Public Reporting of Hospital Patient Satisfaction:

A Review of Survey Methods and Statistical Approaches

Prepared for the Rhode Island Department of Health
Health Quality Performance Measurement and Reporting Program

By
Qualidigm
Middletown, Connecticut

Judith K. Barr, ScD
Senior Scientist

Sara Banks, MPH, PhD
Senior Analyst

August 2002

Acknowledgements:

Jay Buechner, PhD, Rhode Island Department of Health; Cathy Boni, PhD, RN, CPHQ, Hospital Association of Rhode Island; and Tierney Sherwin, MPA, Qualidigm for their helpful comments on a draft version of this report.

Jean Marie Rocha, RN, MPH, Rhode Island Department of Public Health for managing the production of this report.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

An in-depth review of existing public reports on hospital patient satisfaction was conducted to assess the similarities and differences in survey and statistical methodologies, identify issues related to survey distribution and reporting for minority groups, and discuss advantages and disadvantages of various statistical approaches for reporting hospital patient satisfaction data. The relevant literature was reviewed as a context for the description and analysis of the public reports.

Nine public reports were identified that compare hospitals on patient satisfaction. These reports were intended for distribution to the general public. Where available, the companion technical report also was reviewed to provide greater detail on methods. The public reports present data for states (Rhode Island, California, Massachusetts, and Ontario), for regions (Southeast Michigan, Western New York), or for cities (Buffalo, Indianapolis, and Cleveland). Similarities and differences were found among the nine reports in four areas: survey methods, sampling procedures, computation of scores, and reporting scores. Key findings are listed below.

Survey methods

- Rhode Island is the only locale with a legislative mandate; all others are voluntary initiatives.
- All locales except Rhode Island conducted three waves of mailing, but only two had higher mean response rates than Rhode Island.
- Four locales, including Rhode Island, made alternate language surveys available, but methods of distribution varied.
- Only Rhode Island and Ontario used Parkside/Press Ganey as a survey vendor; all others used the Picker Institute.

Sampling procedures

- All locales sampled patients randomly.
- All locales except Ontario stratified the sampling of patients by hospital service.
- All but Ontario selected equal numbers from each hospital.
- All but Ontario required those sampled to be adults and to have had an overnight stay in the hospital.

Computation of scores

- The seven locales using the Picker survey computed “problem scores” or the inverse, “performance scores.” Rhode Island and Ontario computed scores based on 5-point evaluative response sets.

- Seven locales, excluding Rhode Island and Western New York, case-mix adjusted the hospital scores for patient characteristics (e.g., age, gender, health status). None of the locales adjusted for hospital characteristics.

Reporting scores

- All locales except Massachusetts required a minimum number of survey responses, ranging from 1 to 60 per service, for each hospital to be included in the report; Rhode Island required 40 per service.

- All but Ontario and California compared hospital scores to a national normative mean in the public report.

- None of the locales reported actual scores in the public report; all except Massachusetts assigned hospitals comparative ratings, e.g., *above*, *average*, or *below* a normative mean.

- All except Massachusetts used symbols to depict these comparative ratings in the public report; Massachusetts presented the hospital scores in a graphic format in the public report and presented comparative ratings only in their technical report.

- All except Ontario reported results for each domain of care by service type.

Survey and statistical issues in public reporting of hospital patient satisfaction are discussed and include: the large sample size needed to report results by minority status, the use of case-mix adjustment, and the methods for assigning hospitals to normative categories. The findings from this review provide a context for future health quality reporting in Rhode Island.

INTRODUCTION

As part of the Rhode Island Health Quality Performance Measurement and Reporting (HQPMR) program, the first public report on patient satisfaction with hospitals in Rhode Island was published and released in November 2001. This report is the only legislatively mandated public report on hospital patient satisfaction in the country at this time. In addition to the Public Report, a Technical Report, available on the Rhode Island Department of Health (HEALTH) web site, describes the methodology and results in greater detail (www.healthri.org/chic/performance/satisfaction.htm).

Because public reporting of hospital patient satisfaction in Rhode Island is an ongoing process, planning for the second round of measurement and public reporting is underway. To help inform this planning, an in-depth review of existing public reports on hospital patient satisfaction was conducted to complement an earlier publication on the state of public reporting (Department of Health of Rhode Island, 2000). The major purposes of this review were to provide a scan of the various survey and statistical methods used, identify challenges to reporting patient satisfaction for minority populations; and examine the statistical approaches for reporting patient satisfaction data. Specifically, the tasks were to:

- review the literature on issues related to measuring and reporting hospital patient satisfaction;
- assess similarities and differences in survey and statistical methodologies used in public reports of hospital patient satisfaction;
- discuss survey distribution and reporting for minority groups; and
- discuss advantages and disadvantages of various statistical methodologies for reporting hospital patient satisfaction data.

Following the literature review, the methods used for conducting the in-depth review of existing hospital patient satisfaction public reports are described, and the results are presented on the similarities and differences in survey and statistical methodologies used in existing public reports. Then, the Discussion section addresses: (1) issues related to survey distribution and reporting for minority groups, and (2) statistical issues in the computation and public reporting of hospital data (i.e., case-mix adjustment and statistical options for assigning comparative ratings to hospital scores).

LITERATURE REVIEW

To provide a context for the review of existing hospital patient satisfaction public reports and to add to the understanding of the advantages and disadvantages of different methodological approaches, a systematic search and review of the literature was conducted. This review focused on patient satisfaction, specifically in hospitals; the relation between patient characteristics and satisfaction scores; case-mix adjustment of satisfaction rankings; and reporting of minority status

and patient satisfaction. The results also can be used to provide a context for the decisions being made in the Rhode Island HQPMR program.

Patient Satisfaction Measures

One factor that can account for variation in patient perceptions of hospital care is differences in the measures of satisfaction. The patient satisfaction surveys developed by the Picker Institute focus on "experience of care" and take a problem-oriented approach, asking questions about what did or did not happen during the hospitalization with regard to various aspects of care (Cleary, et al., 1991). Other satisfaction surveys take a "satisfaction with care" approach, asking the individual to rate their satisfaction with various aspects of care while they were hospitalized (Finkelstein, et al., 1998; Kane, et al., 1997; Marshall, et al., 1996). These two approaches to assessing patients' views of their hospital experiences may reflect the two complementary but sometimes conflicting goals for developing such information: quality improvement by hospitals and public reporting for use by consumers. To help hospitals direct their quality improvement efforts, specific questions identifying problem areas have been used (Cleary, et al., 1991; Hargraves, et al., 2001). Whether results of these questions are more easily understood by the public in a report on hospital quality than questions asking patients to evaluate their satisfaction or rate the care received (e.g., excellent, good, fair, poor) is a methodological issue that has not been resolved.

Patient Characteristics

Most studies of the relationship of patient characteristics to hospital satisfaction scores have found that several key variables were significantly related to reports of satisfaction, most consistently patient age and self-reported health status. Virtually every study reviewed found these two characteristics to be strongly related to hospital satisfaction, and this finding held for VA hospital patients (Rosenheck, et al., 1997; Young, et al., 2000), for obstetrical patients (Finkelstein, et al., 1998), for different satisfaction measures (Marshall, et al., 1996), and in different countries (Thi, et al., 2002). In general, older patients tended to report greater satisfaction, and sicker patients tended to be less satisfied (Finkelstein, et al., 1998; Hargraves, et al., 2001; Rogut, et al., 1996; Rosenheck, et al., 1997; Thi et al., 2002; Young, et al., 2000). Other patient characteristics that have been significantly related to hospital patient satisfaction include: race/ethnicity (Finkelstein, et al., 1998; Rogut 1996; Young, et al., 2000), gender (Hargraves, et al., 2001; Rosenheck, et al., 1997), education level (Hargraves, et al., 2001), insurance status (Finkelstein, et al., 1998; Rogut, et al., 1996), income (Rogut, et al., 1996; Young, et al., 2000), having a regular physician (Rogut, et al., 1996), and past hospital experience (John, 1992).

