Using Harm-Based Weights for the AHRQ Patient Safety for Selected Indicators Composite (PSI-90): Does It Affect Assessment of Hospital Performance and Financial Penalties in Veterans Health Administration Hospitals?

Qi Chen, Amy K. Rosen, Ann Borzecki, and Michael Shwartz

Objective. To assess whether hospital profiles for public reporting and pay-for-performance, measured by the Agency for Healthcare Research and Quality (AHRQ) Patient Safety for Selected Indicators (PSI-90) composite measure, were affected by using the recently developed harm-based weights.

Data Sources/Study Setting. Retrospective analysis of 2012–2014 data from the Veterans Health Administration (VA).

Study Design. The AHRQ PSI software (v5.0) was applied to obtain the original volume-based PSI-90 scores for 132 acute-care hospitals. We constructed a modified PSI-90 using the harm-based weights developed by AHRQ. We compared hospital profiles for public reporting and pay-for-performance between these two PSI-90s and assessed patterns in these changes.

Principal Findings. The volume-based and the harm-based PSI-90s were strongly correlated ($r = 0.67$, $p < .0001$). The use of the harm-based PSI-90 had a relatively small impact on public reporting (i.e., 5 percent hospitals changed categorization), but it had a much larger impact on pay-for-performance (e.g., 15 percent of hospitals would have faced different financial penalties under the Medicare Hospital-Acquired Condition Reduction Program). Because of changes in weights of specific PSIs, hospital profile changes occurred systematically.

Conclusions. Use of the harm-based weights in PSI-90 has the potential to significantly change payments under pay-for-performance programs. Policy makers should carefully develop transition plans for guiding hospitals through changes in any quality metrics used for pay-for-performance.

Key Words. Patient safety, performance assessment, hospital profiling, composite measures

Public reporting and pay-for-performance programs are currently used by both federal and state agencies to increase transparency and improve health care quality and safety. Public reporting programs encourage transparency and benchmarking via open access to information, while pay-for-performance programs adjust financial reimbursements to hospitals based on a hospital's performance on a specific measure or set of measures (Centers for Medicare and Medicaid Services [CMS] 2016c, d, e). At the federal level, the Centers for Medicare and Medicaid Services (CMS) uses a variety of outcome measures for both public reporting and pay-for-performance. These measures often require routine updates and improvements to meet stakeholders' concerns, such as those related to the validity and reliability of the measures or that aim to better align measures with policy goals.

One of the measures currently used for both public reporting and pay-for-performance is the Agency for Healthcare Research and Quality's (AHRQ) Patient Safety for Selected Indicators (PSI-90); this measure also recently underwent revision in response to stakeholders' concerns (see below). PSI-90 is currently reported on the CMS Hospital Compare website for benchmarking purposes and is also used for pay-for-performance under the Hospital-Acquired Condition (HAC) Reduction Program (Centers for Medicare and Medicaid Services [CMS] 2016c) and the Hospital Value-Based Purchasing (HVBP) program (Centers for Medicare and Medicaid Services [CMS] 2016d) to financially penalize Medicare hospitals.

PSI-90 was designed to provide a simple and transparent single metric that could be used to better understand, communicate, and track patient safety across hospitals in the United States. It is comprised of selected component Patient safety indicators (PSIs), which are calculated using readily available, routinely collected administrative data (Agency for Healthcare Research and Quality [AHRQ] 2015b). Despite the current widespread use of PSI-90 by

Address correspondence to Qi Chen, Center for Healthcare Organization and Implementation Research (CHOIR), 150 S. Huntington Avenue, Boston, MA 02130; e-mail: qc2112@bu.edu or qi.chen2@va.gov. Amy K. Rosen, Ph.D., is with the Center for Healthcare Organization and Implementation Research (CHOIR), VA Boston Healthcare System, Boston, MA, and also with the Department of Surgery, Boston University School of Medicine, Boston, MA. Ann Borzecki, M.D., M.P.H., is with the Center for Healthcare Organization and Implementation Research (CHOIR), Bedford VA Medical Center, Bedford, MA; and also with the Department of Health Policy and Management, Boston University School of Public Health, Boston, MA, and Department of Internal Medicine, Boston University School of Medicine, Boston, MA. Michael Schwartz, Ph.D., is with the Center for Healthcare Organization and Implementation Research (CHOIR), VA Boston Healthcare System, Boston, MA, and also with the Department of Operations and Technology Management, Boston University School of Management, Boston, MA.

CMS and other agencies, it has been criticized for not reflecting the importance of individual PSIs in its weighting scheme, because it takes into account the volume of each PSI rather than its associated harm to patients (National Quality Forum [NQF] 2015b). For example, PSI #15 Accidental Puncture or Laceration is weighted heavily in PSI-90 because it occurs frequently; however, a minor puncture during an operation may result in relatively minor injury to the patient. Thus, weighting by volume may lead to misaligned quality improvement (QI) initiatives (i.e., those targeted toward frequently occurring PSIs rather than the most harmful PSIs). Another concern is that three surgical PSIs (PSI#9 Perioperative Hemorrhage or Hematoma, #10 Postoperative Physiologic and Metabolic Derangement, and #11 Postoperative Respiratory Failure), which may result in severe harm to patients, are not weighted in the composite. Finally, hospitals have expressed concerns that they may be penalized twice for their central venous catheter-related blood stream infections because this safety event is a component of PSI-90, as well as a separate measure in the HAC Reduction Program.

