at the bottom. Further, the top-tier performers were highlighted in the report with a color band (Exhibit 1).

Second, a concerted effort was made to widely disseminate the *QualityCounts* report to the public. It was inserted into the Madison newspaper; there were newspaper stories about the report; and Alliance employers sent it to employees' homes. It was also available on a Web site, and copies were distributed by commu-

EXHIBIT 1
***QualityCounts* Report Format**

Regional hospitals	Surgery[a]	Nonsurgery[a]	Hip/knee	Cardiac	Maternity
Hospital A	(+)	(+)	(+)	(+)	○
Hospital B	(+)	(+)	(+)	○	○
Hospital C	(+)	(+)	(+)	○	(–)
Hospital D	(+)	(–)	(+)	(–)	–[b]
Community hospitals					
Hospital F	(+)	(+)	(+)	○	(+)
Hospital G	(+)	(+)	(+)	(+)	○
Hospital H	(+)	(+)	(+)	○	○
Hospital I	(+)	(+)	(+)	○	○
Hospital J	(+)	(+)	(+)	○	○
Hospital K	(+)	(+)	(+)	○	○
Hospital L	(+)	(+)	(+)	(+)	(–)
Hospital M	(+)	(+)	○	○	–[b]
Hospital N	(+)	(+)	(+)	○	(–)
Hospital O	(+)	(+)	(+)	(–)	○
Hospital P	(+)	(+)	(+)	○	(–)
Hospital Q	(+)	(+)	○	○	(–)
Hospital R	(+)	(+)	(+)	(–)	(–)
Hospital S	(+)	○	(+)	○	–[b]
Hospital T	○	(+)	○	○	–[b]
Hospital U	(+)	○	(+)	○	○
Hospital V	(+)	○	(+)	○	(–)
Hospital W	○	○	–[c]	–[c]	○
Hospital X	○	○	–[c]	–[c]	○
Hospital Y	○	○	○	○	(–)

SOURCE: Mockup of the *QualityCounts* report, with hospital names removed.
NOTES: Plus signs indicate that there were fewer mistakes, complications, and deaths than expected. Circles mean that there was an average number of mistakes, complications, and deaths. Minus signs mean that there were more mistakes, complications, and deaths than expected.
[a] Summary measures.
[b] Hospital doesn't provide this type of care.
[c] Hospital didn't provide enough of this care during the study period to get a rating.

nity groups and at libraries. This was the first public report on hospital quality issued in this region, and it generated substantial public interest.

Finally, the *QualityCounts* report indicated that there was sizable variation in performance in two key clinical areas: maternity and cardiac care. Eight hospitals had poor scores in obstetrics, and three had poor scores in cardiac care (Exhibit 1). Many performance reports fail to show much variation and often do not capture the public's attention. Because of the variations in this report, it may have been more interesting to consumers.

Evaluation Methods

The *QualityCounts* report was sponsored, produced, and disseminated by the Alliance. The evaluation, however, was carried out by university-based researchers

with separate foundation funding.

■ **Design.** The evaluation used an experimental design with two intervention groups and one control group. The primary intervention group was the twenty-four hospitals in south central Wisconsin in the Alliance service area and included in the public *QualityCounts* report. (They also received a more detailed report on their own performance.) These twenty-four hospitals were not randomly selected or assigned but were included as an intervention group because they were in the public report. The other ninety-eight general hospitals in Wisconsin were randomly assigned to either the secondary intervention condition or the control condition. Those assigned to the secondary intervention group received a private report on their own performance (performance information was not made public).[6] Those assigned to the control group received no report. Seven hospitals were eliminated from the analysis because the performance data were unavailable or incomplete, leaving 115 hospitals in the study.[7]

Even though the hospitals in the primary intervention group (public report) were not randomly assigned, we observed no statistically significant differences among the three groups of hospitals in characteristics or size, nor in their pre-report levels of performance (Exhibit 2). The public report was widely disseminated in the fall of 2001, and the private report was also delivered later that fall. In May 2002 a survey of all 115 hospitals was conducted. To avoid sensitization of the hospitals, no baseline measures were carried out prior to the release of the report. Our analysis focuses on differences among the three groups of hospitals (public-, private-, and no-report hospitals) approximately nine months after the report's release.

