Performance Measures

Consumer Rankings and Health Care: Toward Validation and Transparency

Bala Hota, MD, MPH; Thomas A. Webb, BS; Brian D. Stein, MD, MS; Richa Gupta, MBBS, MHSA; David Ansell, MD, MPH; Omar Lateef, DO

The Institute of Medicine (now known as the Health and Medicine Division) has established the critical need to improve patient safety and quality,[1] and to achieve this aim, to use data to measure and improve health care through positive feedback and change.[2] Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, with a goal of improving health care quality through measurement and feedback, enabled the Centers for Medicare & Medicaid Services (CMS) to develop the Hospital Inpatient Quality Reporting program, and link payment rates to measures of quality.[3]

The Agency for Healthcare Research Quality (AHRQ; Rockville, Maryland) Patient Safety Indicators (PSIs) illustrate how measure usage can diffuse through stakeholder groups to become policy. In 2008, in an effort to enable Medicare to pay for quality over quantity, CMS added the PSIs as a measure of inpatient care quality following almost a decade of development and testing of the validity of these measures.[4] Each PSI has well-described criteria for counts of eligible patients (denominators) and cases (numerators), as well as exclusion criteria. Indicator measurement also depends on understanding when events in the hospital occur; adverse events, particularly those related to surgery, must clearly have followed procedures and not preceded them. Thus, it is critical to have accurate data on conditions that are present on admission (POA). Following the lead of CMS, many other private and commercial entities now rank hospitals using PSIs, either with identical or modified methodologies. Perhaps the most well-known use in the lay press of PSIs is in the *U.S. News & World Report* (*USNWR*) patient safety score, which uses seven PSI scores (Table 1, pages 440–441).

Consumer sites have been criticized for presenting too much information, with potential inaccuracies affecting the patient's (or consumer's) ability to choose.[5] Although some components of the *USNWR* ranking have been criticized as being too reliant on reputation,[6] the patient safety score represents a quantitative, objective component of the ranking. For a ranking measure to fulfill the aims of transparency, validity, and credibility,

Article-at-a-Glance

Background: Differences between the Centers for Medicare & Medicaid Services (CMS)–measured rates of safety events for Rush University Medical Center (RUMC; Chicago) and the *U.S. News & World Report* (*USNWR*)–determined patient safety score were evaluated in an attempt to validate the *USNWR* patient safety score–based ranking.

Methods: The *USNWR* findings for Patient Safety Indicators (PSIs) were compared with findings derived from RUMC internal billing data, and sensitivity analyses were conducted using a simulated data set derived from the Healthcare Cost and Utilization Project (HCUP) state inpatient data sets.

Results: Discrepancies were found for PSIs 3 (Pressure Ulcer Rate), 9 (Perioperative Hemorrhage or Hematoma Rate), and 11 (Postoperative Respiratory Failure Rate)—an excess of 0.72, 0.63, and 0.26 cases/1,000 admissions, in *USNWR* versus RUMC, respectively). The sensitivity analysis, which included missing present on admission (POA) flags and dates, resulted in an increase of rates by 1.83 (95% confidence interval [CI] = 1.10–2.56) cases/1,000 hospitalizations, 2.72 (CI = 0.00–5.90) cases/1,000 hospitalizations, and 3.89 (CI = 1.60–6.20) cases/1,000 hospitalizations for PSI 3, 9, and 11, respectively. Regression modeling showed that each 1% increase in transfers was associated with an increase of 0.06 cases of PSI 3/1,000 admissions; each 1,000 increase in admissions was associated with an increase of 0.04 cases of PSI 9/1,000 admissions.

Conclusion: The *USNWR* data set produced inaccurate PSI rates for RUMC, and false-positive event rates were more common among high-transfer and high-volume hospitals. More transparency and validation is needed for consumer-based benchmarking methods. In response to these findings and concerns raised by others, in 2016 *USNWR* made changes to its methodology and data sources and reported them in announcing its 2016–17 Best Hospitals.

Copyright 2016 The Joint Commission

Table 1. Patient Safety Indicators (PSIs) Used in the *U.S. News & World Report* Patient Safety Score*