To respond to these concerns, AHRQ recently developed a new set of weights for PSI-90 (Romano 2015). These new weights incorporate not only the volume of each event but also each indicator's associated harm to the patient (e.g., risk of mortality, risk of readmission) and the disutility (i.e., severity) associated with each individual harm. Specifically, the new weights are based on three components: (i) excess harm associated with each individual PSI; (ii) the estimated preferences for health states reflected by these harms; and (iii) the volume of the PSI complication. Additionally, the three previously excluded surgical PSIs have been assigned weights based on their harm and volume. AHRQ also removed PSI #7 (Central Venous Catheter-Related Blood Stream Infection) from the composite. However, the new PSI software (version 6.0) provides the option to include PSI #7 into the calculation of PSI-90 if users want to include it. This new set of weights will be used in a revised version of PSI-90 ("new PSI-90"), which was endorsed by the National Quality Forum in December 2015 (National Quality Forum [NQF] 2015a), and will be released to the public shortly (AHRQ PSI Software, version 6.0) (Agency for Healthcare Research and Quality [AHRQ] 2016). The new PSI-90 is believed to more accurately reflect the impact of the PSI events; it is currently included in the Measures under Consideration list by CMS (Centers for Medicare and Medicaid Services [CMS] 2016a).

In this article, we examined the extent to which the use of the new harm-based weights versus the original volume-based weights in PSI-90 (herein "harm-based PSI-90" vs. "volume-based PSI-90," respectively) leads to

changes in hospital profiles and payments in the Veterans Health Administration (VA). Since the literature suggests that a hospital's score on the PSI-90 composite measure may be associated with specific hospital characteristics (Rivard et al. 2010; Shin et al. 2014), we also examined whether changes in hospital performance were associated with hospital characteristics such as bed size, number of discharges, and complexity level. In addition, because the harm-based weights were developed based on the potential harm associated with each PSI (e.g., mortality and readmission) at the patient level, we hypothesized that the harm-based PSI-90 would have a stronger correlation with hospital mortality and readmission rates than the volume-based PSI-90.

METHODS

Overview

This is a retrospective study using VA inpatient administrative data between January 1, 2012, and December 31, 2014, and VA 2013 Hospital Quality Report Card data (Veterans Health Administration 2013). We obtained approval from the Institutional Review Board at our facility.

Data Source and Variables

In the VA, the Inpatient Evaluation Center (IPEC) applies the PSI software (currently version 5.0) to administrative data collected from all hospitals to generate individual risk- and reliability-adjusted PSI rates and PSI-90 and reports them back to VA hospitals. IPEC also follows the CMS methodology in reporting mortality and readmission risk- and reliability-adjusted rates in the VA. For this study, IPEC provided us with risk-adjusted rates for 11 component PSIs (#3, 6-15), in-hospital mortality, 30-day mortality, and 30-day readmissions for 132 VA acute-care hospitals (Veterans Health Administration 2016).

We acquired hospital characteristics, including bed size and complexity level, from the VA Hospital Quality Report Card, which is a publicly available, comprehensive, facility-level report of quality and safety data (Veterans Health Administration 2013). Specifically, bed size refers to the total number of available beds at each hospital. Facility complexity level (i.e., Level 1 represents the highest complexity; Level 2, moderately complexity; and Level 3, the lowest complexity) is determined by the VA based on the characteristics of the patient population, the clinical services offered, the educational and

research mission, as well as the administrative complexity at each facility. We also calculated each hospital's number of discharges as the number of patients who were eligible for the 11 component PSIs during the study period.

Development of Harm-Based Weights

As noted above, AHRQ developed harm-based weights based on volume, harm, and disutility (where disutility equals one minus utility) (Romano 2015). Specifically, based on literature review and input from a clinical expert panel, each PSI was linked to potential downstream harms (e.g., mortality, readmission, dialysis, being on a ventilator, living in a skilled nursing facility) over a one-period year after the occurrence of a PSI event (i.e., post-PSI harm state). Next, statistical regression models were developed to estimate the marginal impact of each PSI based on the risk of these harms. Lastly, to determine disutilities, a clinical expert panel ranked these post-PSI harm states, a literature review was conducted to identify patient-reported utilities associated with individual harm states, and a linear regression model was estimated to calibrate clinician rankings to literature-based utilities. The final harm-based weight for each component PSI, q , was calculated as

$$weight_q = \frac{volume_q \sum_{h=1}^H harm_{qh} disutility_{qh}}{\sum_{a=1}^Q volume_a \sum_{h=1}^H harm_{ah} disutility_{ah}}$$

where Q is the total number of component quality indicators, q , in PSI-90; H is the total number of outcome types (harms), h , related to each component indicator; $volume$ is the number of total events within the component indicator in the reference population; $harm$ is the excess risk of each type of outcome (i.e., harm) within each component indicator estimated from a regression model comparing people with and without PSI events in an "at risk" cohort; and $disutility$ is the complement of a utility weight (1-utility_wt) assigned to each excess occurrence of each type of outcome within each component indicator.