EXHIBIT 2
Comparison Of Hospital Characteristics And Baseline Performance In The Three Experimental Conditions

Hospital characteristic	Public-report hospitals (n = 24)	Private-report hospitals (n = 41)	No-report hospitals (n = 46)
Average total beds	97	124	96
Average total ICU beds	10	14	11
Average total inpatient days	19,098	23,304	18,125
Baseline performance			
"Better than expected" on surgery	83%	88%	89%
"Better than expected" on nonsurgery	71	71	70
"Better than expected" on hip/knee	71	77	88
"Better than expected" on cardiac	13	23	9
"Better than expected" on maternity	5	21	26

SOURCE: Wisconsin Bureau of Health Information.
NOTES: None of the differences among hospital types were statistically significant (p = .244–.783). Intensive care unit (ICU) beds include medical-surgical, cardiac, pediatric, mixed, neonatal, and other.

Respondents to the May 2002 survey included chief executive officers (CEOs), medical directors, and quality improvement directors at each hospital. Respondents in the no-report hospitals and those in the private-report hospitals were sent a copy of the public *QualityCounts* report in the advance letter. Participants were also asked in the advance letter to respond to a Web-based survey. Nonrespondents were given the opportunity for a telephone interview. The overall response rate was 62 percent, with 94 percent of the hospitals providing at least one respondent. Seventy-five percent of respondents in the public-report group, 65 percent in the private-report group, and 52 percent in the no-report group completed the survey.

■ **Data analysis.** When the focus of our analysis is on attitudes and concerns regarding public reporting, our unit of analysis is the individual, and all the respondents in these three roles (CEO, quality improvement director, and medical director) are included. However, when the focus is on the hospital's response to the report or hospital quality improvement activities, the unit of analysis is the hospital, and the analysis includes only one respondent per hospital. When the question is about market share or reputation, the CEO is the respondent for the hospital (with the medical director or quality improvement director as backup when the CEO is a nonrespondent). When the question is about specific quality improvement activities, the quality improvement director is the respondent for the hospital (with the medical director or the CEO as backup).

In addition to the analysis reported here, the data were reanalyzed to assess whether the same question posed to a different hospital respondent (for example, a hospital's CEO instead of its quality improvement director) yielded the same or different results. All analyses showed the same pattern of responses regardless of which respondent was reporting. That is, there was a high degree of agreement among the respondents within the hospitals.

Comparison of means with analysis of variance is the primary analytic approach used in the study. For some of the analyses we use quality scores as a control variable. We also sometimes focus only on responses from the low-scoring hospitals (those most in need of improvement). Because obstetrical scores showed the most variance, we use this score to define which are the low-scoring hospitals.

Evaluation Findings

■ **How hospitals view the report.** When asked how hospitals view public reporting on quality performance in general, there were no differences among the respondents in the three experimental conditions (public, private, and no report). On average, respondents in all three conditions were slightly negative about the idea.

However, when asked specifically about the *QualityCounts* report, there were significant differences among the respondents in the three conditions. When we asked about the validity of the data in the report, we found statistically significant differences among the respondents (scores range from 1, "not at all," to 5, "very").

Those in the public-report group were most negative about the validity of the report (mean = 2.1), and those in the private-report group were most positive (mean = 2.6), with the no-report group falling in between (differences among the three groups of hospitals: $F = 4.2$, $df = 2, 152$, $p < .05$). The same pattern was observed when we asked how appropriate the *QualityCounts* report is for the public's use (differences among the three groups of respondents: $F = 6.0$, $df = 2, 168$, $p < .01$) and how useful the report is for quality improvement (differences among the three groups: $F = 6.7$, $df = 2, 169$, $p < .01$). Thus, the public-report hospital respondents were significantly more negative about the public report, its validity, its value for quality improvement, and its appropriateness for the public's use than the other two groups of respondents were. The private-report respondents were the most positive about the public report on these same questions.