Indicator	Indicator Description	Numerator	Denominator	Exclusions	Date-Sensitive Measure	Risk Adjustment Variables
PSI 3	Pressure Ulcer Rate	Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with any secondary diagnosis codes for pressure ulcer at appropriate stage	Surgical or medical discharges, for patients ages 18 years and older. Surgical and medical discharges are defined by specific DRG or MS-DRG codes.	Length of stay < 5 days; principal diagnosis of pressure ulcer; POA pressure ulcer diagnosis; presence of appropriate comorbidities, procedures, or transfer information	Yes—for length of stay and surgical cases	Age; DRG; specific MDC groups; comorbidities
PSI 4	Death Rate among Surgical Inpatients with Serious Treatable Conditions	Number of deaths among cases meeting the inclusion and exclusion rules for the denominator	Surgical discharges, for patients ages 18 through 89 years or MDC 14 (pregnancy, childbirth, and puerperium), with coded OR procedure, appropriate time to procedure, and appropriate inclusion and exclusion criteria for complications	Principal diagnosis of condition; related comorbidities; transfers	Yes—for principal procedures meeting criteria within 2 days of admission; length of stay criteria	Age; DRG groups for low mortality; MDCs; transfers; comorbidities
PSI 6	Iatrogenic Pneumothorax rate	Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with any secondary diagnosis codes for iatrogenic pneumothorax	Surgical and medical discharges, for patients ages 18 years and older. Surgical and medical discharges are defined by specific DRG or MS-DRG codes.	Principal diagnosis of pneumothorax; diagnoses of causative conditions; appropriate comorbidities	No	Sex; age; DRG; comorbidity; lack of procedure day
PSI 9	Perioperative Hemorrhage or Hematoma Rate	Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with secondary diagnosis codes for perioperative hematoma or hemorrhage or procedure codes for control of perioperative hemorrhage or hematoma	Surgical discharges, for patients ages 18 years and older, with any listed procedure codes for an OR procedure. Surgical discharges are defined by specific DRG or MS-DRG codes.	Principal diagnosis or POA code for hemorrhage/hematoma; procedures for control of hematoma/hemorrhage precede other OR procedures; appropriate comorbidities.	Yes—procedures for hematoma/hemorrhage control must follow OR procedure.	Sex; DRG; major diagnostic categories; transfers; comorbidities
PSI 11	Postoperative Respiratory Failure Rate	Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with secondary diagnosis code for acute respiratory failure; or mechanical ventilation after first major OR procedure code, for more than 96 hours; or smaller durations of ventilation occurring 2 or more days after a procedure; or reintubation one or more days after a procedure	Elective surgical discharges, for patients ages 18 years and older, with any listed procedure codes for an OR procedure. Elective surgical discharges are defined by specific DRG or MS-DRG codes with admission type recorded as elective.	Principal diagnosis code or secondary POA code for acute respiratory failure; only OR procedure is tracheostomy; tracheostomy before first OR procedure; appropriate comorbidities.	Yes—procedures indicative of numerator (e.g., tracheostomy) must follow first OR procedure.	Sex; age; DRG; major diagnostic categories; transfers; no procedure days; comorbidities; no point of origin

(continued on page 441)

Copyright 2016 The Joint Commission

Table 1. Patient Safety Indicators (PSIs) Used in the *U.S. News & World Report* Patient Safety Score* (continued)

Indicator	Indicator Description	Numerator	Denominator	Exclusions	Date-Sensitive Measure	Risk Adjustment Variables
PSI 14	Postoperative Wound Dehiscence Rate	Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with any listed ICD-9-CM procedure codes for reclosure of postoperative disruption of the abdominal wall	Discharges, for patients ages 18 years and older, with any listed procedure codes for abdominopelvic surgery	Abdominal wall reclosure occurs on or before day of first abdominal surgery; immunocompromised and other comorbidity diagnoses; length of stay < 2 days.	Yes–length of stay criteria; abdominal wall reclosure must follow first abdominal surgery.	Sex; age; DRG; major diagnostic categories; comorbidities
PSI 15	Accidental Puncture or Laceration Rate	Discharges, among cases meeting the inclusion and exclusion rules for the denominator, with any secondary diagnosis codes for accidental puncture or laceration during a procedure	Surgical and medical discharges, for patients ages 18 years and older. Surgical and medical discharges are defined by specific DRG or MS-DRG codes.	Principal or secondary POA diagnosis of accidental puncture or laceration during a procedure; spine surgery; pregnancy	No	Sex; age; DRG; major diagnostic categories; comorbidities

DRG, diagnosis-related group; MS, Medical-Severity; MDC, major diagnostic category; POA, present on admission; OR, operating room; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification.

* Agency for Healthcare Research and Quality. Patient Safety Indicators Technical Specifications Updates – Version 5.0, March 2015. Accessed Sep 2, 1016. http://www.qualityindicators.ahrq.gov/modules/PSI_TechSpec.aspx.

the integrity of the measure must be verifiable and reproducible.

Unfortunately, too little external validation of publicly reported consumer rankings occurs. Agreement between ranking systems often is poor,[6] with high performers in one system ranked lower in other systems. Without an understanding of the nuanced differences in data sources or methodologies, rankings become proprietary, and health care systems and consumers lose trust in measurement systems.