The final harm-based weights that were endorsed by NQF are shown in Table 1. Of note, the weight for PSI #7 was forced to be 0 by AHRQ so that hospitals would not be double counted for their rates of central venous catheter-related blood stream infection (National Quality Forum (NQF) 2015a). In addition, the weights in the final PSI software version 6.0 may be slightly different from the weights endorsed by the National Quality Forum.

Table 1: The Volume-Based Versus Harm-Based Weights Used in the Patient Safety for Selected Indicators (PSI-90)

<i>Component Patient Safety Indicator (PSI)</i>	<i>Volume-Based Weights Used in the PSI Software, version 5.0</i>	<i>Harm-Based Weights Endorsed by the National Quality Forum*</i>
PSI #3 Pressure Ulcer Rate	0.033006	0.03633
PSI #6 Iatrogenic Pneumothorax Rate	0.075069	0.09736
PSI #7 Central Venous Catheter-Related Blood Stream Infection Rate	0.037684	0 [†]
PSI #8 Postoperative Hip Fracture Rate	0.001796	0.00879
PSI #9 Perioperative Hemorrhage or Hematoma Rate	0	0.15026
PSI #10 Postoperative Physiologic and Metabolic Derangement Rate	0	0.04915
PSI #11 Postoperative Respiratory Failure Rate	0	0.21544
PSI #12 Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate	0.3379	0.18429
PSI #13 Postoperative Sepsis Rate	0.057308	0.24132
PSI #14 Postoperative Wound Dehiscence Rate	0.018205	0.0089
PSI #15 Accidental Puncture or Laceration Rate	0.43903	0.00815

*The weights in the final PSI software version 6.0 may be slightly different from the weights endorsed by the National Quality Forum.

[†]Since central venous catheter-related blood stream infection rate itself is currently used for payment adjustment, PSI #7 is excluded from PSI-90 (i.e., weighted as 0) so that hospitals will not be penalized twice for this measure. However, the new PSI software (version 6.0) does provide the option to include PSI #7 into the calculation of PSI-90 if users want to include it.

Calculating Volume-Based and Harm-Based PSI-90s

PSI-90 is calculated using the weighted average of 11 component indicators (i.e., PSIs). Specifically, indirect risk adjustment is used to calculate an observed to expected ratio (O/E), where the expected is determined from a risk adjustment model with demographic and clinical variables. The ratio is then reliability-adjusted, resulting in shrinkage toward the overall mean.

As described earlier, we obtained hospital rates for each of the 11 component PSIs from IPEC. To calculate the volume-based PSI-90, we first applied the SAS program PSI_COMPOSITE included in the *PSI software* (version 5.0) (Agency for Healthcare Research and Quality [AHRQ] 2015b); this measure incorporates the original volume-based weights. We then replaced the volume-based weights by the harm-based weights in the

PSI_COMPOSITE program and reran the program to generate the harm-based PSI-90. The volume-based and harm-based weights are shown in Table 1.

Hospital Profiling and Payment

Similar to the approach used by the CMS Hospital Compare website to profile hospitals (Centers for Medicare and Medicaid Services [CMS] 2016e), we computed a 95 percent confidence interval (CI) for each hospital's PSI composite score and then assigned hospitals to performance categories based on a comparison of each hospital's 95 percent CI with the national VA PSI composite score. If the hospital's CI included the national VA composite score, then the hospital's performance was considered to be "no different from" the national average. However, if the hospital's CI was below or above the overall VA's composite score, the hospital was considered performing "better than" or "worse than" the national average, respectively.

We also examined the impact on payment under the HAC Reduction Program and HVBP program using each of the PSI-90 measures. Specifically, under the current HAC Reduction Program (Centers for Medicare and Medicaid Services [CMS] 2016c), hospitals in the worst performing quartile based on the total HAC score will have a 1 percent payment reduction applied to all Medicare discharges in FY16; PSI-90 contributes 25 percent to this score. Thus, consistent with this approach, we categorized hospital performance into quartiles based on the volume- and harm-based PSI-90s.

To estimate the potential financial impact of the PSI-90 change on hospitals, we assumed that all VA hospitals would receive payment under the CMS Acute Care Hospital Inpatient Prospective Payment System (IPPS). Thus, to simulate payment, we calculated the total hospital payment based on FY2016 IPPS Final Rule Regulations (Centers for Medicare and Medicaid Services [CMS] 2016b). Under the CMS IPPS, base operating and capital rates are adjusted by an area wage index to reflect the expected differences in local market prices for labor, which is intended to measure differences in hospital wage rates among labor markets by comparing the average hourly wage for hospital workers in each urban or statewide rural area to the nationwide average. In FY2016, the national operating labor share was 69.2 percent and 62 percent for hospitals with a wage index greater and less than 1, respectively. To simplify our simulation under the HAC Reduction Program, we did not consider the various wage index factors between hospitals that were used in the CMS calculation, but instead set the wage index as 1 for all VA hospitals. Thus, the

final payment for each hospital admission was calculated as Base Rate (\$5,466) $\times$ the Diagnosis Related Group relative weight. The total payment for each hospital was the sum of the payments for all admissions at this hospital in 1 year. The financial penalty for hospital i under the HAC Reduction Program, if any, was calculated as 1 percent $\times$ 25 percent $\times$ total hospital payment at hospital i .