A few studies found that hospital characteristics were related to patient reports of satisfaction. For example, differences by hospital service have been noted, with obstetrical patients most satisfied and surgical patients more satisfied than medical patients (Cleary, et al., 1991; Rogut, et al., 1996; Young, et al., 2000). Other hospital characteristics include: teaching status (Finkelstein, et al., 1998; Young, et al., 2000), rural location (Young, et al., 2000), and nurse staffing levels (Rogut, et al., 1996).

Case-Mix Adjustment

To address these differences, case-mix adjustment has been applied to the results in some studies to control for differential distribution of patient characteristics among hospitals when those characteristics are related to the outcome, i.e., satisfaction. Without some adjustment, these differences might systematically bias responses and lead to reporting and rating variation that is unrelated to quality of care (Elliott, et al., 2001). The literature on case-mix adjustment of patient satisfaction scores, similar to the process of risk adjustment for patient outcomes, is mixed. Findings from these analyses indicate that patient characteristics typically explain little of the variation among hospital patient satisfaction scores, for example, less than 10% and as little as 2% (Cleary, et al., 1991; Finkelstein, et al., 1998; Hargraves, et al., 2001; Kane, et al., 1997; Rosenheck, et al., 1997; Young, et al., 2000). Two studies reported that up to 15% of the variance was explained by patient characteristics (Rogut, et al., 1996; Young, et al., 2000). Despite the generally small contribution of these factors to the explanation of hospital differences in patient satisfaction and the mostly minor differences in hospital rankings after adjustment, many authors recommend adjusting for patient characteristics to avoid the possibility of bias and the concern that hospitals may have about the appearance of bias (Finkelstein, et al., 1998; Hargraves, et al., 2001; Rosenheck, et al., 1997; Young, et al., 2000).

An alternative suggestion, based on an analysis of CAHPS data, is to stratify reports of ratings of care for key variables, e.g., report ratings separately for those with excellent or very good self-rated health status and those with poor, fair, or good health status (Elliott, et al., 2001). This approach, although more costly, would address the problem of case-mix adjusting that might eliminate real differences in care among providers. While stratified reporting may be useful when the effects of case-mix adjustment vary across hospitals, it may be impractical when the subgroups are small in number and also may yield overly complex results for consumers (Zaslavsky, 2001). Alternative statistical approaches, such as indirect estimation and hierarchical modeling, have been suggested when sample sizes are small (Zaslavsky, 2001).

The findings for hospital characteristics generally have not been strong and have not been used for case-mix adjustment; these factors may account for differences in patient perceptions and, therefore, would not be subject to adjustment. The differences by type of service can be addressed by stratifying the reports of patient perceptions and presenting them separately for each service (Hargraves, et al., 2001). For example, obstetrical patients are more likely than others to be satisfied regardless of hospital, so that stratifying by service would account for differences in the proportion of obstetrical patients and comparisons can be made across hospitals within service type.

Minority Reporting

While there are reports of health status, access to care, treatment, and survival for different minority groups (for example, see: Bach, et al., 2002; Bradley, et al., 2002; Collins, et al., 2002; Lannin, et al., 1998; Schneider, et al., 2002), relatively few reports are available on patient satisfaction by minority groups. A survey of inpatient psychiatric care in Britain compared white British patients with ethnic minority patients on satisfaction with care using semi-structured interviews (Callan & Littlewood, 1998). Data from the 1996-97 Community Tracking Survey

(outpatient) assessed patient ratings of satisfaction with physician style and trust in the physician, and reported racial/ethnic differences in the scores (Doescher, et al., 2000). Patient assessments of primary care in one state were analyzed for differences by patient ethnicity (Taira, et al., 2001). Data from the CAHPS survey of patient satisfaction in health plans were analyzed and reported by different racial/ethnic groups (Morales, et al., 2001). In one study of hospital patient satisfaction, the Mexican-American patients were surveyed, but there was no comparison to non-minority patients (Hennessy & Friesen, 1994).

These studies are important as a context for understanding differences among minority groups and between minority and non-minority patients. However, they do not address the issues related to reporting satisfaction comparatively across hospitals. Most of these reports are at an aggregate level with large samples; even the CAHPS survey analyses do not report results for minority/non-minority members by health plan, in part because of small numbers at the plan level and because of potential inter-plan differences in minority/non-minority expectations for care. National attention is being given to these issues. A recent report by the Institute of Medicine (2002) sets forth the racial and ethnic disparities in health care and points to the need for data collection and reporting to monitor progress. The National Quality Forum (2002) recently addressed measurement issues and challenges in reporting healthcare quality for minority populations and recommended improved race and ethnicity data collection practices for quality measurement.

METHODS

The initial step in the review of the public reports was to identify any additional reports since the prior review (Department of Health of Rhode Island, 2000). This current review is limited to public reports of hospital patient satisfaction that present comparative data. A combination of web site searches, a literature search, and key informant telephone calls was used to determine the existence of public reports comparing patient satisfaction across hospitals. Reports were solicited from the source or downloaded from the web sites. If a methodological or technical report had been produced, it was also obtained whenever possible.

The next step was to compare and contrast these reports on several key dimensions relevant to data collection and reporting of patient satisfaction survey data. These dimensions were divided into four categories. The first, Survey Facts, includes: the locale of the survey, the sponsoring organization, the survey vendor, the number and timing of report cycles, the number of hospitals surveyed, the survey procedure (e.g., number of mailings), the response rate (i.e., mean and range), and information on the use of alternate language surveys for non-English-speaking populations. The second category, Sampling Procedures, includes: number of patients sampled, sampling criteria (i.e., inclusion and exclusion), and sampling stratification procedures. The third category, Computation of Scores, includes: calculation of hospital-level scores, and case-mix adjustment (includes the variables tested, whether adjustment was actually performed, the degree of change in hospital scores due to the adjustment, and the explanatory power of the regression model). The final category, Reporting of Scores, includes: the minimum number of survey returns to report the hospital's scores, the number of hospitals included in the report, the normative score used (e.g., national or state average), and the methodology and format for

comparative reporting of hospital scores. Information not available in the public reports was gathered from the companion technical reports, where available.

Finally, as part of this review process, telephone calls with key informants were used to clarify points and complete information that was not available in the public or technical reports. Contacts knowledgeable about the hospital reports in their areas were reached in Michigan, Colorado, Massachusetts, Ontario, and Buffalo. (See Appendix A: List of Contacts.) The main purpose of these calls was to fill in the gaps in information on methods of data collection and statistical analysis of scores. Examples of the data collection questions include: the percentage of hospitals in a given locale that participated in the survey, the minimum number of survey responses required for the hospital to be included in the public report, and the methods used to determine whether someone needed an alternative language survey. Examples of the statistical questions include: the methods for sample stratification, the amount of change in hospital scores after adjustment for case-mix differences among hospitals, and the statistical basis for classifying hospitals into performance categories (e.g., as average, above, or below a normative mean).