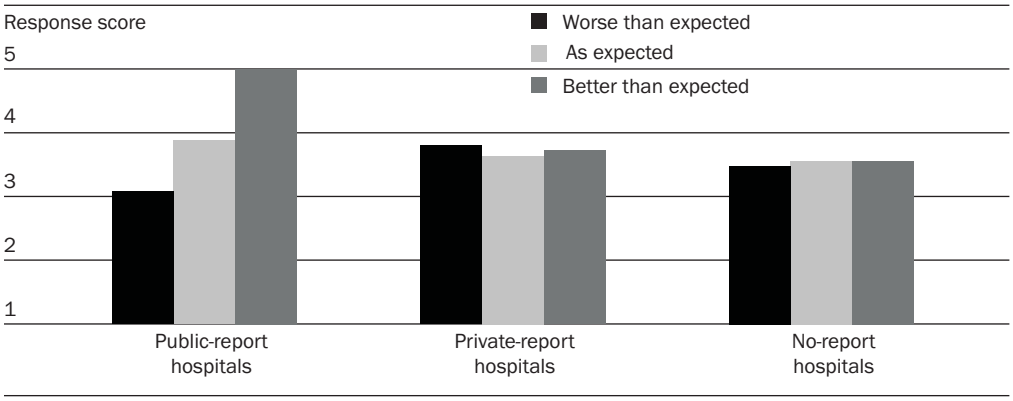
We also examined the relationship between performance scores (in obstetrics) and views regarding the validity of the data. Respondents in public-report hospitals with the lowest performance scores (in obstetrics) were most negative about the validity of the data ($r = .35$, $n = 43$, $p < .05$). There was no such relationship, however, among the private-report and no-report hospitals.

Even though respondents in the three groups varied in their views regarding the value of the *QualityCounts* report, they had similar suggestions when asked about how the report could be improved (an open-ended question). Respondents in all three groups raised concerns about the validity of the data, the quality and consistency of coding practices, and the lack of transparency of the risk-adjustment methodology used.

■ **Concerns about public image and market share.** Respondents were asked about the likelihood that the public report (or "one like it," in the case of the private and no-report hospitals) would affect their hospital's public image. Exhibit 3 shows the findings broken out by experimental condition and by hospitals' scores. As before, the scores used for this analysis are the hospitals' scores for obstetrics. For private- and no-report hospitals, performance scores were unrelated to what respondents thought a report (such as *QualityCounts*) would do to their hospital's public image. Most thought that it would neither detract from nor enhance it. However, among the public-report hospitals, responses differed significantly depending on their scores. Those with poor scores were more likely to say that the report would detract from, while those with good scores were more likely to say that the report would enhance, their public image. When the same question was asked about market share, there were no differences across the three conditions, and score level was unrelated to responses. Thus, among the public-report hospitals, the report was viewed as potentially affecting their public image but not their market share.

■ **Quality improvement efforts.** The respondents were asked where they had put most of their quality improvement efforts in the past year (from a list of seventeen potential areas).[8] There were no differences by experimental condition, and all three groups responded similarly. Pain control, restraints, and fall prevention were

EXHIBIT 3
Respondents' Belief That Public Reports Of Hospital Performance Will Enhance Or Detract From A Hospital's Public Image, On A Five-Point Scale, Among Hospitals In Three Study Conditions