For the HVBP program (Centers for Medicare and Medicaid Services [CMS] 2016d), PSI-90 is one of seven measures used in the outcome domain; this domain contributes 40 percent to the total performance score. Hospitals could receive up to a 1.75 percent payment adjustment based on two components of their performance in 2016: (i) achievement (i.e., compared to a national benchmark), and (ii) improvement in 2016 (i.e., compared to their own previous performance). Due to the lack of longitudinal data in this study and the complexity of combining the PSI component with other measures in the outcome domain, we used a simplified approach that is conceptually similar to that used by CMS, but which considered only the achievement part of the performance score. (We have used this approach in previous work; Rosen et al. 2013). Under the CMS method, M is defined as the median PSI-90 score, B as the benchmark score (mean of the top 10 percent of hospitals), and P as the performance of an individual hospital. A hospital's performance score is then calculated as a linear transformation of $(P-M)/(B-M)$, and payment is calculated as an additional linear transformation of the score combined with scores from other dimensions. We modified the CMS method and calculated a hospital's performance score as $(P - \max)/(\min - \max)$, where "max" is the highest (worst) PSI composite score across our hospitals and "min" the lowest (best). Instead of allocating financial penalties (which, given the complexity of the HVPB, is very difficult), we calculated the percentage of a hypothetical payment pool that would be awarded to each hospital based on its performance score divided by the sum of all of the performance scores of all hospitals.

Analyses

We first examined the correlation of the volume-based and harm-based PSI-90 and then the absolute change in hospital ranks when using the two versions of the composite. We then assessed changes in hospital profiles for public reporting and in payments under the HAC Reduction Program and HVBP program when using the volume-based PSI-90 versus the harm-based PSI-90. We also conducted univariate analyses to examine whether disagreements in

hospital rankings were associated with hospital bed size, volume, or complexity level. Finally, we assessed whether the harm-based PSI-90 was more strongly correlated with mortality and readmission rates than the volume-based PSI-90, and further examined this relationship in a multivariate regression model, adjusting for hospital characteristics. All analyses were performed using SAS software version 9.3 (SAS Institute Inc. Cary, NC, USA).

RESULTS

The volume-based and harm-based PSI-90s were strongly correlated ($r = 0.67, p < .0001$). The average absolute change in hospital ranks was 20.5 (range from 0 to 108). We found relatively small differences between the use of the volume-based versus the harm-based PSI-90s for public reporting (i.e., 6 of 132 [5 percent] hospitals changed categorization to “better than,” “worse than,” or “no different from” the national average). Specifically, of the four hospitals identified as performing “better than average” using the volume-based PSI-90, 3 (75 percent) were identified as “no different from the national average” of “average-performing” using the harm-based PSI-90. Of the 121 average-performing hospitals, 1 (1 percent) was reclassified to “worse than the national average”; among the 7 hospitals that performed worse than the national average, 2 (29 percent) improved and were reclassified as average-performing hospitals using the harm-based PSI-90 (see Table 2).

We found moderate agreement in the distribution of hospitals across quartiles (method used for the HAC Reduction Program) based on which PSI

Table 2: Changes in Hospital Profiles for Public Reporting Based on the Volume-Based Versus the Harm-Based Patient Safety for Selected Indicators (PSI-90)

<i>Hospital Profiles Based on the Volume-Based PSI-90</i>	<i>Hospital Profiles Based on the Harm-Based PSI-90</i>			
	<i>Better Than</i>	<i>Average-Performing</i>	<i>Worse Than</i>	<i>Total</i>
Better than*	1	3	0	4
Average-performing†	0	120	1	121
Worse than‡	0	2	5	7
Total	1	125	6	132

*Better than: upper 95% confidence interval (CI) of hospital PSI composite is lower than national average composite.

†Average-performing: 95% CI of hospital PSI composite overlaps with national composite.

‡Worse than: lower 95% CI of hospital PSI composite is higher than national average composite.

composite was used (i.e., kappa = 0.40 with 95 percent CI of [0.29, 0.52]); there was 55 percent agreement between measures on hospital quartile assignment. Specifically, of the 33 hospitals that would have received a financial penalty based on the volume-based PSI-90, 10 (30 percent) would not have been penalized under the HAC Reduction Program using the harm-based PSI-90, yielding an exemption from a total of \$444,000 in penalties among these hospitals (ranging from \$7,000 to \$126,000 per hospital). Conversely, among 99 hospitals that would not have received financial penalties based on the volume-based PSI-90, 10 (10 percent) would have been penalized using the harm-based PSI-90, yielding a total of \$636,000 in penalties (ranging from \$7,000 to \$163,000 per hospital) (see Table 3).