RESULTS

Four *a priori* criteria were established for determining which hospital reports to include in the statistical scan. Reports were required to: (1) measure patient satisfaction, (2) address hospital performance, (3) be publicly reported, and (4) provide comparative data. The following nine hospital reports met these four criteria: the Rhode Island Department of Health report (RI); the California Institute for Health Systems Performance/California HealthCare Foundation report (CA); the Massachusetts Health Quality Partnership report (MA); the Ontario Hospital Association report (ONT); four Hospital Profiling Project reports in Southeast Michigan (SEMI), Buffalo (BUF), Indianapolis (IND) and Cleveland (CLE); and the Niagara (Western New York) Health Quality Coalition report (WNY). (See Appendix B: List of Public Reports.) Of these nine reports, four were state/province-based (RI, CA, MA, ONT), two were regional (SEMI, WNY), and three were city-based (BUF, IND, CLE). In terms of sponsorship, four of the nine reports were sponsored by an employer (SEMI, BUF, IND, CLE), two by a coalition (MA, WNY), one by a hospital association (ONT), one by a health department (RI), and one by an independent public corporation/philanthropic organization (CA). In addition to producing reports for the general public, RI, CA, MA, and ONT produced companion technical reports available to the public on web sites. (See Appendix B.)

Results of this review have been summarized into four tables, corresponding to four broad categories of dimensions: Survey Facts (Table 1), Sampling Procedures (Table 2), Computation of Scores (Table 3), and Reporting of Scores (Table 4). (See Methods section for a description of the dimensions presented in each of these tables.) The tables are structured such that the nine reports are listed in the row headings and the data collection and statistical dimensions under comparison are listed in the column headings of the tables.

Survey Facts (Table 1)

Vendor. Seven of the nine locales worked with the Picker Institute of Boston, MA as their survey vendor, and used a version of the Picker patient satisfaction survey. RI and ONT selected Parkside Associates, based in Park Ridge, Illinois, which was acquired in December 2000 by Press Ganey Associates of South Bend, Indiana. (RI used two surveys from Parkside: the inpatient survey for the 11 general hospitals and the rehabilitation hospital, and the psychiatric survey for the adult psychiatric hospital.)

Number and timing of report cycles. SEMI and WNY have published four rounds of hospital patient satisfaction reports, with the most recent released in 2001 and 2002 respectively. BUF, IND, and CLE each have published three reports, with the most recent in 2001. ONT has published two reports, most recently in 2001. Three locales (MA, 1998; CA, 2001; and RI, 2001) have published their first round of patient satisfaction reports.

Number of hospitals surveyed. RI was able to achieve 100% participation from acute care and selected specialty hospitals (an adult psychiatric and a rehabilitation hospital) in the State because of its legislative mandate to produce a public report. Other locales, however, relied on voluntary participation of hospitals. Participation ranged from 9% of invited hospitals in IND to 95% of the acute care hospital systems in ONT and 100% of the larger hospitals in the Buffalo-Niagara region of WNY.

Survey procedure. All but RI used a 3-wave survey mailing strategy to maximize response rates. This process included an initial mailing of the survey, a reminder post-card mailing 1-2 weeks later, and a re-mailing of the survey 2-3 weeks following. RI opted for a 2-wave mailing procedure (initial mailing plus reminder postcard one week later) for discharges from the acute care and rehabilitation hospitals. Because of laws governing patient confidentiality, patients at the psychiatric facility in RI were handed a survey with a business reply envelope by a staff person prior to discharge, and no reminder postcards were sent.

Response rate. The average response rates across the locales were consistently in the 40% range, with ONT reporting 40% on the low end and WNY reporting 49% at the high end. No information was available in the reports of the Hospital Profiling Projects (SEMI, BUF, IND, and CLE) regarding response rates. Higher response rates in RI would require additional costs for a third wave mailing, including a copy of the questionnaire and a second cover letter, and/or a telephone follow-up for non-respondents after the scheduled mailings have been sent (Rea & Parker, 1997).

Alternate language survey. Four of the nine locales made alternate language surveys available for patients who preferred this option to completing the survey in English. RI communicated the availability of a Spanish survey in a bilingual English-Spanish cover letter that accompanied the initial survey mailing. The cover letter stated that the Spanish survey could be requested by calling a toll-free number. ONT developed a bilingual English-French survey in addition to the English one, and allowed hospitals to select either version depending on their patient characteristics. CA developed Spanish and Chinese surveys and offered hospitals three options for acquiring them: (1) provide instructions in the cover letter (i.e., call toll-free number) for how patients can obtain the alternate survey, (2) have hospitals pre-identify patients for

whom English is not their primary language, or (3) mail both English and alternate surveys. Most California hospitals declined use of an alternate language survey, although a few hospitals implemented the first option. In MA, Spanish, Russian, Khmer, and Portuguese surveys were made available to hospitals. Some hospitals chose to pre-identify patients, and other hospitals chose to have patients request an alternate language survey. The Hospital Profiling Project (SEMI, BUF, IND, and CLE) and WNY determined that there was no need for an alternate language survey.

Sampling Procedures (Table 2)

Sampling stratification. All but ONT used a stratified sampling design in which equal numbers of patients would be selected from each hospital service type (i.e., medical, surgical, obstetrical) being surveyed. RI's sampling plan also included patients from two specialty hospitals (i.e., rehabilitation and psychiatric). In ONT, although a simple random sample was used in most hospitals, some of the larger hospitals conducted a stratified sampling procedure (specifics not described).

Number of patients sampled. All locales randomly sampled patients to complete the survey. All but ONT designed the sample to select equal numbers of patients from each hospital or hospital service type. Locales using Picker as their vendor sampled 600 patients per hospital, or 200 per service type. For CA hospitals that had fewer than 2500 annual discharges, 300 per hospital were sampled; for the smallest hospitals with fewer than 300 eligible discharges, 100% were selected. RI sampled 25 patients per service type per hospital per week, equivalent to 325 patients per service type (or 975 patients per hospital for those hospitals with all three services represented in the survey and 650 patients per hospital for hospitals with only two services). For hospitals in RI with fewer than 25 patients per service per week, 100% of the patients per service type were selected. Most hospitals in ONT sampled 500 patients per hospital; however, some large hospitals with multiple programs or sites sampled significantly more than 500 patients per hospital.

Sampling criteria. All locales required an overnight stay in either a medical or surgical unit to be eligible for sampling, and all but ONT required the sampled patient to be an adult (age 18 or older). Only ONT *excluded* obstetrical patients, and only RI *included* patients from the psychiatric and rehabilitation hospitals. All locales except ONT and WNY specified inclusion of patients discharged to a personal residence, excluding those discharged to a nursing home.

Computation of Scores (Table 3)

Calculation of hospital-level scores. All locales used a stratified reporting approach, although the stratification variable differed. Only ONT calculated and reported scores separately for small, community, and teaching hospitals. All other locales calculated and reported scores separately by hospital service.

The seven locales that used Picker as their survey vendor computed hospital-level "problem scores" (BUF, CLE, IND, and SEMI) or their arithmetic inverse, "performance scores" (CA, MA, and WNY). A problem score is the weighted average of the percent of patients who

reported problems to the survey questions within a given domain of care (i.e., each of the subscales that summarizes different components of care). A performance score is the weighted average of the percent of patients who gave the "best possible" response to the questions within a given domain. This scoring system dichotomizes a 3-point scale into two categories: never a problem and always/sometimes a problem. The two locales that used Parkside/Press Ganey (RI and ONT) computed hospital-level scores based on the entire response scale. First, individual responses were transformed from an ordinal scale to an interval scale by assigning numerical values to categorical responses (Excellent=100, Good=75, Fair=50, Poor=25, Very Poor=0). Next, patient-level domain scores were computed by taking the mean of responses to questions in each domain. Finally, hospital-level domain scores were calculated by taking the mean of the patient-level domain scores.