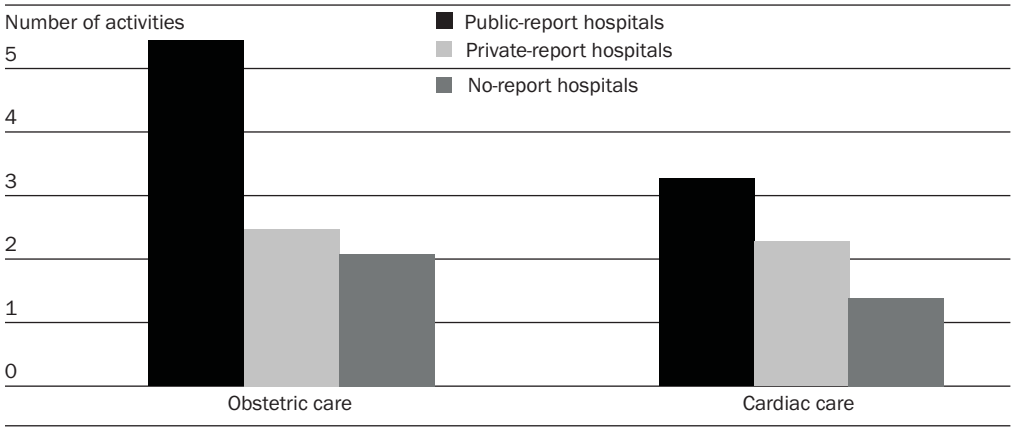
SOURCE: Survey data collected by the authors.
NOTES: Responses are for obstetric performance only, as an indicator of overall performance. The response score ranged from 1 (very likely to detract) to 5 (very likely to enhance), with 3 denoting neither enhance nor detract. For public-report hospitals, results were significant at $F = 5.8$, $df = 2,23$, $p < .05$. Results for the other two study-condition groups were not statistically significant.

the most commonly mentioned areas of effort among all three hospital groups (data not shown). Further, there were no differences by experimental condition in terms of the importance of different strategies to their overall quality improvement efforts. Board discussion on quality; the use of cross-functional work groups; and the use of incentives, rewards, and recognition to support quality improvement efforts were similar across the three groups of hospitals (data not shown). Thus, the main targets for quality improvement activities and the strategies used to support these activities are largely the same among the hospitals.

However, when it came to quality improvement efforts specific to the areas included in the public (and private) reports, a different pattern emerged. Respondents were asked to indicate whether their hospital was engaged in each of seven possible quality improvement activities in obstetrics (Exhibit 4). The public-report hospitals reported a significantly higher number of quality improvement activities in obstetrics (an average of 3.4 activities out of 7), while the private-report hospitals reported an average of 2.5 activities, and the no-report hospitals, 2.0 activities (significant differences are between the public-report group and the private- and no-report group combined: $F = 4.4$, $df = 1, 100$, $p < .05$). When respondents were asked about four cardiac quality improvement activities, a similar pattern was observed: Public-report hospitals reported an average of 2.5 activities; private-report hospitals, 1.5 activities; and no-report hospitals, 1.4 activities (differences across all three hospital groups: $F = 4.3$, $df = 2, 99$, $p < .05$).

Exhibit 4 shows just the hospitals with low scores in obstetrics and in cardiac care and the number of quality improvement efforts under way at these hospitals.

EXHIBIT 4
Average Number Of Quality Improvement Activities In Obstetrics And Cardiac Care, Among Hospitals With Worse-Than-Expected Scores, In Three Study Conditions



SOURCE: Survey data collected by the authors.
NOTES: Quality improvement activities for obstetrics included best practices around cesarean sections, best practices around vaginal birth after cesarean (VBAC), reducing third- or fourth-degree laceration, reducing hemmorhage, reducing prenatal complications, reducing postsurgical complications, and other. Activities for cardiac care included best practices for acute myocardial infarction, best practices for heart failure, best practices for community-acquired pneumonia, and other. For obstetrics, differences among hospitals in the three study conditions were statistically significant, $F = 8.3$, $df = 2, 31$, $p < .01$. For cardiac care, $F = 1.7$, $df = 2, 11$, ns.

Low-scoring public-report hospitals show the highest level of quality improvement activities, the private-report hospitals an intermediate level, and the no-report hospitals the lowest level. Although not all of these quality improvement activities were reported to have been initiated in the past year, more were initiated among the public-report hospitals than among the other two groups.

For hospitals with poor cardiac scores, a similar pattern is observed. However, because there are only fourteen hospitals with poor cardiac scores, the differences do not reach statistical significance (Exhibit 4).