The percentage of the payment pool that each hospital would have received under the HVBP program using the volume-based versus the harm-based PSI-90 varied widely (Figure 1) For example, Hospital #33 would have changed from 0.9 percent using the volume-based PSI-90 to 0.2 percent of the

Table 3: Changes in Hospital Payment under the Hospital Acquired Condition Reduction Program Based on the Volume-Based Versus Harm-Based Patient Safety for Selected Indicators (PSI-90)

<i>Hospital Payment Based on Volume-Based PSI-90</i>	<i>Hospital Payment Based on Harm-Based PSI-90</i>				
	<i>Best</i>	<i>2nd</i>	<i>3rd</i>	<i>Worst*</i>	<i>Total</i>
Best	20	5	5	3 [†]	33
2nd	8	14	9	2 [‡]	33
3rd	3	9	16	5 [§]	33
Worst*	2 [¶]	5 ^{**}	3 ^{††}	23	33
Total	33	33	33	33	132

Note. Kappa = 0.40 with 95% CI of (0.29, 0.52).

*Hospitals in the worst performing quartile based on the total HAC score will have 1% payment reduction applied to all Medicare discharges in FY16; PSI-90 contributes 25% to this score.

[†]These three hospitals would have been subject to \$13,000, \$76,000, and \$76,000 penalties under the HAC Reduction Program, respectively, if the harm-based PSI-90 was used.

[‡]These two hospitals would have been subject to \$9,000 and \$78,000 penalties under HAC Reduction Program, respectively, if the harm-based PSI-90 was used.

[§]These five hospitals would have been subject to \$7,000, \$34,000, \$75,000, \$105,000, and \$163,000 penalties under HAC Reduction Program, respectively, if the harm-based PSI-90 was used.

[¶]These two hospitals would have avoided \$7,000 and \$15,000 penalties under HAC Reduction Program, respectively, if the harm-based PSI-90 was used.

^{**}These five hospitals would have avoided \$12,000, \$14,000, \$25,000, \$55,000, and \$66,000 penalties under HAC Reduction Program, respectively, if the harm-based PSI-90 was used.

^{††}These three hospitals would have avoided \$49,000, \$72,000, and \$126,000 penalties under HAC Reduction Program, respectively, if the harm-based PSI-90 was used.

Figure 1: Changes in Hospital Payment under the Hospital Value Base Purchasing Based on the Volume-Based Versus the Harm-Based Patient Safety for Selected Indicators (PSI-90)

Notes. Hospitals are arrayed on the horizontal axis from best performer (#1) to worst performer (#132) using the volume-based PSI-90. The vertical axis indicates the percent of the payment pool awarded to each hospital based on their achievement on the PSI-90. Examples of changes in hospital's payment: Hospital #33 received 0.9 percent of the payment pool based on its achievement on the volume-based PSI-90. The payment decreased to 0.2 percent of the payment pool based on the harm-based PSI-90. This was a 78 percent decrease, that is, $(0.9 - 0.2 \text{ percent}) / 0.9 \text{ percent}$. Hospital #83 received 0.7 percent of the payment pool based on its achievement on the volume-based PSI-90. The payment increased to 1.1 percent of the payment pool based on the harm-based PSI-90. This was a 57 percent increase, that is, $(1.1 - 0.7 \text{ percent}) / 0.7 \text{ percent}$.

payment pool using the harm-based PSI-90 (a decrease in 78 percent), while Hospital #83 would have increased from 0.7 percent to 1.1 percent of the payment pool (a 57 percent increase). Overall, 71 percent of hospitals would have faced changes of greater than 20 percent in HVBP payments, and 85 percent of hospitals would have faced changes of more than 10 percent under the HVBP program. The large fluctuations for some of the lower-performing hospitals may reflect the instability of percentage changes when the percentage of the payment pool based on the volume-based PSI-90 was very low.

Disagreement in rankings between the two PSI-90s was not associated with hospital bed size ($p = .32$), number of discharges ($p = .58$), or hospital complexity level ($p = .56$). Changes in hospital profiles occurred systematically due to the changes in the weights of specific PSIs. For example, two hospitals changed from the "best" quartile using the volume-based PSI-90 to the "worst" quartile based on the harm-based PSI-90 (see Table 2) because they had relatively high rates of PSI #9 Perioperative Hemorrhage or Hematoma, PSI#11 Postoperative Respiratory Failure, and PSI#13 Postoperative Sepsis. These three PSIs were weighted much higher in the harm-based PSI-90 than

Table 4: Correlations between Hospitals' Patient Safety for Selected Indicators (PSI-90) and Rates of Mortality and Readmissions

	<i>Volume-Based PSI-90 Score</i>		<i>Harm-Based PSI-90 Score</i>	
	<i>Pearson r</i>	<i>p-value</i>	<i>Pearson r</i>	<i>p-value</i>
In-hospital mortality rate	−0.052	.59	−0.03	.76
30-day mortality rate	−0.105	.25	−0.052	.56
30-day readmission rate	0.004	.64	0.04	.66

in the volume-based PSI-90 (i.e., 0.15 vs. 0; 0.22 vs. 0; and 0.24 vs. 0.06, respectively). On the other hand, three hospitals were re-categorized from the “worst” to the “best” quartiles (see Table 2) because they had very high rates of PSI#15 (Accidental Puncture or Laceration), which had a weight of 0.44 in the volume-based PSI-90 compared to 0.01 in the harm-based PSI-90.