Case-mix adjustment. All but RI and WNY chose to case-mix adjust their hospital-level scores for key patient characteristics. The common patient characteristics used by all seven locales were age, gender, and self-reported health status. Six of these seven locales (CA, MA, SEMI, BUF, IND, and CLE) adjusted for education as well. Additionally, ONT adjusted for whether the patient or a proxy completed the survey, and for the number of hospitalizations in the previous two years. Although RI tested for age, gender, insurance, length of stay, and self-reported health status, case-mix adjustment was thought to be unnecessary because of the small differences that were found. WNY decided against case-mix adjusting their hospital scores in order to have the scores represent the actual case mix of the hospitals.

The seven locales that did case-mix adjust used multiple regression as the method for adjusting, and, except for ONT, developed a separate adjustment model for each hospital-service (3) by survey-domain (7) combination -- for a total of 21 models. ONT produced a separate adjustment model for each of their 10 survey domains. Overall, patient characteristics accounted for a small portion of the variance in patient satisfaction, ranging from 1% to 7%. Further, in all cases where this information was stated in the reports, the degree of shift in hospital scores due to case-mix adjustment was minimal. For instance, CA and MA reported an average shift of less than one percentage point between unadjusted and adjusted hospital scores, with a maximum shift across all domains within any service of 5 (CA) or 3 (MA) percentage points.

Reporting of Scores (Table 4)

Minimum number of survey returns and number of hospitals reported. All locales except MA established a minimum number of hospital survey responses required to compute a hospital score; if responses fell below this number, the hospital would not be included in the report. The four Hospital Profiling Project reports as well as WNY required a 30% response rate from each hospital. Consequently, two of the 15 hospitals surveyed in SEMI and six of the 18 hospitals surveyed in BUF were not included in the report. RI required 40 surveys per service, and all hospitals met that minimum (with the exception of one hospital for the surgical area). ONT required 100 surveys per hospital and added two other requirements: completion of 50% of the 65 survey questions, and at least one question on five of the nine domains completed; nine ONT hospitals failed to meet this minimum and were not included in the report. CA required only two survey returns per hospital because their intent was to include as many hospitals as possible in

the report; all CA hospitals met this minimum. MA chose not to establish a minimum, although the representatives indicated they would recommend doing so.

Taking these requirements into account, CA reported on the largest number of hospitals (N=113), and ONT reported on the next largest number (N=86). All hospitals surveyed by MA (N=58), WNY (N=15), RI (N=13), CLE (N=7), and IND (N=2) were publicly reported. SEMI reported 13 hospitals and BUF reported 12.

Normative scores. All locales compared their hospital scores to a normative score. The "normative score" or "normative mean" is the criterion to which each hospital's score is compared, rather than comparing hospital scores directly to one another. The "normative score" refers to an average score based on a normative distribution that can be calculated from survey data. This term can be distinguished from the term "benchmark", which often connotes an external, evidence-based standard. Currently, there is no universally-accepted benchmark for patient satisfaction measurement. The seven locales that used Picker as their survey vendor used domain-specific national averages for each type of hospital service as their normative scores; the norms were calculated as the arithmetic mean of the domain scores by hospital service for the hospitals in Picker's national database from the prior two years. (CA included its hospital scores in the calculation of the national averages; other Picker locales excluded themselves.) In addition to the national average, CA and MA used state averages as a second normative score, calculated as the mean of the domain scores by hospital service for all patients (CA) or for all hospitals (MA) participating in the survey. CA used only the state normative mean, and not the national mean, for presenting hospital-specific comparisons in the general public report.

Of the two locales that contracted with Parkside, RI used national averages for their normative scores, calculated as the domain-specific mean scores for each hospital service, computed from Parkside's national database. ONT used Province averages for their normative scores, computed as the mean of the domain scores by hospital type (i.e., teaching, community, and small) for all patients responding to the survey. Although all locales compared their hospital scores to a normative score, only MA displayed the normative scores in the public report, and only CA displayed them in the technical report. Hospital scores were compared to the normative score for the purpose of computing comparative ratings.

Comparative reporting of hospital scores. None of the locales displayed actual hospital scores in the public reports, although CA and MA displayed them in technical reports and RI published data reports on the web site. In the public reports, eight locales assigned comparative ratings to hospital scores in relation to a normative mean -- essentially, assigning hospitals to a performance category; thus, only the comparative ratings were displayed in these public reports. The exception was MA, which used a graphic display of hospital scores in the public report and presented comparative ratings only in the technical report. All locales except ONT chose a three-level categorization scheme corresponding to three performance categories, e.g., below average, average, and above average. Although ONT used three levels in the first public report released in 1999, in the second report released in 2001, a five-level categorization scheme was used to represent below average, somewhat below average, average, somewhat above average, and above average.

The comparative ratings of hospital scores were visually displayed in the public reports as symbols and presented in tabular format. Stars were used as the symbol except by RI, which used diamonds as the symbol. Only MA used a graphic format in the public report to compare hospital scores and normative scores; hospital scores were plotted as solid circles (along with the 95% confidence intervals) on a horizontal axis demarcated by a numerical scale, with the normative scores displayed as vertical bars.

The method for assigning the comparative ratings to hospitals varied across the locales. CA and WNY computed a 95% confidence interval around each hospital's score and compared the confidence interval to the normative mean. If the 95% confidence interval did not overlap the normative mean, that hospital was flagged as above or below average, accordingly. (MA used the same approach for the technical report.) ONT followed the same methodology but also computed a wider 99.9% confidence interval around each hospital's score. The two confidence intervals were then compared to the normative mean to produce five performance categories: "above average" if the 99.9% confidence interval was fully above the normative mean; "somewhat above average" if the wider 99.9% confidence interval overlapped the normative mean but the narrower 95% confidence interval was fully above it; "below average" and "somewhat below average" if the 99.9% and 95% confidence intervals of a hospital's score, respectively, were fully below the normative mean *and* the hospital's score value was lower than the scores of all hospitals designated "average"; and "average" for all remaining hospitals.

The four locales of the Hospital Profiling Project (SEMI, BUF, IND, and CLE) assigned hospitals to three performance categories by comparing the hospital's observed score (i.e., without a confidence interval) to a tolerance region of 2 standard deviations around the normative mean. Thus, hospitals with observed scores falling within the 2-standard-deviation tolerance region around the normative mean were designated "average"; hospitals with observed scores falling outside the tolerance region were flagged "above average" or "below average", accordingly.

RI blended the two statistical methods of computing a tolerance region around the normative mean and computing a confidence interval around the hospital's observed score. First, a 1-standard-deviation tolerance region was computed around the normative mean. Hospitals with observed scores falling within this tolerance region were designated "average." Hospitals with scores outside this region were subjected to a second test by computing a 95% confidence interval around each hospital score. If the confidence interval overlapped the normative mean, that hospital was reassigned to the "average" category. However, if the confidence interval was fully below or above the normative mean, that hospital was flagged as "below average" or "above average", respectively. Thus, two steps were required to flag a hospital as "below average" or "above average": (1) its observed score fell outside the 1-standard-deviation tolerance region around the normative mean, and (2) the 95% confidence interval around the hospital's observed score was completely below or above the normative mean.