Hemorrhage after delivery was perceived by the hospitals to be a major factor in the low obstetrical scores. If we look at the low-scoring hospitals and their quality improvement activities aimed specifically at reducing hemorrhage after delivery, 88 percent of the public-report hospitals reported quality improvement activities to reduce hemorrhage; only 27 percent did so in the private-report group; and 9 percent in the no-report group. Even though this analysis only includes thirty-four hospitals, these differences are statistically significant at the .001 level ($F = 10.0$, $df = 2, 31$). It is also interesting to note that although the private-report group was as informed about their own performance in obstetrics as the public-report hospitals were, they were far les likely to act on it (difference between public- and private-report hospitals: $F = 10.7$, $df = 1, 21$, $p < .01$).

Finally, the public-report hospital respondents were asked what they thought their scores would be in two years, when the next public report is issued. Three-fourths of the low-scoring hospitals (using the obstetrics scores) said that they

thought their scores would improve. The other 25 percent said that they were unsure about what would happen. Fifty-eight percent of the hospitals that scored in the "as expected" range also thought that their scores would improve. When asked how their scores would improve, 62 percent listed quality improvement efforts (many indicated other strategies as well); 15 percent indicated that they would improve only because of changes in coding practices. Thus, most of the hospitals were optimistic that they could improve their scores through attention to quality improvement.

Discussion

The findings provide strong evidence that making performance information public stimulates quality improvement in the areas where performance is reported to be low. Since quality improvement efforts among the public-report hospitals appear to be significantly greater than in hospitals given only private reports, there is added value to making performance information public.

The findings also indicate that making performance information public generated negative attitudes, anger, and distrust among the hospitals included in this report. A national set of standard measures, appropriately evidence-based and vetted by a credible organization, could help to remove some of the tension around the selection of performance measures for use in a public report. A national set of standard measures would also allow hospitals to have some confidence that what they are being held accountable for is fair and appropriate.

The National Quality Forum is working on a process to establish such a set of measures. Until they are available, however, there are strategies that could reduce some of the negative hospital response. For example, if more collaboration is sought with the hospitals in the reporting process and if they receive a preliminary round of reporting that is only privately shared, hospitals may react less negatively. Also, increasing the validity checks on the data and adopting more transparent methodologies for risk adjustment might allay some other objections.

Our findings suggest that the public-report hospitals viewed the report as affecting their public image but not their market share. Because the hospitals' concern for public image appears to be a key motivator for improvement, an important role for consumers in stimulating quality improvement may simply be increased attention to hospital quality. In addition, however, the hospitals' lack of concern with market share was unexpected and warrants further research.

■ **Study limitations.** Two other issues regarding further research are related to the generalizability of our findings. First, a limitation of our study design is the nonrandom assignment of the primary intervention group. Although the twenty-four hospitals in the public-report group were no different at baseline in performance or characteristics from the other two groups that were randomized, they may have responded differently to a public report because of the presence of the purchasing coalition and their collective experience of responding to the coalition's efforts in

"The public report must be designed so that it is immediately obvious who the top and bottom performers are."

promoting improvements. In other words, it is possible that because of prior activities, the Alliance had sensitized the hospitals that received the public report to quality issues, and, as a result, they were more responsive. Future research should examine whether hospitals are more responsive to reports sponsored by purchasers.

Second, our analysis examined quality improvement activities approximately nine months after a report's dissemination. Future research should address both the extent to which such activities continue and, even more importantly, the degree to which the observed increased quality improvement efforts yield actual improvements in outcomes.

Our measurement of quality improvement efforts was based on self-reports from the hospitals, and future examinations should, of course, include other data sources. Events occurring after the collection of the data reported in this analysis indicate, however, that the self-reported data gathered in our design are accurate. Very public and visible quality improvement efforts have grown out of the public performance report. For instance, a statewide coalition of perinatal care providers has initiated an educational campaign aimed at all hospitals in Wisconsin that offer obstetrical services. The educational campaign focuses on preventing, recognizing, and treating postpartum hemorrhage, a major source of low scores on maternity care. In addition, the Alliance has established a fund to invest in projects to improve providers' ability to measure and improve care. Hospitals can apply for funds to undertake specific projects. Two of the projects funded thus far address improvements in obstetrical outcomes, including reducing postpartum hemorrhage. All of these activities appear to be a direct result of the public report and act as corroborating evidence that quality improvement efforts were stimulated in the clinical areas showing poor performance.