We did not find a significant correlation between PSI-90 and the other outcome measures regardless of weighting method (see Table 4). Even after adjusting for hospital characteristics (i.e., bed size, number of discharges, and complexity level), we found no significant relationships between in-hospital mortality rate and the frequency-based and the harm-based PSI-90s ($p = .27$ and $.23$, respectively), between 30-day mortality rate and the two PSI-90s ($p = .41$ and $.45$, respectively), and between 30-day readmission rate and the two PSI-90 at the hospital level ($p = .68$ and $.74$, respectively).

DISCUSSION

An important finding from our study was that the type of weighting used to calculate PSI-90 significantly affected hospital profiles. Changes in profiles were not associated with hospital characteristics; instead, they were associated with differences in those component PSI rates that were weighted differently under the two approaches. Additionally, associations at the hospital level between the harm-based PSI-90 and the other outcome measures (i.e., mortality and readmission) were similar to those for the volume-based PSI-90.

CMS will shortly adopt the new PSI-90, which incorporates the harm-based weights, as well as several major revisions of specific component PSIs, for public reporting and pay-for-performance (National Quality Forum [NQF] 2016a). Although we were unable to follow AHRQ's exact methodology in calculating the new composite measure due to the unavailability of the new software, we were able to assess the impact of the new harm-based

weights on hospital profiles. Based on our findings, hospital profiles will likely be affected by this significant change; in fact, we expect that the impact will be greater in the pay-for-performance program than in the public reporting program. In our analysis, only 5 percent of the hospitals changed their classification for public reporting (i.e., “better than,” “no difference than,” and “worse than” average), while 14 percent of hospitals received a different financial penalty under HAC Reduction Program, yielding changes of more than \$1 million penalties. Furthermore, when we simulated a bonus payment program similar in spirit to the HVBP, 70 percent of the hospitals were awarded a difference of greater than 20 percent in their payment using the new harm-based PSI-90. These differing results between public reporting and pay-for-performance were due to the different methodologies used under each of these programs (CIs are used for public profiling while point estimates are used for pay-for-performance) (Agency for Healthcare Research and Quality [AHRQ] 2015b; National Quality Forum [NQF] 2015b; Centers for Medicare and Medicaid Services [CMS] 2016d).

The Measure Applications Partnership (MAP), a multistakeholder partnership that guides the U.S. Department of Health and Human Services on selection of performance measures for federal health programs (National Quality Forum [NQF] 2016b), supports the use of the new PSI-90 for public reporting and pay-for-performance. They recommended, though, that the new PSI-90 should be used for public reporting prior to including it in pay-for-performance programs since there is scant evidence as to how the recent revisions to PSI-90 will affect assessment of individual hospitals’ performance (National Quality Forum [NQF] 2016a). However, our findings suggest that this strategy, while theoretically sound, may not be that practical since switching from the volume-based weights to the harm-based weights in PSI-90 had a much larger impact on hospital payment in the pay-for-performance programs than on hospital classification in the public reporting programs. CMS should consider replicating our analyses to examine whether similar changes to hospital profiles occur in the private sector.

We recommend that CMS should carefully plan a “phase-in” period during the transition of PSI-90. First, CMS should develop educational materials explaining these potential changes to hospitals prior to implementing the new PSI-90 in any of their programs. For example, a comprehensive component could be added to the AHRQ Quality Indicators Toolkit (Agency for Healthcare Research and Quality [AHRQ] 2015a) for Hospitals to guide hospitals through the calculation of the harm-based PSI-90 and to help hospitals interpret the changes to their PSI-90 score. Since hospital differences in PSI-

90 performance using the harm-based PSI-90 versus the volume-based PSI-90 may reflect either between or within hospital differences in quality, or variation in coding or documentation with respect to individual PSI components, better coding and documentation should be encouraged to provide more accurate data for calculating hospitals' "true" performance on patient safety based on PSI-90. Second, our findings suggest that changes in hospital profiles were not associated with selected hospital characteristics, despite the previous literature in this area, nor did they occur randomly. Rather, these changes were due to the changes in the weights assigned to different PSIs. For example, hospitals that had relatively higher rates of PSIs #9 Perioperative Hemorrhage or Hematoma, #11 Postoperative Respiratory Failure, and/or #13 Postoperative Sepsis had a greater likelihood of being assessed as performing poorly by the harm-based PSI-90, while hospitals with relatively higher rates of PSI #15 Accidental Puncture or Laceration were more likely to be assessed as performing better. Therefore, CMS should also consider providing individual component PSI rates to hospitals, in addition to overall PSI-90 rates, as part of their educational materials to help hospitals better understand why their hospital profiles changed, as well as to facilitate more meaningful QI initiatives (e.g., hospitals could target those PSIs that are most harmful rather than those that are most frequent). Third, CMS should also consider blending in the new PSI-90 with the previous PSI-90 results for a period of time (e.g., 1 year) so that hospitals will not face a sudden increase or decrease in their PSI-90 scores. Based on the MAP suggestion mentioned above, CMS may wish to use this "blended" PSI-90 score for a longer period for pay-for-performance than for public reporting so that hospitals have sufficient time to understand the transition to the harm-based PSI-90 before they face its potentially different financial penalties. Lastly, specifically for public reporting, CMS might consider including PSI #7 (Central Venous Catheter-Related Blood Stream Infection Rate) as an option when calculating PSI-90, as the new PSI software (version 6.0) does provide this option. As mentioned previously, PSI #7 was removed from the PSI-90 to avoid double penalties in pay-for-performance. However, consideration should be given to including this PSI for hospital reporting purposes to help hospitals design QI plans.