DISCUSSION

Consideration of the detailed information reviewed about the nine public reports of hospital patient satisfaction raises a number of issues that can inform decision making about the public

reporting of data in Rhode Island. To address those issues, this discussion focuses on advantages and disadvantages relevant to three topics: minority representation in surveying and reporting, case-mix adjustment of satisfaction scores, and statistical options for assigning hospital scores to performance categories.

Minority Representation

Surveying. While some locales have made their surveys available in different languages, there is no agreement on the best way to determine who needs an alternate language questionnaire. The argument for finding ways to identify preferred language and distribute alternate language questionnaires to minority patients selected for the survey is that a higher response rate can be expected if language needs are met. Current methods make it difficult to identify such patients with any certainty. For example, the cover letter in Spanish used in Rhode Island identified only 33 respondents in the Round 1 survey, perhaps because it required the respondents to make a phone call and give their addresses. Moreover, in most areas, including Rhode Island, hospitals do not have a consistent or reliable method for recording race/ethnicity or language needs. This issue is not limited to non-English languages, but may include low literacy level, compromised vision, or other impairments. In Rhode Island, five hospitals are using the current pilot phase of the Round 2 survey to test alternative methods of determining language needs. The intent is to code when Spanish is the primary language so that Spanish language questionnaires can be mailed to the patient initially, rather than requiring a call to request this survey. Once the pilot is complete, all hospitals have agreed to participate in a debriefing to review the success of the coding processes and to define a common process for coding language that can be used by all the hospitals.

Reporting. An advantage to developing reporting that shows the satisfaction of different racial/ethnic groups is that it can help identify disparities in perceptions of care and be used to design targeted quality improvement efforts. Disadvantages relate to the difficulties in obtaining large enough samples for adequate reporting of different minority groups. Moreover, some evidence suggests that it is not race or ethnicity that may be the important variable, but age, socioeconomic status, health status, and related variables (e.g., insurance status, education) that make a difference in one's expectations and perceptions of the hospital experience (Callan & Littlewood, 1998; Cleary, et al., 1991; Finkelstein, et al., 1998; Kane, et al., 1997; Young, et al., 2000). Exploratory work in Rhode Island suggests that a major obstacle to reporting satisfaction by minority status is obtaining large enough sample sizes to make valid statistical comparisons between minority and non-minority patients. The issues are in the process of being addressed, and no conclusions or recommendations have been determined. Additional work is needed to fully understand the possibilities for reporting hospital patient satisfaction for the minority population in Rhode Island. These issues are of concern nationally (National Quality Forum, 2002), providing a context for the experience in Rhode Island.

Case-Mix Adjustment

All locales acknowledged that adjustment made little difference in hospital patient satisfaction scores. Taking into account patient characteristics, such as age, gender, self-reported health

status, and education, only a small proportion of the variance in satisfaction scores was explained by these variables. When scores were case-mix adjusted for patient characteristics, there was very little shift in hospital ratings. The importance of reporting results separately by type of service (i.e., medical, surgical, obstetrical) was recognized, so that any differences in patient satisfaction would be apparent. Few studies addressed hospital characteristics such as teaching status and size (Finkelstein, et al., 1998; Young, et al., 2000).

An advantage of using case-mix adjustment is the view that it provides a level of comfort to the hospitals that their ratings are not the result of differences in their mix of patients (Hargraves, et al., 2001). Case-mix adjustment also has the potential to reduce response bias related to differences in patient populations by accounting for differential distributions of patients across hospitals. Several disadvantages of case-mix adjustment have been identified. For example, adjustment may suggest to the reader that it was done to mediate real differences, when there actually was little or no effect of the adjustment on scores. Further, case-mix effects may vary from one hospital to another (Zaslavsky, 2001), and adjustment may mask real population differences that need to be addressed and may even undermine the goals of customer feedback (Young, et al., 2000). Finally, the literature shows that, although some patient characteristics are statistically correlated with (i.e., related to) patient satisfaction, the amount of the variation in scores attributable to these characteristics is so small as not to be meaningful, suggesting that other unmeasured factors also should be considered and sources of patient assessments identified for quality improvement.

Publicly Reporting Hospital Scores

An important consideration in public reporting of hospital patient satisfaction is how to present hospital performance to the public. Questions can be raised about whether to use a normative score and what kind to use, whether to report actual scores or display them as symbols, and how to assign comparative ratings to hospitals based on their patient satisfaction scores. These questions, and the statistical issues they generate, must be considered because scores may need to be translated in ways that make them easier for the public to understand.

Using a normative score for comparison. All nine public reports on hospital patient satisfaction used normative scores for comparison, although the norms used varied from national to state/province averages, or a combination of the two. An advantage of using normative scores is that there is a “criterion” to which individual hospitals can be compared, rather than comparing hospitals only to one another. A national norm provides a more broad-based comparative value; on the other hand, state or province norms may be viewed as more relevant because they reflect performance in the same region or locale. For Rhode Island, a national norm (i.e., a national average) was used to provide a more stable comparison than one based on the small number of hospitals in the State. An important consideration is whether a national norm truly reflects similar populations or conditions as in the reporting area. Providing data on the characteristics of the hospitals that constitute the national norm may alleviate such concerns.

Scores versus symbols. Once the decision is made to compare the scores to a normative score, how to present that comparison must be decided. Two commonly used approaches are graphs and symbols. Graphs depict the scores being compared and may include the normative

score used for comparison. Symbols summarize the data, typically classify the hospital scores into categories in relationship to the normative score (e.g., 3 stars for above the norm, 2 stars for same as, 1 star for below), and typically display the data in tabular format. The difficulty of determining the most appropriate visual presentation is apparent in the extensive research on consumer responses to reports of the CAHPS survey results (Goldstein & Fyock, 2001; Harris-Kojetin, et al., 2001). For example, initially, health plan results were reported as symbols, in this case stars, using a 3-level rating; based on cognitive testing, the decision was to report the actual plan scores as bars and omit the stars from the Medicare Internet site and handbook (Goldstein & Fyock, 2001). Among the nine public versions of the hospital patient satisfaction reports reviewed, only MA used a graph format to display their data; the remaining locales displayed symbols.

Assigning comparative ratings. When a comparative rating method for the scores is used, a question to consider is how the categories are determined. Among the nine reports reviewed, three different options for classifying hospital scores were used. Each option is a variation based on two decision criteria: placing a "tolerance region" around the normative mean, and/or placing a confidence interval around each individual hospital's score. The tolerance region is intended to limit small, inconsequential differences that have no practical meaning, by setting a boundary around the normative mean to delineate the minimum distance a hospital score must be from the mean in order to be flagged as above or below. The confidence interval takes into account both the sampling variability around each hospital's score (i.e., the variation in responses to survey questions) and the size of the hospital sample (i.e., the number of patients who responded to the survey). The width of the confidence interval varies accordingly. The three options used are:

- Option 1 -- flags a hospital as above or below the normative mean if the hospital score is outside a "tolerance region" (e.g., 2 standard deviations) around the normative mean (used by Southeast Michigan, Buffalo, Indianapolis, and Cleveland).

- Option 2 -- flags a hospital as above or below the normative mean if the 95% CI confidence interval around the hospital score does not overlap the normative mean (used by California, Massachusetts, and Western New York; Ontario used a variant of this option).

- Option 3 -- flags a hospital as above or below the normative mean if the hospital score is outside a "tolerance region" (i.e., 1 standard deviation) around the normative mean *and* the 95% confidence interval around the hospital score does not overlap the normative mean (used by Rhode Island).