■ **Key elements for success.** While the findings indicate that making performance public stimulates quality improvement activities, it is likely that three key ingredients, all of which occurred with our public reporting condition, are necessary to observe this result in other settings. First, the report must be widely disseminated in the community. Second, the hospitals need to know that a future public report will be produced and widely disseminated again within a year or two. Finally, the public report itself must be highly evaluable—that is, the public report must be designed so that it is immediately obvious who the top and bottom performers are, to stimulate quality improvement efforts. A public report that is very easy to understand and use has the potential to affect hospitals' public image.

The need to widely disseminate a report and the need to let it be known that a future report will be issued seem, on the face of it, effective strategies that do not need to be investigated. However, the necessity that the report be highly evaluable

to stimulate quality improvement efforts may require further investigation. This is a strategy that report sponsors will find difficult to implement; providers will vigorously oppose an approach that explicitly ranks or identifies top and bottom performers. However, it may be this very strategy that makes the difference between motivation to improve and no motivation.

The authors acknowledge the Robert Wood Johnson Foundation, which provided funding for this work through the Health Care Financing and Organization (HCFO) Initiative. The authors further acknowledge the panel of experts who provided input on the evaluation instrument development: Arnold Epstein, Jinnet Fowles, David Knutson, Harlan Krumholz, Jerod Loeb, Elizabeth Mort, and Thomas Schmidt.

NOTES

1. M.N. Marshall et al., "The Public Release of Performance Data: What Do We Expect to Gain? A Review of the Evidence," *Journal of the American Medical Association* (12 April 2000): 1866–1874.
2. E.L. Hannan et al., "Improving the Outcomes of Coronary Artery Bypass Surgery in New York State," *Journal of the American Medical Association* (9 March 1994): 761–766; G.T. O'Connor et al., "A Regional Intervention to Improve the Hospital Mortality Associated with Coronary Artery Bypass Graft Surgery: The Northern New England Cardiovascular Disease Study Group," *Journal of the American Medical Association* (20 March 1996): 841–846; D. Baker et al., "Mortality Trends for Medicare Patients Hospitalized with Six Medical Conditions during a Program to Publicly Report Hospital Performance" (Paper presented at the annual meeting of AcademyHealth, Atlanta, June 2001); J.D. Clough et al., "Lack of Relationship between the Cleveland Health Quality Choice Project and Decreased Inpatient Mortality in Cleveland," *American Journal of Medical Quality* (March 2002): 47–55; and W.A. Ghali et al., "Statewide Quality Improvement Initiatives and Mortality after Cardiac Surgery," *Journal of the American Medical Association* (5 February 1997): 379–382.
3. D.R. Longo et al., "Consumer Reports in Health Care: Do They Make a Difference in Patient Care?" *Journal of the American Medical Association* (19 November 1997): 1579–1584.
4. Marshall et al., "The Public Release of Performance Data."
5. J.H. Hibbard et al., "Strategies for Reporting Health Plan Performance Information to Consumers: Evidence from Controlled Studies," *Health Services Research* (April 2002): 291–313.
6. The private report was the same as the more detailed report that the public-report hospitals received on their own performance. The only difference was that the private report had less information about how peer hospitals performed.
7. Seven hospitals were eliminated because of inadequate data. From the private-report group, two hospitals were eliminated because they had closed; two hospitals had too few patients to calculate performance scores or did not offer the services evaluated (ability to calculate at least three of five performance indicators was required); and two hospitals shared administrative structures where the performance data were not separable. From the no-report group, one hospital was eliminated because it had too few indicators of performance. The problems with the adequacy of the data were discovered after the randomization of hospitals. However, even after the elimination of these seven hospitals, there still were no significant differences among the experimental hospital groups in terms of characteristics or performance.
8. The seventeen quality improvement areas included computerized physician order entry, staffing strategies, fall reduction, pain relief, restraints, computerized medical records, other information systems infrastructure, pediatric services, obstetrics/gynecology (maternity services), psychiatric services, emergency services, anesthesiology, radiology, orthopedics, cardiac services, intensive care unit, and other.