In an effort to evaluate a difference found between our CMS–measured rates of safety events at Rush University Medical Center (RUMC; Chicago) and the *USNWR*–determined patient safety score, our institution undertook a comparison of institutional data with the data used in *USNWR* patient safety score ranking and then determined the impact that the gaps could have on our national ranking. We specifically sought to understand discrepancies between measured quality as reported to CMS and *USNWR* ranking on the patient safety score. In addition, hypothesizing that hospital characteristics such as volume and transfer rates were associated with differences in performance characteristics of PSIs, we conducted a sensitivity analysis to assess the impact of these factors on quality indicator rates.

Methods

To evaluate the validity of the *USNWR* rankings on the patient safety score, we conducted two analyses. First, from September through December 2015, we compared the *USNWR* findings for PSIs with findings derived from RUMC billing data. Second, in May 2016, we conducted analyses using a data set created from the Healthcare Cost and Utilization Project (HCUP) state inpatient data sets (SIDS) in which we simulated the effect of missing data.

Comparison of *U.S. News & World Report* Results with RUMC Billing Data

To evaluate the RUMC ranking in the *USNWR* patient safety score, we obtained data from the *USNWR* analytic vendor. According to the vendor, these data were the publicly available federal fiscal years 2011–2013 MedPAR limited data set (mLDS),[7] which is a well-established data set used for research. The data set provided was restricted to our institution's records. We compared the data therein to our internal claims data for Medicare patients in the relevant time period. This was conducted with the knowledge and support of *USNWR* and was governed by a data use agreement that allowed aggregate analyses but not linkage to RUMC data sets or re-identification of individual records. To obtain the data set (which entailed a fee), we were referred by *USNWR* to the subcontractor tasked with providing the patient safety score for its annual rankings.

Descriptive statistics were obtained by examining the prevalence of admissions, charge codes, POA flag presence and absence, and total PSI events judged by the AHRQ Quality Indicators software (version 4.5a, AHRQ SAS QI software

Copyright 2016 The Joint Commission

[SAS; Cary, North Carolina]) for both the *USNWR* data set and RUMC data verified to have been sent for the time period to CMS. The overall rates of PSIs were calculated for RUMC and in the *USNWR* data set and qualitatively compared. The same denominator was used for RUMC and *USNWR* counts of events. Point estimates of rate differences were calculated by measuring the absolute difference between RUMC and *USNWR* rates.

Confidence intervals (CIs) and statistical significance for rate differences were calculated using Wald asymptotic confidence limits, which are based on the normal approximation to the binomial distribution and a two-sided Wald asymptotic test of equality for the risk differences. Risk differences and statistical tests were calculated using the PROC FREQ procedure in SAS and the RISKDIFF option.

Sensitivity Analysis

A sensitivity analysis was conducted using the AHRQ HCUP SIDS for New York and Florida for calendar year 2013. The HCUP data are a family of health care databases developed by AHRQ and have all-payer, encounter-level de-identified information. The data are in the appropriate format for use with the AHRQ Quality Indicators software. Data from New York and Florida were combined to generate a single data set of hospital discharges. These data were sampled to generate a representative sample of hospitals to be used in the analysis. Sampling was achieved via a stratified sample of 100 hospitals based on equal representation of hospitals with both high- and low-admission frequency and transfer rates. Hospitals were stratified as above or below median rates of transfer and admission. In addition, low-admission hospitals (< 365 admissions per year, or 1 admission per day) were a priori excluded.

Two analyses were conducted. First, an assessment of the change in PSI 9 (Perioperative Hemorrhage or Hematoma Rate) and PSI 11 (Postoperative Respiratory Failure Rate) was conducted with days after admission for procedures present (that is, similar to RUMC data) and absent (that is, the data from the mLDS). This analysis was achieved by using the AHRQ Quality Indicators software to measure PSIs 9 and 11—first, with procedure days present, and then with dates of procedures removed and a change of the associated AHRQ Quality Indicators software flag for the procedure days to absent. Second, a simulation was run to examine the impact of removing POA indicators on rates of PSI 3 (Pressure Ulcer Rate).

Because 10.1% of POA flags were inappropriately missing from the mLDS, we produced simulated data sets with 100 replications in which a random 10.1% of POA flags were deleted for each hospital in each iteration of the simulation. After deletion, PSI 3 measures were recalculated. The POA setting in the AHRQ Quality Indicators software was set to "present" in all analyses for PSI 3. The AHRQ Quality Indicators software has a setting to allow for data sets in which the POA flag is known to be absent. In those situations, the POA setting in the software is set to "absent." Because the mLDS is intended to have present POA flags, the setting of the software was set to "present" by the *USNWR* analytics vendor in its analyses. As such, for our analysis, the simulation PSI 3 results were used to generate a bootstrapped mean and 95% CI for absolute PSI 3 rates and the differences between simulated data sets with missing data and the original sample rate.

For both sensitivity analyses, statistical associations were examined between the absolute change in rate and hospital characteristics, including true rate of PSI, admission rate, and transfer rate. For the PSI 3 analysis, a mixed-effects, random-intercept model was used, with "hospital" treated as a random effect to account for the 100 within-hospital replications used to generate the simulation data set. The fixed effects included in the model were transfer rate, admission rate, and baseline PSI rate as independent variables. For PSI 9 and PSI 11, a linear regression model was used to model the absolute change in rate between date-present and date-absent data sets; because no replications were needed for this set, a random effect was not included. This model also included transfer rate, admission rate, and baseline PSI rate. Associations between dependent and independent variables were tested using maximum likelihood estimation and ordinary least squares regression for mixed (PSI 3) and linear (PSIs 9 and 11) models, respectively; final regression models included variables that were significant at the $p \leq 0.05$ level using two-sided tests. For PSI 3, models were built using the PROC MIXED procedure, and for PSI 9 and PSI 11, the PROC REG procedure.

All analyses were conducted using the AHRQ PSI Quality Indicators software, version 4.5a, and SAS v 9.3. This version of the software was also used by the analytic vendor for *USNWR* in its analyses to generate the patient safety score. The data used for this project were de-identified; as a result, the evaluation was exempt from review by the Institutional Review Board.

Results

U.S. News & World Report Data Set Analysis

Overall Comparison. PSIs with inclusion and exclusion criteria used by *USNWR*, with *USNWR* and RUMC rates based on billing data, are listed in Table 1 and Table 2 (page 443). In

Copyright 2016 The Joint Commission

Table 2. Patient Safety Indicator (PSI) Rates, *U.S. News & World Report* Results Compared with Medicare-Reported Results*

Indicator	*U.S. News* Count	Rush Count	Count Difference	*U.S. News* Rate	Rush Rate	Rate Difference, 95% CI	*P* Value
PSI 3	25	1	24	0.75	0.03	0.72 (0.71–0.72)	< 0.0001
PSI 4	24	29	-5	0.72	0.86	-0.14 (-0.16– -0.14)	< 0.0001
PSI 6	26	24	2	0.78	0.72	0.06 (0.05– 0.07)	< 0.0001
PSI 9	106	85	21	3.16	2.53	0.63 (0.56–0.69)	< 0.001
PSI 11	80	71	9	2.38	2.12	0.26 (0.21–0.33)	< 0.001
PSI 14	0	0	0	0	0	0	–
PSI 15	92	93	-1	2.74	2.77	-0.03 (-0.10–0.03)	0.39

Rush, Rush University Medical Center; CI, confidence interval.

* PSI descriptions can be found in Table 1 (pages 440–441).

four of seven PSI categories tested, the results from *USNWR* matched a reanalysis of internal claims data using the AHRQ Quality Indicators software within a range of -5 to +2 cases in absolute counts. In the remaining three PSI categories, *USNWR* data showed substantially greater PSI events than a reanalysis of internal claims data (Table 2).

Evaluation of the mLDS. The mLDS, on review, was found to have three main deficiencies. First, POA flags were missing from a subset of records. Of the 35,122 admissions in the data set, 3,538 records (10.1%) were missing a POA indicator for all diagnoses. As a comparison, in internal claims data for Medicare patients in the relevant time period, 0.3% of records had no POA recorded for the admission. Second, the dates of procedure codes were missing from the mLDS, as expected. Therefore, the data set did not have any indication for the dates when procedures occurred but had quarter and year present, consistent with a limited data set. Finally, truncation of data appeared to be present within the data set used by *USNWR* for an early time period. In the first quarter of federal fiscal year 2011, 4 fewer diagnoses on average were present in the mLDS than in RUMC billing data; for subsequent periods between the second quarter 2011 to the fourth quarter 2013, 0.5 to 1.5 fewer diagnoses on average were present.

Sensitivity Analysis

Comparison of HCUP Hospital Characteristics to RUMC. The data set used for the sensitivity analysis included 100 hospitals, with an average of 11,289 (95% CI = 9,064–13,515) admissions annually, an admission transfer rate of 12% (95% CI = 8%–17%), and a Medicare payer mix of 48% (95% CI = 45%–51%). In comparison, RUMC has approximately 34,000 admissions annually, with an 11% admission transfer rate and a 33% Medicare payer mix.

Data Quality and Changes in PSI Rates. As shown in Table 3 (below), missing date information resulted in false-positive event detection and elevated rates of PSIs 9 and 11 detected by the AHRQ Quality Indictors software. Missing POA information led to additional elevation in rates of PSI 3 and PSI 9 because of false-positive event detection. In regression models, higher transfer rates were significantly associated with higher rates of false-positive PSI 3 events in the setting of randomly missing POA flags, while high-admission hospitals had higher rates of false-positive PSI 9 events in the setting of missing dates of procedures (Table 4, page 444). Figures 1 and 2 (page 445) show the predicted rate increases due to increased transfers

Table 3. Results of Sensitivity Analyses Using AHRQ State Inpatient Database Sample*

Measure	PSI 3	PSI 9	PSI 11
HCUP Sample (Baseline)	0.74 (0.42–1.07)†	3.18 (2.45–3.91)	10.35 (7.96–12.75)
HCUP Sample (Dates removed)	0.75 (0.43–1.08)	4.24 (3.44–5.04)	14.22 (10.86–17.58)
HCUP Sample (POA removed)	2.58 (1.75–3.40)	5.90 (2.70–9.11)	14.25 (10.93–17.57)
Difference (Dates removed vs. baseline)	0.01 (0.00–0.02)	1.06 (0.68–1.44)	3.86 (1.57–6.17)
Difference (POA removed vs. baseline)	1.83 (1.10–2.56)	2.72 (0.00–5.90)	3.89 (1.60–6.20)

AHRQ, Agency for Healthcare Research and Quality; PSI, Patient Safety Indicator; HCUP, Healthcare Cost and Utilization Project; POA, present on admission.

* Indicator descriptions can be found in Table 1.

† Mean, confidence interval.

Copyright 2016 The Joint Commission

and admissions for PSI 3 and PSI 9, respectively.

Table 4. Regression Models of Hospital Characteristics for False-Positive Rate (Patient Safety Indicators [PSIs] 3, 9, and 11)

Measure	PSI 3	PSI 9	PSI 11
Intercept (in Multivariate Model)	1.15 (0.36–1.95)†	0.56 (0.03–1.08)†	–
Transfer Rate (% of Admissions)	0.06 (0.03–0.10)‡	-0.01 (-0.03–0.01)§	-0.08 (-0.22–0.05)§
Total Admissions (1,000's)	-0.04 (-0.11–0.02)§	0.04 (0.01–0.08)†	-0.07 (-0.27–0.14)§
Baseline PSI 3 Rate	0.28 (-0.17–0.73)§		
Baseline PSI 9 Rate		-0.04 (-0.14–0.07)§	
Baseline PSI 11 Rate			0.02 (-0.18–0.22)§

* Model developed on simulated data set with dates removed (PSI 9) and present on admission flag removed for 10.1% of records and dates removed for all records (PSI 3). PSI descriptions can be found in Table 1.
† $p < 0.05$, included in multivariate model.
‡ $p < 0.001$, included in multivariate model.
§ $p > 0.05$, not included in multivariate model.

Discussion

Using the source data for the *USNWR* patient safety score calculation, we validated the measures of PSI events used to calculate the RUMC patient safety score and compared this with RUMC billing data–based PSI calculations. We found that the rankings of safety published by *USNWR* were inaccurate because of several data quality issues. First, the data set being used to generate ranking of institutions by *USNWR* is a de-identified, research-related data set unsuitable for use to measure health care–associated events in an unbiased manner. The data set lacks the essential fields (dates of procedures) to enable measurement of time-dependent PSIs (PSIs 9 and 11), which led to increased rates of these PSIs with procedure days removed (Table 3). Second, for 10.1% of the records, the data set appeared to have missing values for indicators for events that were present on admission. For RUMC, these missing values yielded an excess of 0.72 cases of PSI 3/1,000 admissions, and in the sensitivity analysis, resulted in an even larger increase of 1.85 cases/1,000 admissions on average across the data set. Regression modeling showed that, for PSI 3, transfer rates were associated with increased false-positive detections; for each 1% increase in transfers, an increase in 0.06 cases/1,000 admissions detected. Similarly, for PSI 9, increased admissions resulted in more false-positive events; for each 1,000 increase in admissions, an increase in 0.04 cases/1,000 admissions was detected. In response to our findings and issues found by others, *USNWR* acknowledged in January 2016 the potential impact of missing POA information and absence of dates on their hospital rankings. Subsequently, *USNWR* announced changes to its methodology and data sources in June 2016[8] and then reported them in detail[9,10] in association with the announcement of the 2016–17 Best Hospitals.[11]

In the evaluation reported in this article, we were unable to determine whether missing POA flags in the mLDS occurred at random, was a finding unique to our institution or a finding across all institutions, or why these flags were missing. We speculate that the missing POA flags may have been present in the mLDS provided to *USNWR*'s analytics vendor at inception, although determination of the root cause may not be feasible. The consequence of the missing data in combination with the absence of dates, as demonstrated in the sensitivity analysis, is that for centers with large inpatient transfer populations, missing POA information will have a greater impact on false-positive detections of pressure ulcer than for hospitals without high transfer rates. Similarly, because of the absence of dates in the mLDS, increased admission rates would be more likely to result in an increase in false-positive detections of PSI 9 in high- rather than low-volume centers. On the basis of regression modeling, increases of 10% in transfer rates could increase PSI 3 rates by 1 case/1,000 admissions—a change that would be sufficient to move a hospital from the mean rate to the lowest 5th percentile of the distribution of hospitals (that is, an increase from 2.58 to 3.58 cases/1,000 admissions). We did not directly assess the impact of the truncation of diagnoses from the data set, which was limited to the oldest time period in the mLDS. Given the use of risk adjustment (Table 1), missing diagnoses could add additional increase to rates in complex centers, should missing diagnoses in analytic sets be present. Nevertheless, centers with larger transfer rates or higher admission rates could expect to see systematic bias in their *USNWR*–reported patient safety scores, and therefore, the impact of bias on scoring and ranking.

AHRQ PSIs have known flaws that threaten their validity. As a consequence, their utility for benchmarking must be approached with caution. A recent meta-analysis demonstrated that the positive predictive value of PSIs was poor when compared with chart review, reflecting coding errors, and that overall, data demonstrating the validity of the indicators are lacking.[12] Studies have also documented the impact of dates of service and POA indicators on validity of the PSI measures. For example, Rosen et al. found PSIs to have a positive predictive value of approximately 40%–50%, which improved to 75% when the presence of a condition at admission was known;

Copyright 2016 The Joint Commission

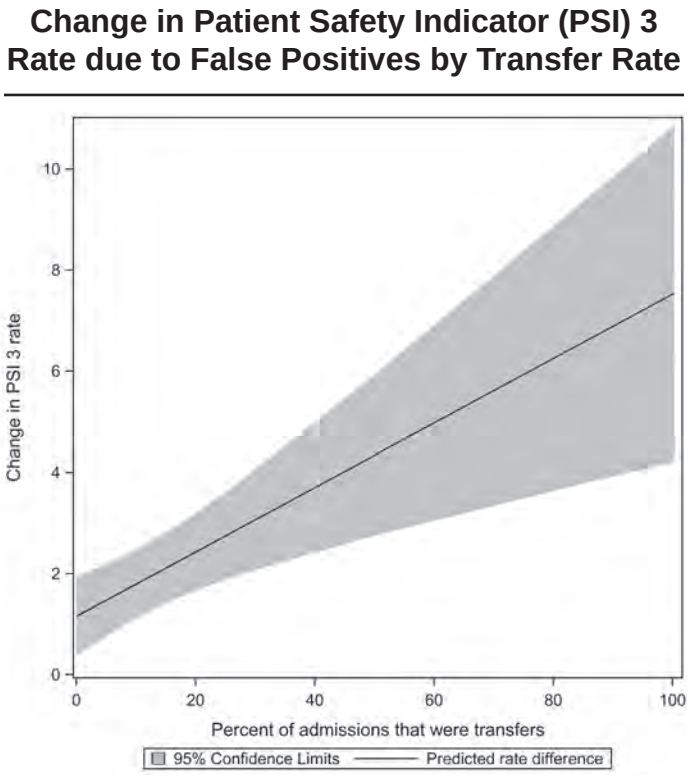

Figure 1. *The predicted rate increase for PSI 3 (Pressure Ulcer Rate) due to increased transfers is shown. The graph shows model-based estimates of the predicted rate difference and 95% confidence intervals.*

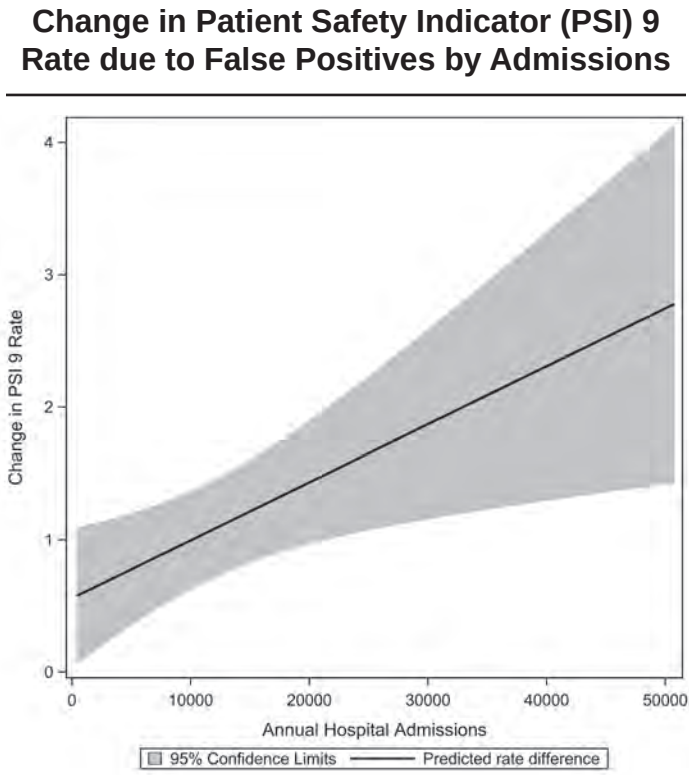

Figure 2. *The predicted rate increase for PSI 9 (Perioperative Hemorrhage or Hematoma Rate) due to increased admissions is shown. The graph shows model-based estimates of the predicted rate difference and 95% confidence intervals.*

absence of the POA flag falsely increased PSI detections by up to 50%.[4] Similar sensitivities and positive predictive values were found in previous studies for surgical and nonsurgical PSIs.[13–15] These limits on performance characteristics of the PSIs for detecting events jeopardize their validity even in settings in which data capture is complete.[16,17] AHRQ acknowledges these issues in its AHRQ Quality Indicators Toolkit and identifies strategies to reduce false positives through improved coding practice.[18] More recently, concerns about a lack of clinical relevance and bias regarding the use of the PSI 90 composite measure have been described.[19]

The *USNWR* rankings appear to be further hampered by the use of a data set in which missing data limit the meaning of AHRQ PSI results. Through the use of a large national data set that has been curated and made available by AHRQ, we have shown that the data gaps present in the *USNWR* data set produce biased estimates of PSI results, with greater false-positive detection rates for hospitals with higher admission and transfer admission rates. Our evaluation extends the existing literature by demonstrating not only that missing POA flags and procedure days yield inflated rates and poorer positive predictive value but that such missing data introduce bias by differentially worsening the PSI rates of high-volume/high-transfer hospitals as compared with low-volume/low-transfer centers.

Although consumer use of benchmarking data has increased over time, benchmarking publications have been found to be confusing for consumers.[5] We believe that a critical issue is the difficulty of independently validating benchmarking results. In a recent review of nine consumer ranking systems, transparency and reproducibility were noted to be uncommon among consumer systems; these systems are likely to confuse consumers and health systems, as deeper understanding of results is needed to explain meaning.[20] Obtaining data sets can be challenging, and revalidation complex. Moreover, bias based on data quality issues, as we have demonstrated, further threatens the trust and validity of the resulting published metrics.

The use of safety and quality measurement to rate hospitals and modify payment is an important incentive to develop and maintain high-quality care. The implementation of ranking and measurement by CMS through payments and public reporting can be an effective and transparent data-driven system to promote quality- and value-based care. Where caution must

Copyright 2016 The Joint Commission

be applied is in the downstream application of these measures, where transparency of methods may not exist and errors in analysis may be unchallenged. Consumer groups and lay publications that seek to measure and rank hospitals should be commended for the ambition to bring order to the confusing business space of health care, but the enormity of the task being undertaken by these entities should be acknowledged and the potential pitfalls of nontransparent data analysis recognized. For individual centers, there would seem to be a critical need for self-advocacy and communication to clarify the accuracy of consumer ranking systems. Trustworthy, transparent, nongovernmental health care rankings are possible, but they must be conducted in a manner in which access to data, nonproprietary methods, and ease of replication are facilitated. Until these principles can be widely adopted, we are likely entering an era in which the institutions subject to these rankings will be obliged to actively validate, through attempts to reproduce results, consumer rankings of importance to ensure that the respective methodologies are correct. J

The authors gratefully acknowledge the assistance and support of Robert Finke, Trustee, Rush University Medical Center, and the Quality of Care Committee, Rush University Medical Center.

Bala Hota, MD, MPH, is Associate Professor of Medicine and Chief Research Information Officer, Rush University Medical Center, Chicago. **Thomas A. Webb, BS,** is Manager of Clinical Resource Management, Rush University Medical Center. **Brian D. Stein, MD, MS,** is Assistant Professor of Medicine and Associate Chief Medical Officer, Rush University Medical Center. **Richa Gupta, MBBS, MHSA,** is Chief Quality Officer and Associate Vice President, Performance Improvement and Clinical Effectiveness, Rush University Medical Center. **David Ansell, MD, MPH,** is Senior Vice President for System Integration, Rush University Medical Center. **Omar Lateef, DO,** is Associate Professor of Medicine and Chief Medical Officer and Vice President, Rush University Medical Center. Please address correspondence to Bala Hota, Bala_Hota@rush.edu.

References

1. Institute of Medicine. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press, 2000.
2. Institute of Medicine. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. Washington, DC: National Academies Press, 2013.
3. Centers for Medicare & Medicaid Services. Reporting Hospital Quality Data for Annual Payment Update (RHQDAPU). Nov 2004. Accessed Sep 2, 2016. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/downloads/HospitalFactSheetAP.pdf.
4. Rosen AK, et al. Validating the patient safety indicators in the Veterans Health Administration: Do they accurately identify true safety events? *Med Care*. 2012;50:74–85.
5. PwC Health Research Institute. Scoring Healthcare: Navigating Customer Experience Ratings. Accessed Sep 2, 2016. http://www.pwc.com/us/en/health-industries/publications.html.
6. White C, Reschovsky JD, Bond AM. Understanding differences between high- and low-price hospitals: Implications for efforts to rein in costs. *Health Aff (Millwood)*. 2014;33:324–331.
7. Centers for Medicare & Medicaid Services. MedPAR Limited Data Set (LDS)—Hospital (National). (Updated: Jun 28, 2016.) Accessed Sep 2, 2016. https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/MEDPARLDSHospitalNational.html.
8. U.S. News & World Report L.P. Upcoming Changes to the *U.S. News* Patient Safety Score. Harder B, Comarow A. Jun 28, 2016. Accessed Sep 2, 2016. http://health.usnews.com/health-news/blogs/second-opinion/articles/2016-06-28/upcoming-changes-to-the-us-news-patient-safety-score.
9. U.S. News & World Report L.P. Methodology: *U.S. News & World Report* 2016-17 Best Hospitals Procedures & Conditions Ratings. Dougherty G, et al. Aug 2, 2016. Accessed Sep 2, 2016. http://static.usnews.com/documents/health/best-hospitals/BHPC_Methodology_2016-17.pdf.
10. U.S. News & World Report L.P. Methodology: *U.S. News & World Report* 2016-17 Best Hospitals: Specialty Rankings. Olmsted M, et al. Aug 9, 2016. Accessed Sep 2, 2016. http://static.usnews.com/documents/health/best-hospitals/BH_Methodology_2016-17.pdf.
11. U.S. News & World Report L.P. *U.S. News & World Report* Announces the 2016–17 Best Hospitals. Aug 2, 2016. Accessed Sep 2, 2016. http://www.usnews.com/info/blogs/press-room/articles/2016-08-02/us-news-announces-the-201617-best-hospitals.
12. Winters BD, et al. Validity of the Agency for Health Care Research and Quality Patient Safety Indicators and the Centers for Medicare and Medicaid hospital-acquired conditions: A systematic review and meta-analysis. *Med Care*. Epub 2016 Apr 25.
13. Kaafarani HM, et al. Validity of selected Patient Safety Indicators: Opportunities and concerns. *J Am Coll Surg*. 2011;212:924–934.
14. Ramanathan R, et al. Validity of Agency for Healthcare Research and Quality Patient Safety Indicators at an academic medical center. *Am Surg*. 2013;79:578–582.
15. Romano PS, et al. Validity of selected AHRQ Patient Safety Indicators based on VA National Surgical Quality Improvement Program data. *Health Serv Res*. 2009;44:182–204.
16. Rosen AK, et al. Using estimated true safety event rates versus flagged safety event rates: Does it change hospital profiling and payment? *Health Serv Res*. 2014;49:1426–1445.
17. Kubasiak JC, et al. Patient Safety Indicators for judging hospital performance: Still not ready for prime time. *Am J Med Qual*. Epub 2015 Dec 29.
18. Agency for Healthcare Research and Quality. AHRQ Quality Indicators Toolkit: Documentation and Coding for Patient Safety Indicators. Accessed Sep 2, 2016. http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/systems/hospital/qitoolkit/b4-documentationcoding.pdf.
19. Rajaram R, Barnard C, Bilimoria KY. Concerns about using the Patient Safety Indicator-90 composite in pay-for-performance programs. *JAMA*. 2015 Mar 3;313:897–898.
20. Hwang W, et al. Finding order in chaos: A review of hospital ratings. *Am J Med Qual*. 2016;31:147–155.

Copyright 2016 The Joint Commission