A fourth option proposed by a statistical consultant (Pezzulo, 2001) also uses both criteria and would flag a hospital as above or below the normative mean if the confidence interval (e.g., 70%) around the hospital score is outside the tolerance region (e.g., 1 standard deviation) around the normative mean (i.e., the confidence interval and the tolerance region do not overlap).

Statistical issues. The statistical issues surrounding these options are related to how accurately each option classifies hospitals as average, above, or below the norm. Two types of statistical "error" can occur during this classification process, referred to as Type-I and Type-II error. With a Type-I error, an average hospital is misclassified as above or below average; with a

Type-II error, an above- or below-average hospital is misclassified as average. Any classification scheme has the potential to commit a Type-I error, a Type-II error, or some "mix" of the two. Typically, the two types of potential error operate in opposing directions, such that as one type of error is decreased, the other type increases.

Because Option 1 ignores the sampling variability around each hospital's score, it is biased toward Type-I error, meaning that there is a greater chance for an average hospital to be incorrectly classified as above or below average than for an above or below average hospital to be misclassified as average. For example, a hospital's observed score may lie below the tolerance region, but its confidence interval may overlap the tolerance region. Thus, because sampling variability (i.e., the confidence interval) is not taken into account, this hospital would be classified as below average. For Option 2, the absence of a tolerance region around the normative mean allows hospital scores with only small differences from the normative mean to be flagged as above or below the mean; this option is also biased toward a Type-I error. In Option 3, the tolerance region around the normative mean and the confidence interval around the hospital score are applied independently, so that there is still a bias toward Type-I error. The potential for Type I bias appears to be greater with Option 2 than with Options 1 or 3 (Pezzulo, 2001). It should be noted that all hospitals in the Rhode Island public report have sample sizes large enough (i.e., 175-200) so that, with a 95% confidence interval, the relative mix of Type-I to Type-II error becomes more balanced and the bias toward Type-I error is reduced. In the fourth option, the tolerance region and confidence region are interrelated. For small confidence intervals (less than 40%), the ratio of Type-I to Type-II error is balanced; however, as the size of the confidence interval is increased for all sample sizes, the bias toward Type-II error increases, meaning that a non-average hospital might be classified as average.

The options described above each present a different relative mix of potential error and thus may be biased toward one type of error or another (i.e., tipping the balance toward either Type-I or Type-II error). There is no consensus on the ideal mix of the two types of error; it depends on what classification outcomes are perceived as acceptable. Furthermore, it is important to remember that both types of error operate in a symmetrical fashion. The same Type-I error can misclassify an average hospital as either above or below average. Similarly, the same Type-II error can misclassify both below average and above average hospitals as average. Additionally, the potential for classification error changes with different sample sizes. Decisions about the appropriate criteria to use in comparing hospitals to a national or regional norm are important because of this potential for statistical error in categorizing the hospitals. If Type-I errors are more likely, then hospitals may be flagged as different when their satisfaction scores are within an average range. If Type-II errors are more likely, then more hospitals will be classified as average, and there will be less discrimination (or variation) among the hospitals to help consumers see where there may be differences. It is this balance or mix of the potential for statistical error that can help guide these decisions.

Summary

Nine public reports were identified that compare hospitals on patient satisfaction. These reports present data for states (Rhode Island, California, Massachusetts, and Ontario), for regions (Southeast Michigan, Western New York), or for cities (Buffalo, Indianapolis, and Cleveland). Similarities and differences were found among the nine reports in four areas: survey methods, sampling procedures, computation of scores, and reporting of scores. Although the approaches for sampling patients were similar, the use of different vendors meant different questionnaires and scoring of responses. While all used a normative comparison for each hospital, the methods for assigning any hospital to a comparison category varied in important ways.

In Rhode Island, several statistical approaches have evolved to date and are being used for reporting hospital patient satisfaction, hospital clinical performance, and nursing home clinical measures. Although all include an external benchmark for comparison, there are varied decisions about setting tolerance limits around the benchmark and confidence intervals around each facility's score. Moreover, the specifics of these approaches -- the source of the benchmark, the size of the tolerance region around the normative mean, and the size of the confidence interval around each hospital's score -- vary for these measures and settings. This variation is attributable to many factors, and it may be necessitated by differences in the quality measures (e.g., clinical vs. satisfaction), in the data sources, and in the settings of care. As the Health Quality Performance Measurement and Reporting Program in Rhode Island continues to expand, the issues raised by these differences will be addressed to meet the goals of quality improvement and public accountability.

REFERENCES

Bach PB, Schrag D, Brawley OW, et al. Survival of Blacks and Whites After a Cancer Diagnosis. *Journal of the American Medical Association* 2002; 287(16): 2106-2113.

Bradley CJ, Given CW, & Roberts C. Race, Socioeconomic Status, and Breast Cancer Treatment and Survival. *Journal of the National Cancer Institute* 2002; 94(7): 490-496.

Callan A and Littlewood R. Patient Satisfaction: Ethnic Origin or Explanatory Model? *Journal of Social Psychiatry* 1998; 44(11): 1-11.

Cleary PD, Edgman-Levitan S, Roberts M, et al. Patients Evaluate Their Hospital Care: A National Survey. *Health Affairs* 1991 (Winter): 254-267.

Collins KS, Hughes DL, Doty MM, et al. Diverse Communities, Common Concerns: Assessing Health Care Quality for Minority Americans: Findings from The Commonwealth Fund 2001 Health Care Quality Survey. New York: The Commonwealth Fund, March 2002.

Department of Health of Rhode Island. Health Care Quality Performance Measurement. A Review of the Current State of Public Reporting on Health Care Quality Performance: States, Hospitals, and Coalitions. July 2000.

Doescher MP, Saver BG, Franks P, & Fiscella K. Racial and Ethnic Disparities in Perceptions of Physician Style and Trust. *Archives of Family Medicine* 2000; 9: 1156-1163.

Elliott MN, Swartz R, Adams J, et al. Case-Mix Adjustment of the National CAHPS® Benchmarking Data 1.0: A Violation of Model Assumptions? *HSR: Health Services Research* 2001; 36(3): 555-573.

Finkelstein BS, Singh J, Silvers JB, et al. Patient and Hospital Characteristics Associated with Patient Assessments of Hospital Obstetrical Care. *Medical Care* 1998; 36(8): AS68-AS78.

Goldstein E, & Fyock J. Reporting of CAHPS® Quality Information to Medicare Beneficiaries. *HSR: Health Services Research* 2001; 36(3): 477-488.

Hargraves JL, Wilson IB, Zaslavsky A, et al. Adjusting for Patient Characteristics When Analyzing Reports From Patients About Hospital Care. *Medical Care* 2001; 39(6): 635-641.

Harris-Kojetin LD, McCormack LA, Jael EF, et al. Creating More Effective Health Plan Quality Reports for Consumers: Lessons from a Synthesis of Qualitative Testing. *HSR: Health Services Research* 2001; 36(3): 447-476.

Hennessy LL & Friesen MA. Perception of Quality of Care in a Minority Population: A Pilot Study. *Journal of Nursing Care Quality* 1994; 8(2): 32-37.

Institute of Medicine. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. BD Smedley, AY Stith, and AR Nelson, Editors. Washington, D.C.: National Academy Press, 2002.

John J. Patient Satisfaction: The Impact of Past Experience. *JHCM: Journal of Health Care Marketing* 1992; 12(3): 56-64.

Kane RL, Maciejewski M & Finch M. The Relationship of Patient Satisfaction with Care and Clinical Outcomes. *Medical Care* 1997; 35(7): 714-730.

Lannin DR, Mathews HF, & Mitchell J. Influence of Socioeconomic and Cultural Factors on Racial Differences in Late-Stage Presentation of Breast Cancer. *Journal of the American Medical Association* 1998; 279(22): 1801-1807.

Marshall GN, Hays RD & Mazel R. Health Status and Satisfaction with Health Care: Results from the Medical Outcomes Study. *Journal of Consulting and Clinical Psychology* 1996; 64(2): 380-390.

Morales LS, Elliott MN, Weech-Maldonado R, et al. Differences in CAHPS® Adult Survey Reports and Ratings by Race and Ethnicity: An Analysis of the National CAHPS® Benchmarking Data 1.0. *HSR: Health Services Research* 2001; 36(3): 595-617.

National Quality Forum. Improving Healthcare Quality for Minority Patients: Workshop Proceedings. Washington, D.C.: National Forum for Health Care Quality Measurement and Reporting, 2002.

Pezzulo JC. Letter to Rhode Island Department of Health. Georgetown University, Washington, DC, October 3, 2001.

Rea LM & Parker RA. Designing and Conducting Survey Research: A Comprehensive Guide. San Francisco: Jossey-Bass Publishers, 1997.

Rogut L, Newman LS & Cleary PD. Variability in Patient Experiences at 15 New York City Hospitals. *Bulletin of the New York Academy of Medicine* 1996; 73(2): 314-334.

Rosenheck R, Wilson NJ & Meterko M. Influence of Patient and Hospital Factors on Consumer Satisfaction with Inpatient Mental Health Treatment. *Psychiatric Services* 1997; 48(12):1553-1561.

Schneider EC, Zaslavsky AM & Epstein AM. Racial Disparities in the Quality of Care for Enrollees in Medicare Managed Care. *Journal of the American Medical Association* 2002; 287(10): 1288-1294.

Taira DA, Safran DG, Seto TB, et al. Do Patient Assessments of Primary Care Differ by Patient Ethnicity? *HSR: Health Services Research* 2001; 36(6): 1059-1071.

Thi PL, Briancon S, Empereur F, & Guillemin F. Factors Determining Inpatient Satisfaction with Care. *Social Science and Medicine* 2002; 54(4): 493-504.

Young GJ, Meterko M & Desai KR. Patient Satisfaction with Hospital Care: Effect of Demographic and Institutional Characteristics. *Medical Care* 2000; 38(3): 325-334.

Zaslavsky AM. Statistical Issues in Reporting Quality Data: Small Samples and Casemix Variation. *International Journal for Quality in Health Care* 2001; 13(6): 481-488.

Table 1: Survey Facts

Locale	Sponsoring Organization	Survey Vendor	Number & Timing of Report Cycles	Number of Hospitals Surveyed	Survey Procedure	Response rate		Alternate Language Surveys
						Mean	Range	
Rhode Island (RI)	RI Dept of Health	Parkside/ Press Ganey	1st report (2001); Legislative mandate	11 general, 2 specialty (100% of general hospitals in RI)	2-wave mailing (for psych: one-wave hand-out)	43%	31-53%	Spanish survey available by request. Option offered in cover letter.
California (CA)	Calif. Institute for Health Systems Performance & Calif. HealthCare Foundation	Picker	1st report (2001)	113 (30% of eligible hospitals)	3-wave mailing	43%	20-62%	Spanish & Chinese surveys available. 3 options: pt requests from cover letter, hospital pre-identifies, or double mailing. Most hospitals declined options or chose pt request.
Massachusetts (MA)	Massachusetts Health Quality Partnership	Picker	1st report (1998); 2nd report (2000) not published due to problems.	58 (76% of acute care hospitals in MA)	3-wave mailing	47%	28-59%	Spanish, Russian, Khmer, & Portuguese surveys available. Some hospitals pre-identified patients; others provided option to request in cover letter.
Ontario (ONT)	Ontario Hospital Association	Parkside/ Press Ganey	1st report (1999); 2nd report (2001)	95 hospital systems (95% of eligible acute care hospital systems in ONT)	3-wave mailing	40%	18-59%	English and bilingual (English/ French) surveys available. Hospitals selected their preferred version.
SE Michigan (SEMI)	Hospital Profiling Project	Picker	4th annual report (2001)	15 (52% of invited hospitals in region)	3-wave mailing	No info	No info	Determined no need for alternate language survey.
Buffalo (BUF)			3rd annual report (2001)	18 (78% of invited hospitals in city)				
Indianapolis (IND)			3rd annual report (2001)	2 (9% of invited hospitals in city)				
Cleveland (CLE)			3rd annual report (2001)	7 (30% of invited hospitals in city)				
Buffalo-Niagara region (WNY)	Niagara Health Quality Coalition	Picker	4th report (2002)	15 (100% of larger hospitals in region)	3-wave mailing	49%	42-60%	Determined no need for alternate language survey.

Table 2: Sampling Procedures

Locale	Number of Patients Sampled	Sampling Criteria		Sampling Stratification
		Inclusion	Exclusion	
RI	25 patients per service type per hospital per wk (x 13 wks = 325 patients/service type/hospital). 100% of patients if < 25 patients/service type/wk	(1) medical, surgical, OB, rehab, or psych; (2) adult; (3) overnight stay; (4) discharged to home	(1) deceased; (2) received survey in previous 6 months; (3) transferred to another facility; (4) for psych: those unable to complete survey on own	Selected equal numbers from 3 services plus a rehabilitation and a psychiatric hospital
CA	600 patients/hospital; 300 patients/hospital if < 2,500 annual discharges; 100% of eligible patients if < 300 eligible discharges	(1) medical, surgical, or OB; (2) adult; (3) overnight stay; (4) discharged to home	(1) admitted for psych or substance abuse treatment; (2) admitted for observation; (3) patients who died or whose baby died	Selected equal numbers from 3 services
MA	600 patients/hospital	(1) medical, surgical, or OB; (2) adult; (3) overnight stay; (4) discharged to home	(1) admitted for psych or substance abuse treatment; (2) patients who died or whose baby died; (3) patients treated in multiple settings	Selected equal numbers from 3 services
ONT	Approximately half of participating hospitals surveyed 500 patients/hospital; the other half surveyed > 500 patients/hospital	(1) medical or surgical; (2) adult and pediatric; (3) overnight stay	(1) deceased; (2) outpatients; (3) psychiatry; (4) obstetrics	Some larger hospitals conducted stratified sampling procedure for surveying multiple programs or multiple sites, yielding larger sample sizes. Otherwise, simple random sample of 500/hospital.
SEMI BUF IND CLE	200 medical, 200 surgical, 200 OB patients per hospital	(1) medical, surgical, or OB; (2) adult; (3) overnight stay; (4) discharged to home	(1) deceased	Selected equal numbers from 3 services
WNY	600 patients/hospital	(1) medical, surgical, or OB; (2) adult	(1) deceased; (2) surveyed in the past 6 months	Selected equal numbers from 3 services