As described above, the harm-based weights were designed using empirical evidence on the excess harm associated with each individual PSI and the estimated preferences for health states reflected by these harms. Although the literature suggests that the occurrence of PSI events may be associated with mortality and readmission at the individual case level (Rosen et al. 2013; Ramanathan et al. 2014; Ricciardi et al. 2016), we did not find that the harm-

based PSI-90 correlated better with mortality and readmission rates at the hospital level. One possible explanation is that aggregating data from the patient level to the hospital level may weaken associations (Robinson 2009; Te Grotenhuis, Eisinga, and Subramanian 2011). Future studies are needed to confirm this finding.

Our study has a few limitations. First, we did not construct the actual new PSI-90 composite measure due to the unavailability of the new software. Thus, our findings illustrate the impact of using the harm-based weights on hospital profiles, which may be different from the impact of using the new PSI-90 algorithm. Future studies should replicate our analyses with the actual new PSI-90 composite measure (which incorporates specific modifications of individual PSIs in addition to the new harm-based weights). Second, as previously noted, we did not have longitudinal data to assess the improvement score used by the HBVP. However, our study also has several strengths. This is, to our knowledge, the first study to examine the impact of adopting the harm-based PSI-90 on hospital profiles and to simulate financial penalties under HAC Reduction Program. Second, because we postulated that assessment of hospital performance would change as a result of revisions to the PSI-90 methodology, applying these new weights to a different database other than the one used for its development (the Healthcare Cost and Utilization Project) was a useful step in further evaluating the effect of the revisions of the measure.

CONCLUSION

The newly developed harm-based weights have better face validity than the current volume-based weights because they incorporate patient harm (a key element of patient safety). However, the use of the harm-based PSI-90 significantly affected particular hospital profiles (i.e., decreased performance for hospitals with relatively high rates of Perioperative Hemorrhage or Hematoma, Postoperative Respiratory Failure, and/or Postoperative Sepsis, and improved performance of hospitals with relatively high rates of Accidental Puncture or Laceration). In addition, the use of the new PSI-90 had a larger impact on CMS's pay-for-performance programs than on public reporting, due to methodological differences between the programs in how they profile facilities. Hospitals may need assistance in understanding why their performance profiles may change and the potential implications (e.g., financial penalties) that may be associated with these changes. Thus, policy makers

should carefully develop and disseminate plans when changing methodologies in the use of quality metrics to guide hospitals through these transitional periods.

ACKNOWLEDGEMENTS

Joint Acknowledgement/Disclosure Statement: This work was funded by the VA National Center for Patient Safety, Patient Safety Center of Inquiry on Measurement to Advance Patient Safety (MAPS) #XVA 68-023, PI Rosen.

Disclosures: This paper was accepted for a podium presentation by AcademyHealth Annual Meeting, June 2016.

Disclaimers: This paper does not represent the opinions of the VA. The views expressed are those of the authors, and no official endorsement by the VA is intended or should be inferred.

REFERENCES

- Agency for Healthcare Research and Quality (AHRQ). 2015a. "AHRQ Hospital QI Toolkit" [accessed on July 19, 2015]. Available at <http://www.qualityindicators.ahrq.gov/resources/Toolkits.aspx>
- Agency for Healthcare Research and Quality (AHRQ). 2015b. "Patient Safety for Selected Indicators" [accessed on April 22, 2015]. Available at http://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V50/TechSpecs/PSI_90_Patient_Safety_for_Selected_Indicators.pdf
- Agency for Healthcare Research and Quality (AHRQ). 2016. "AHRQ Quality Indicators Software v6.0" [accessed on July 18, 2016]. Available at <http://www.qualityindicators.ahrq.gov/software/Default.aspx>
- Centers for Medicare and Medicaid Services (CMS). 2016a. "CMS Measures Inventory" [accessed on April 22, 2016]. Available at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/CMS-Measures-Inventory.html>
- Centers for Medicare and Medicaid Services (CMS). 2016b. "FY 2016 IPPS Final Rule Hope Page" [accessed on July 19, 2016]. Available at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page.html>
- Centers for Medicare and Medicaid Services (CMS). 2016c. "Hospital-Acquired Condition Reduction Program" [accessed on April 22, 2016]. Available at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/HAC-Reduction-Program.html>
- Centers for Medicare and Medicaid Services (CMS). 2016d. "Hospital Value-Based Purchasing" [accessed on 2016]. Available at <https://www.cms.gov/Medicare/>

Quality-initiatives-patient-assessment-instruments/hospital-value-based-purchasing/index.html