Table 3: Computation of Scores

Locale	Calculation of Hospital-Level Scores	Case-Mix Adjustment			
		Variables tested	Adjustment Actually Performed	Degree of Shift in Hospital Scores Due to Case-mix Adjustment	Explanatory Power of the Regression Models
RI	Responses transformed from ordinal to interval scale (Excellent=100, Good=75, Fair=50, Poor=25, Very Poor=0). Patient-level domain score = mean of responses to questions in domain. Hospital-level domain score = mean of patient-level domain scores.	(1) age, (2) gender, (3) insurance, (4) LOS, (5) self reported health status, (6) service type	Decided against case-mix adjustment.	N/A	N/A
CA	Hospital-level domain score = "Performance score" [1]	(1) age, (2) gender, (3) self-reported health status, (4) education, (5) additional variables which were not significant	Hospital score adjusted separately for each service-domain combination (n=21). Multiple regression. Age, gender, self-reported health, and education. Roll-up scores for medical, surgical, and maternity adjusted for service type.	Average shift < 1 percentage point. Maximum shift across all domains within any service = 5 percentage points.	Ranged from 1% to 7% across the 21 models
MA	Hospital-level domain score = "Performance score" [1]	(1) age, (2) gender, (3) self-reported health status, (4) education	Hospital score adjusted separately for each service-domain combination (n=21). Multiple regression. All tested variables.	Average shift < 1 percentage point. Maximum shift across all domains within any service = 3.4 percentage points.	Ranged from 1% to 5% across the 21 models
ONT	Responses transformed from ordinal to interval scale (Excellent=100, Good=75, Fair=50, Poor=25, Very Poor=0). Patient-level domain scores = mean of responses to questions in domain. Hospital-level domain scores = mean of case-mix-adjusted patient-level scores. The 10th indicator is a composite measure, formed by combining a weighted sum of 9 scales.	(1) age, (2) gender, (3) self-reported health status, (4) whether other filled out survey, (5) # hospitalizations in previous 2 years	Hospital score adjusted separately for each domain (n=10). Multiple regression. 6 domains adjusted by all 5 patient characteristics; remaining 4 domains adjusted by first four characteristics.	Average shift = 2.0 percentage points. Maximum domain shift = 3.0 percentage points; minimum domain shift = 1.2 percentage points.	Ranged from 0.7% to 3.4% across the 10 models
SEMI	Hospital-level domain score = "Problem score" [2]	(1) age, (2) gender, (3) self-reported health status, (4) education	Hospital score adjusted separately for. each service-domain combination (n=21). Multiple regression. All tested variables	Statistically insignificant shift	No information
BUF					
IND					
CLE					
WNY	Hospital-level domain score = "Performance score" [1]	N/A	Decided against case-mix adjustment.	N/A	N/A

[1] Performance score = the weighted average of the percent of patients who gave the best possible response to each of the questions within a given domain (inverse of "Problem Score")
[2] Problem score = the weighted average of the percent of patients who reported a problem for any question within a given domain.

Table 4: Reporting of Scores

Locale	Minimum # Survey Returns to Report Hospital Scores	Number of Hospitals in Report	Normative Score	Comparative Reporting of Hospital Scores
RI	40 surveys/service type	13	National average - from Parkside national database; domain- and hospital-service-specific	3 level rating scheme. 2 steps required to flag above or below norm: 1 standard deviation (SD) tolerance region computed around normative mean. Hospitals with scores within 1 SD are designated average. Hospitals with scores outside 1 SD are given 2nd test: 95% confidence interval (CI) computed around hospital score. If CI does not overlap normative mean, hospital is designated above or below average, accordingly; else, hospital is designated average. Reporting stratified by hospital service.
CA	2 surveys/hospital	113	(1) State average - mean performance scores of all patients; (2) National average - from Picker national database; domain- and hospital-service-specific (includes CA)	3 level rating scheme. 95% CI computed around hospital score; no tolerance region around normative mean. Hospitals with CI overlapping normative mean are designated average; hospitals with CI fully above or below norm are designated above or below average, respectively. Reporting stratified by hospital service.
MA	No minimum	58	(1) State average - mean scores of all participating hospitals; (2) National average - from Picker national database; domain- and hospital-service-specific	3 level rating scheme. 95% CI computed around hospital score; no tolerance region around normative mean. Hospitals with CI overlapping normative mean are designated average; hospitals with CI fully above or below norm are designated above or below average, respectively. Reporting stratified by hospital service.
ONT	(1) 100 surveys/ hospital; (2) 50% of the 65 questions and at least 1 question on 5 of the 9 indicators completed.	86	Province average - mean scores of all participating hospitals	5 level rating scheme. Hospitals with 99.9% CI fully above normative mean are designated above average; those with only 95% CI fully above norm are designated somewhat above average; hospitals with 95% and 99.9% CIs fully below norm AND with scores lower than scores of all average hospitals are assigned somewhat below and below average, respectively. Remaining hospitals are designated average. Reporting stratified by hospital type (i.e., teaching, community, small).
SEMI	30% response per hospital	13	National average - from Picker national database; domain- and hospital-service-specific	3 level rating scheme. 2 SD tolerance region computed around normative mean; no CIs computed around hospital scores. Hospitals with scores above or below 2 SD are designated. above or below average, respectively. Remaining hospitals are designated average. Reporting stratified by hospital service.
BUF		12		
IND		2		
CLE		7		
WNY	30% response per hospital	15	National average - from Picker national database; domain- and hospital-service-specific	3 level rating scheme. 95% CI computed around hospital score; no tolerance region around normative mean. Hospitals with CI overlapping normative mean are designated average; hospitals with CI fully above or below norm are designated above or below average, respectively. Reporting stratified by hospital service.

APPENDIX A: LIST OF CONTACTS

California Institute for Health Systems Performance, Marcia Nelson

Canada Institute for Health Information, Jennifer Zelmer

Canada Institute for Health Information, Jeremy Chrystman

Ford Motor Company, Diane Bechel

Health Action Council of NE Ohio, Pat Casey

Massachusetts Health Quality Partners, Inc., Barbra Rabson

Massachusetts Hospital Association, David Smith

Niagara Health Quality Coalition, Bruce Boissonnault

Niagara Health Quality Coalition, Joe Allen

APPENDIX B: LIST OF PUBLIC REPORTS

California Health Care Foundation and California Institute for Health Systems Performance (2001), *Results from the Patients' Evaluation of Performance (PEP-C) Survey: What Patients Think of California Hospitals.* (http://www.chcf.org/topics/view.cfm?itemID=12691).

Hospital Profiles, 2001, *Atlanta Hospital Profiles: A Consumer Guide.* (http://www.hospitalprofiles.org/pdf/atlanta.pdf).

Hospital Profiles, 2001, *Buffalo Hospital Profiles: A Consumer Guide.* (http://www.hospitalprofiles.org/pdf/buffalo.pdf).

Hospital Profiles, 2001, *Cleveland Hospital Profiles: A Consumer Guide.* (http://www.hospitalprofiles.org/pdf/cleveland.pdf).

Hospital Profiles, 2001, *Indianapolis Hospital Profiles: A Consumer Guide.* (http://www.hospitalprofiles.org/pdf/indy.pdf).

Hospital Profiles, 2001, *Southeast Michigan Hospital Profiles: A Consumer Guide.* (http://www.hospitalprofiles.org/pdf/somichigan.pdf).

Massachusetts Health Quality Partners, Inc., 1998, *Statewide Patient Survey Project.* (http://www.mhqp.org/statewidesurvey.html).

National Health Quality Coalition, 2002. Quality of Care as Reported by Hospital Patients. (http://www.myhealthfinder.com/).

Ontario Hospital Association, 2001, *Hospital Report 2001: Acute Care.* (http://secure.cihi.ca/hreports/toc.shtml).

Rhode Island Department of Health, 2001, *A Report of Patient Satisfaction with Hospital Care in Rhode Island.* (http://www.healthri.org/chic/performance/satisfaction.htm).