Centers for Medicare and Medicaid Services (CMS). 2016e. "Medicare Hospital Compare" [accessed on April 22, 2016]. Available at <https://www.medicare.gov/hospitalcompare/search.html>

National Quality Forum (NQF). 2015a. "Measure 0531 Patient Safety for Selected Indicators (modified version of PSI90)" [accessed on April 22, 2015]. Available at <http://www.qualityforum.org/QPS/QPSTool.aspx#qpsPageState=%7B%22TabType%22%3A1,%22TabContentType%22%3A3,%22SearchCriteriaForStandard%22%3A%7B%22TaxonomyIDs%22%3A%5B%5D,%22SelectedTypeAheadFilterOption%22%3A%7B%22ID%22%3A0,%22FilterOptionLabel%22%3A%220531%22,%22TypeOfTypeAheadFilterOption%22%3A1,%22TaxonomyId%22%3A0,%22FilterOptionLabel%22%3A%220531%22%7D,%22Keyword%22%3A%220531%22,%22PageSize%22%3A%2225%22,%22OrderType%22%3A3,%22OrderBy%22%3A%22ASC%22,%22PageNo%22%3A1,%22IsExactMatch%22%3Afalse,%22QueryStringType%22%3A%22%22,%22ProjectActivityId%22%3A%220%22,%22FederalProgramYear%22%3A%220%22,%22FederalFiscalYear%22%3A%220%22,%22FilterTypes%22%3A0,%22EndorsementStatus%22%3A%22%22%7D,%22SearchCriteriaForPortfolio%22%3A%7B%22Tags%22%3A%5B%5D,%22FilterTypes%22%3A0,%22PageStartIndex%22%3A1,%22PageEndIndex%22%3A25,%22PageNumber%22%3Anull,%22PageSize%22%3A%2225%22,%22SortBy%22%3A%22Title%22,%22SortOrder%22%3A%22ASC%22,%22SearchTerm%22%3A%22%22%7D,%22ItemsToCompare%22%3A%5B%5D,%22SelectedStandardIdList%22%3A%5B%5D,%22StandardID%22%3A321,%22EntityTypeID%22%3A3%7D>

National Quality Forum (NQF). 2015b. "Voting Draft Report: NQF Endorsed Measures for Patient Safety" [accessed on April 22, 2015]. Available at https://www.qualityforum.org/Projects/n-r/Patient_Safety_Measures_2015/Voting_Memo.aspx

National Quality Forum (NQF). 2016a. "MAP 2016 Considerations for Implementing Measures in Federal Programs - Hospitals" [accessed on April 22, 2016]. Available at <https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=i81688d&ItemID=>

National Quality Forum (NQF). 2016b. "Measure Applications Partnership" [accessed on April 22, 2016]. Available at <http://www.qualityforum.org/map/>

Ramanathan, R., P. Leavell, L. G. Wolfe, and T. M. Duane. 2014. "Agency for Healthcare Research and Quality Patient Safety Indicators and Mortality in Surgical Patients." *American Journal of Surgery* 80 (8): 801–4.

Ricciardi, R., J. Nelson, T. D. Francone, P. L. Roberts, T. E. Read, J. F. Hall, D. J. Schoetz, and P. W. Marcello. 2016. "Do Patient Safety Indicators Explain Increased Weekend Mortality?" *Journal of Surgical Research* 200 (1): 164–70.

Rivard, P. E., A. Elixhauser, C. L. Christiansen, Z. Shibe, and A. K. Rosen. 2010. "Testing the Association between Patient Safety Indicators and Hospital Structural Characteristics in VA and Nonfederal Hospitals." *Medical Care Research and Review* 67 (3): 321–41.

- Robinson, W. S. 2009. "Ecological Correlations and the Behavior of Individuals." *International Journal of Epidemiology* 38 (2): 337–41.
- Romano, P. 2015. "Weighting Component Measures in AHRQ's PSI-90 Composite." AHRQ 2015 Research Conference. Washington, DC.
- Rosen, A. K., S. Loveland, M. Shin, M. Shwartz, A. Hanchate, Q. Chen, H. M. Kaafarani, and A. Borzecki. 2013. "Examining the Impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: The Case of Readmissions." *Medical Care* 51 (1): 37–44.
- Shin, M. H., J. L. Sullivan, A. K. Rosen, J. L. Solomon, E. J. Dunn, S. L. Shimada, J. Hayes, and P. E. Rivard. 2014. "Examining the Validity of AHRQ's Patient Safety Indicators (PSIs): Is Variation in PSI Composite Score Related to Hospital Organizational Factors?" *Medical Care Research and Review* 71 (6): 599–618.
- Te Grotenhuis, M., R. Eisinga, and S. V. Subramanian. 2011. "Robinson's Ecological Correlations and the Behavior of Individuals: Methodological Corrections." *International Journal of Epidemiology* 40 (4): 1123–5.
- Veterans Health Administration. 2013. "Hospital Report Card" [accessed on April 22, 2013]. Available at <http://www.va.gov/health/hospitalReportCard.asp>
- Veterans Health Administration. 2016. "Inpatient Evaluation Center" [accessed on April 22, 2016]. Available at <http://vawww.va.gov/ipec>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix