Are Patient Safety Indicators Related to Widely Used Measures of Hospital Quality?

Thomas Isaac, MD, MBA, MPH¹ and Ashish K. Jha, MD, MPH^{1,2,3}

¹VA Boston Healthcare System, Boston, MA, USA; ²Department of Health Policy and Management, Harvard School of Public Health, Boston, MA, USA; ³Division of General Medicine, Brigham and Women's Hospital, Boston, MA, USA.

CONTEXT: Patient safety indicators (PSIs) are screening tools that use administrative data to identify potential complications of care and are being increasingly used as measures of hospital safety. It is unknown whether PSIs are related to standard quality metrics.

OBJECTIVE: To examine the relationship between select PSIs and measures of hospital quality.

DESIGN, SETTING, AND PARTICIPANTS: We used the 2003 MedPAR dataset to examine the performance of 4,504 acute-care hospitals on four medical PSIs among Medicare enrollees.

MAIN OUTCOME MEASURES: We used bivariate and multivariate techniques to examine the relationship between PSI performance and quality scores from the Hospital Quality Alliance program, risk-adjusted mortality rates, and selection as a top hospital by US News & World Report.

RESULTS: We found inconsistent and usually poor associations among the PSIs and other hospital quality measures with the exception of "failure to rescue," which was consistently associated with better performance on all quality measures tested. For example, hospitals in the top quartile of failure to rescue performance had a 0.9% better summary performance score in acute myocardial infarction (AMI) processes and a 22% lower mortality rate in AMI compared to hospitals in the bottom quartile of failure to rescue ($p < 0.01$ for both comparisons). Death in low mortality DRG, decubitus ulcer, and infection due to medical care generally had poor or often inverse relationships with the other quality measures.

CONCLUSIONS: With the exception of failure to rescue, we found poor or inverse relationships between PSIs and other measures of healthcare quality. Whether the lack of relationship is due to the limitations of the PSIs is unknown, but suggests that PSIs need further validation before they are employed broadly.

KEY WORDS: patient safety indicators (PSIs); hospital quality scores; unsafe medical care.

J Gen Intern Med 23(9):1373-8

DOI: 10.1007/s11606-008-0665-2

© Society of General Internal Medicine 2008

INTRODUCTION

Unsafe medical care is a major cause of morbidity and mortality in US hospitals.^{1,2} Despite the interest in monitoring the frequency of adverse events, ongoing data collection efforts are time-consuming and expensive. In response, the Agency for Healthcare Research and Quality (AHRQ) created the Patient Safety Indicators (PSIs), tools that utilize billing information to screen for preventable complications of inpatient care. PSIs flag potential inpatient complications by using information from ICD-9 codes, demographics, length of stay, and other data in each discharge.³ Several studies have used the PSIs to identify risk factors and track trends of inpatient safety events on both local and national levels.⁴⁻⁶

Although the PSIs are not well-validated, they are increasingly being used to measure hospital safety. HealthGrades, an organization that publicly rates hospital quality, uses PSIs for its measurement of hospital safety.⁷ In addition, payers such as the Center for Medicare and Medicaid Services (CMS) are increasingly using PSIs to reward hospitals with lower complication rates in pay-for-performance initiatives.⁸ Although validation studies of PSIs are ongoing, the few studies that have examined whether PSI performance is related to other measures of hospital quality have shown inconsistent relationships.^{4,9,10}

Given the increasing demands to grade hospital performance, we sought to determine whether the medical PSIs were related to widely used measures of hospital quality: process measures of care used by the Hospital Quality Alliance (HQA); in-hospital risk-adjusted mortality for common medical conditions; and selection as "America's Best" by the US News rating system. This information can help payers, policy makers, consumers, and others who use these data for hospital comparisons understand whether these metrics identify similar top-performing institutions. If the PSIs are not related to other widely used metrics of hospital quality, it may suggest shortcomings in the measurement system or raise questions about how the dimensions of hospital quality and safety are related.

METHODS

Data. We used the 2003 MedPAR Part A 100 percent file, which contains discharge data for all hospitalizations of Medicare beneficiaries enrolled in the fee-for-service program. We linked the MedPAR discharge dataset to the 2003 annual survey of the American Hospital Association and limited our analyses to enrollees age 65 or older admitted to hospitals that care for

Received December 19, 2007

Revised March 21, 2008

Accepted April 29, 2008

Published online June 24, 2008

general medical or surgical patients. We used the risk-adjustment software developed by Elixhauser, available from the AHRQ, to define the presence or absence of 30 comorbidities in approximately 10.4 million discharges from 4,504 hospitals.¹¹ We then applied the PSI software, Version 3.0, to the dataset.

Although the AHRQ software calculates 20 different hospital-level PSIs, we excluded surgical or obstetric indicators and focused on the PSIs that relate to care for medical inpatients. The four PSIs examined were: death in low mortality DRG, decubitus ulcer, failure to rescue, and selected infection due to medical care. Death in low mortality DRG refers to in-hospital deaths per 1,000 patients in DRGs with less than 0.5% mortality; decubitus ulcer refers to cases that developed during hospitalization per 1,000 discharges with a length of stay of 5 or more days; failure to rescue refers to deaths per 1,000 patients that developed specified complications of care during hospitalization (such as pneumonia, sepsis, gastrointestinal bleed, etc.); selected infection due to medical care refers primarily to cases of infection due to intravenous lines and catheters. Further details about the PSI software, including specific patient exclusions made by each indicator, are available from AHRQ.³

We used the PSI software to calculate risk-adjusted PSI rates for each hospital. The risk-adjustment accounts for baseline differences in patient age, gender, modified DRG, and comorbidities among hospitals.³ The AHRQ software does not perform risk adjustment for death in low mortality DRG rates since all patients eligible for this event are considered low risk at baseline. We followed AHRQ guidelines and excluded hospitals with less than 30 cases in the denominator (patients at risk for a particular event) in order to create stable estimates.¹²

Relationship of PSIs and Process Metrics. We examined the relationship between hospital performance on the PSIs and its performance on HQA quality metrics collected between July 2005 through June 2006.¹³ Led by CMS, HQA is a collaboration of leading oversight organizations trying to improve the quality of hospital care by publicly reporting performance data.¹⁴ CMS collects process measures and imposes financial penalties for non-participation. We examined the ten core measures that apply to three medical conditions: acute myocardial infarction (AMI), congestive heart failure (CHF), and pneumonia. We created a summary performance score in each of the three clinical conditions using methodology endorsed by The Joint Commission.¹⁵ Summary scores were calculated if a hospital reported at least 30 discharges for any one of the indicators pertaining to the condition. The summary score is the sum of the numerators divided by the sum of the denominators of all the indicators in the clinical condition. We examined the relationships between each HQA summary score and each PSI.

Relationship of PSIs and In-Hospital Mortality. We examined whether performance on PSIs was related to in-hospital mortality rates in the six medical conditions identified by AHRQ as conditions where mortality rate is a marker of quality. We used limited license APR-DRG software to create modified DRGs and a severity index for each discharge in the MedPar dataset.¹⁶ We then calculated risk-adjusted in-hospital mortality rates using the AHRQ Inpatient Quality

Indicator Software, version 3.0, on the same MedPar dataset. We examined mortality for six conditions: AMI, CHF, pneumonia, gastrointestinal bleed, stroke, and hip fracture. Further information regarding ICD-9 codes, inclusion, and exclusion criteria used by the Inpatient Quality Indicators are available from the AHRQ.¹⁷ We excluded all hospitals with less than 30 patients at risk in order to have more stable mortality estimates. We examined mortality rates for AMI with and without transfers. The results were qualitatively similar, and the results with transfers are shown.

PSI Performance of US News Hospitals. We examined whether hospitals selected by US News & World Report in 2004 as a top-50 hospital in cardiac disease and cardiac surgery or as a top-50 hospital in respiratory disorders performed better on the PSIs. US News & World Report measures comprehensive quality by equally weighting three factors: reputation among experts, structural characteristics such as availability of health technologies, and risk-adjusted mortality rates.¹⁸ We chose to examine the top performing cardiac and pulmonary hospitals since these conditions represent the bulk of inpatient medical care. We examined the performance on each PSI among the combined group of 75 hospitals.

Statistical Analyses. We used Spearman correlations to examine the relationship between individual PSIs and HQA summary scores and then assigned hospitals into categories based on performance in each PSI. For two PSIs, death in low-mortality DRG and selected infection due to medical care, nearly half of the hospitals had zero events. For these two PSIs, we dichotomized hospitals into those with events and those without events. We tested for variance equality and used the appropriate t-tests to compare the HQA performance of hospitals with events against those without events. For failure to rescue and decubitus ulcer, we categorized hospitals into quartiles based on their complication rate and then calculated the average HQA summary scores within each quartile. We examined for linear trend in HQA performance across the PSI quartiles using analysis of variance.

We followed a similar approach when we examined the relationship among hospital PSI rates and mortality rates, beginning with Spearman correlations and then calculating mean mortality rates in each PSI performance category described above. We tested for variance equality and used appropriate t-tests to compare the mortality rates in the two categories for death in low mortality DRG and selected infection due to medical care. For failure to rescue and decubitus ulcer, we examined for linear trend in mortality rates across the PSI quartiles using analysis of variance. In order to facilitate interpretation, we calculated relative mortality rates by dividing the mortality rate in each category by the mortality rate in the worst PSI performance category.

For our analysis of US News Hospitals, we used a non-parametric (Wilcoxon) test to compare the median performance of US News hospitals to non-US News hospitals. We chose a non-parametric approach since the number of US News hospitals was relatively small and the PSI data were not normally distributed.

In secondary analyses, we used linear regression to examine the PSIs' relationships with HQA processes and mortality rates accounting for baseline differences in the following hospital

characteristics: bed size, teaching status (member of the Council of Teaching Hospitals or not), regional US location (Northeast, Midwest, South, West), urban or rural location, presence or absence of medical intensive care unit, presence or absence of cardiac care unit, percent Medicare patients, and percent Medicaid patients. We also examined the relationship between failure to rescue and in-hospital mortality by excluding deaths that were considered both failure to rescue cases and deaths in each of the six medical conditions. All analyses were conducted using SAS 9.1 (Carey, NC).

RESULTS

Of the 4,504 hospitals that cared for medical and surgical patients, there was a sufficient number of discharges to calculate at least one PSI in 4,414 hospitals (Fig. 1). The hospitals with insufficient discharges to calculate any PSIs were generally small, non-teaching hospitals in the western region of the US (data not shown). The rates of adverse events varied widely by indicator and across hospitals (Fig. 2).

Relationship of PSIs and Hospital Quality Alliance Metrics. We were able to examine the relationship between PSIs and HQA measures in 3,594 hospitals. We found a negative weak relationship between failure to rescue and the three HQA summary scores, indicating that better performance on this PSI was related to better HQA scores (correlation coefficients ranging between -0.06 to -0.08; $p < 0.01$ for all three comparisons). Correlations among the other PSIs and the three condition-specific summary scores were also weak but inconsistent in direction, ranging between -0.15 and 0.19. When we examined the relationship between individual PSIs in categories and HQA summary scores, we found similar patterns (Table 1): Better performance on failure to rescue was consistently associated with better performance on all three HQA summary process measures, although the magnitude of the difference was consistently small and of unclear clinical significance. For example, hospitals in the top quartile of failure to rescue scored 0.9% higher in overall AMI

Figure 1. Flow chart of selection of hospitals.

* Rates show number of events per 1,000 at risk discharges, $\pm$ interquartile ranges.

Figure 2. Median and interquartile ranges of medical PSI event rates.*

performance compared to the worst performing quartile (92.8% compared to 91.9%; p for trend=0.03). The relationship between other PSIs and HQA scores were inconsistent, often inverse of what might be expected, and of small clinical magnitude. After adjustment for hospital characteristics, the relationships between failure to rescue and the HQA measures were no longer significant, and many of the other relationships between the PSIs and the HQA processes were also attenuated (data not shown).

Relationship of PSIs and In-Hospital Mortality. Better failure to rescue performance was consistently associated with lower mortality for all six conditions, with Spearman correlation coefficients ranging between 0.20 and 0.38 ($p < 0.01$ for all relationships). The other PSIs had weak and inconsistent relationships with risk-adjusted mortality (coefficients ranging between -0.11 and 0.07). When we categorized PSIs into quartiles, we found a similar pattern: Hospitals with better failure to rescue rates had between a 22% and 31% lower relative mortality rate in all six conditions compared to hospitals with the worst failure to rescue rates (Tables 2 and 3; $p < 0.01$ for all comparisons). After excluding the deaths that were considered both failure to rescue cases and deaths in each of the six medical conditions, we found that hospitals in the top quartile of failure to rescue performance still had between an 18% and 26% lower rate of deaths ($p < 0.001$ for all comparisons). The relationships between other PSIs and mortality rates were inconsistent and often inverse. After adjusting for hospital characteristics, better performance on failure to rescue continued to be associated with lower mortality for all six conditions ($p < 0.001$ for all comparisons), although the magnitude of some of the relationships were somewhat attenuated.

PSI Performance of U.S. News Hospitals. Hospitals ranked as the nation's best hospitals in cardiac and pulmonary care had better performance on "failure to rescue" than non-US News hospitals (Table 4; median rate per 1,000 at risk discharges was 106.8 vs. 127.5 respectively; $p < 0.001$). However, these US News hospitals performed worse in all three of the other PSIs compared to non-US News hospitals ($p < 0.01$ for all comparisons).

Table 1. Relationship Between Risk-Adjusted PSI Performance and Mean HQA Summary Performance Scores

Risk-adjusted PSI category	HQA Mean Performance Scores					
	AMI		CHF		Pneumonia	
	Score	P-value*	Score	P-value*	Score	P-value*
Death in low mortality DRG [†]						
No events (top 60%)	91.3	<0.001	76.0	<0.001	84.1	0.13
Any events (bottom 40%)	92.9		80.6		84.5	
Decubitus ulcer						
Best performing quartile (top 25%)	92.0	0.25	76.7	0.12	85.6	<0.001
2nd quartile (26–50%)	92.5		78.1		84.7	
3rd quartile (51–75%)	92.8		79.0		84.1	
4th quartile (bottom 25%)	91.6		77.8		82.6	
Failure to rescue						
Best performing quartile (top 25%)	92.8	0.03	82.2	0.03	85.4	<0.001
2nd quartile (26–50%)	92.6		81.3		84.4	
3rd quartile (51–75%)	92.8		81.1		84.4	
4th quartile (bottom 25%)	91.9		80.7		83.8	
Infection due to medical care						
No events (top 46%)	88.2	<0.001	71.9	<0.001	83.8	0.01
Any events (bottom 54%)	92.7		80.7		84.5	

*P value is p for trend across categories for decubitus ulcer and failure to rescue. For death in low mortality DRG and infection due to medical care, p value is calculated using t-tests with equal variances comparing two categories

†Death in low mortality DRG is defined by observed rate, not risk-adjusted rate

DISCUSSION

We examined the relationship between hospital performance on all four medical PSIs and performance on well-known quality measures of inpatient medical care and found that the relationship varied depending on the PSI. Better performance on failure to rescue was consistently associated with better performance in process measures, risk-adjusted mortality rates, and being a US News hospital. However, the other three medical PSIs tested had inconsistent, weak, and often inverse relationships with HQA quality measures, risk-adjusted mortality rates, and selection as a top hospital by the US News service.

Our study supports existing work that has shown that PSIs are generally poorly related to other quality metrics.^{4,9,10} Although our study was not designed to validate whether PSIs measure hospital safety or accurately capture adverse events, our findings heighten existing concerns regarding the use of PSIs as a safety measurement tool. A critical review of the use of administrative data for public reporting previously concluded that it is uncertain whether differences in PSI rates among hospitals reflect actual lower complication rates in hospitals.¹⁹ The AHRQ and others have also consistently advised caution in using the PSIs for public quality reporting and payment.^{12,19} Policy-makers, payers, and rating agencies should take pause

Table 2. Relationship Between Risk-Adjusted PSI Performance and Risk-Adjusted Mortality Rates

Risk-adjusted PSI category	Risk-Adjusted Mortality Rates					
	AMI		CHF		Pneumonia	
	Rate Ratio*	P-value [†]	Rate Ratio*	P-value [†]	Rate Ratio*	P-value [†]
Death in low mortality DRG [‡]						
No events (top 60%)	1.00	0.92	1.04	0.06	1.02	0.19
Any events (bottom 40%)	1.00 (ref)		1.00 (ref)		1.00 (ref)	
Decubitus ulcer						
Best performing quartile (top 25%)	1.02	0.67	1.10	<0.01	0.99	0.33
2nd quartile (26–50%)	0.98		1.08		1.00	
3rd quartile (51–75%)	0.99		0.98		0.97	
4th quartile (bottom 25%)	1.00 (ref)		1.00 (ref)		1.00 (ref)	
Failure to rescue						
Best performing quartile (top 25%)	0.78	<0.001	0.70	<0.001	0.69	<0.001
2nd quartile (26–50%)	0.86		0.81		0.79	
3rd quartile (51–75%)	0.92		0.87		0.86	
4th quartile (bottom 25%)	1.00 (ref)		1.00 (ref)		1.00 (ref)	
Infection due to medical care						
No events (top 46%)	1.09	0.007	1.15	<0.001	1.05	0.005
Any events (bottom 54%)	1.00 (ref)		1.00 (ref)		1.00 (ref)	

*Mortality rate ratio calculated by dividing mortality rate of worst PSI performance category by mortality in other PSI performance categories

†P value is p for trend across categories for decubitus ulcer and failure to rescue. For death in low mortality DRG and infection due to medical care, p value is calculated using T tests with equal variances comparing two categories

‡Death in low mortality DRG is defined by observed rate, not risk-adjusted rate

Table 3. Relationship Between Risk-Adjusted PSI Performance and Risk-Adjusted Mortality Rates

Risk-adjusted PSI category	Risk-adjusted Mortality Rates					
	Stroke		GI Hemorrhage		Hip Fracture	
	Rate Ratio*	P-value†	Rate Ratio*	P-value†	Rate Ratio*	P-value†
Death in low mortality DRG‡						
No events (top 60%)	1.05	0.004	1.03	0.31	1.05	0.12
Any events (bottom 40%)	1.00 (ref)		1.00 (ref)		1.00 (ref)	
Decubitus ulcer						
Best performing quartile (top 25%)	1.11	<0.001	0.96	0.17	1.09	0.03
2nd quartile (26–50%)	1.05		0.94		1.04	
3rd quartile (51–75%)	1.00		0.96		0.99	
4th quartile (bottom 25%)	1.00 (ref)		1.00 (ref)		1.00 (ref)	
Failure to rescue						
Best performing quartile (top 25%)	0.75	<0.001	0.75	<0.001	0.75	<0.001
2nd quartile (26–50%)	0.87		0.79		0.82	
3rd quartile (51–75%)	0.90		0.91		0.90	
4th quartile (bottom 25%)	1.00 (ref)		1.00 (ref)		1.00 (ref)	
Infection due to medical care						
No events (top 46%)	1.16	<0.001	1.05	0.30	1.24	<0.001
Any events (bottom 54%)	1.00 (ref)		1.00 (ref)		1.00 (ref)	

*Mortality rate ratio calculated by dividing mortality rate of worst PSI performance category by mortality in other PSI performance categories

†P value is p for trend across categories for decubitus ulcer and failure to rescue. For death in low mortality DRG and infection due to medical care, p value is calculated using t tests with equal variances comparing two categories

‡Death in low mortality DRG is defined by observed rate, not risk-adjusted rate

before extending the use of PSIs as a grading or payment tool until this tool has been better validated.

Previous work examining whether claims data identify safety lapses also raised concerns: One study found that only 27% of medical complications flagged by administrative data were actually adverse events, and only 16% of these had associated deficiencies in care.²⁰ Another study found that 58% of medical cases flagged by administrative data as inpatient complications were actually present on admission.²¹ Although refinements in billing data may improve their specificity in identifying certain safety lapses, their validity will remain dependent upon accurate and complete coding of events.^{22,23}

We found a consistent relationship between better performance in failure to rescue and other quality measures examined. Failure to rescue may behave differently than other PSIs since it examines the effectiveness in rescuing patients from complications of care rather than preventing complications.³ The association between failure to rescue and condition-specific in-hospital mortality may not be surprising given that failure to rescue is a type of in-hospital mortality rate. However, the relationships persisted after excluding deaths considered to be failure to rescue cases in each of the medical conditions.

Our finding that the three other medical PSIs tested have poor relationships with quality measures has several potential

interpretations. Preventing complications of medical care may represent a different domain of quality than processes focused on effective care or mortality rates. Alternatively, it might be that PSIs inaccurately capture adverse events and therefore fail to identify safer hospital care. Given the tremendous variability in the coding of secondary ICD-9 diagnoses across hospitals, it is possible that hospitals committed to safety-reporting identify and code more complications. Alternatively, some hospitals may code complications more aggressively due to financial incentives.^{19,24} Teasing these issues apart and understanding why some hospitals perform better in certain PSIs are critical before using them in pay-for-performance incentives or public reporting programs.

Our study has important limitations. First, we do not examine the sensitivity or specificity of the PSI in capturing preventable events, and therefore, this is not meant as a rigorous validation study. Second, each of the quality metrics we used for comparison has its own limitations. The HQA measures focus on only three medical conditions, and the data analyzed come from a later time period than the PSI data. Further, HQA scores have improved over time, which can create difficulties in correctly capturing a hospital's performance at a given instance.²⁵ Risk-adjusted mortality rates have their own limitations: administrative data may not adequately capture

Table 4. Comparison of PSI Rates in US News and Non-US News Hospitals

	US News Hospitals		Non-US News Hospitals		P value†
	Median Rate*	Number of Hospitals	Median Rate*	Number of Hospitals	
Death in low mortality DRG	2.0±2.4	75	0±3.8	3,645	<0.001
Decubitus ulcer	23.8±15.1	75	19.4±15.9	3,948	0.007
Failure to rescue	106.8±30.5	75	127.5±55.8	2,135	<0.001
Infection due to medical care	3.1±1.9	75	0.7±2.1	4,339	<0.001

*Rates are median values of risk-adjusted rates per 1,000 at-risk cases ± interquartile range

†P value calculated using Wilcoxon testing

differences in baseline risk, and in-hospital mortality rates may be biased in favor of institutions that commonly transfer out sick patients. US News only ranks a few top-performing institutions. Further, none of the metrics we chose for comparison were designed to specifically measure hospital safety. Third, we only examined 4 of the possible 20 PSIs, although we chose all PSIs closely related to caring for medical (as opposed to surgical or obstetric) patients. Fourth, the use of Medicare data to analyze the PSIs' relationships to other metrics may not generalize easily to rates for non-elderly patients. However, there are benefits in using Medicare data: it is the single largest payer of hospital care, and its well publicized pay-for-performance demonstration is using two PSIs.⁸ Further, given CMS's influence, if it were to broadly adopt the use of PSIs for grading or paying hospitals, it is likely that others would follow. Finally, since our study examines the hospital as the unit of analysis, conclusions about complications rates for individual patients admitted to a hospital cannot be made.

In conclusion, we examined the relationship between select PSIs and other measures of hospital quality and found that low rates of failure to rescue were consistently related to better performance on other inpatient quality measures. However, the other PSIs were usually inversely related to quality metrics. Understanding the factors that underlie this inconsistent relationship is critical and should give pause to policy makers who use PSIs to grade and pay hospitals.

Acknowledgements: We are grateful to E. John Orav, PhD, for his statistical input, Jeffery Geppert, JD, for his help with the PSI software and Drs. Saul Weingart and Lisa Iezzoni for their insightful comments on earlier versions of the manuscript.

This material is the result of work supported with resources and the use of facilities at the Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System.

Conflict of Interest: None disclosed.

Corresponding Author: Thomas Isaac, MD, MBA, MPH; VA Boston Healthcare System, 150 South Huntington Avenue, Boston, MA 02115, USA (e-mail: txi001@gmail.com).

REFERENCES

1. Kohn LT, Corrigan J, Donaldson M, eds. To Err is Human: Building a Safer Health Care System. In Washington, D.C.: National Academy Press: 1999.
2. Brennan TA, Leape LL, Laird NM, et al.. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med. 1991;324(6):370-6.
3. AHRQ Quality Indicators - Guide to Patient Safety Indicators. In Rockville, MD, Agency for Healthcare Research and Quality, 2003.
4. Romano PS, Geppert JJ, Davies S, Miller MR, Elixhauser A, McDonald KM. A national profile of patient safety in U.S. hospitals. Health Aff (Millwood). 2003;22(2):154-66.
5. National Healthcare Quality Report, 2005. In Rockville, MD, AHRQ, 2005.
6. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA. 2003;290(14):1868-74.
7. HealthGrades Quality Study: Fourth Annual Patient Safety in American Hospitals Study. In 2007.
8. Premier Hospital Quality Incentive Demonstration. In 2004.
9. Thornlow DK, Stukenborg GJ. The association between hospital characteristics and rates of preventable complications and adverse events. Med Care. 2006;44(3):265-9.

10. Miller MR, Pronovost P, Donithan M, et al. Relationship between performance measurement and accreditation: implications for quality of care and patient safety. Am J Med Qual. 2005;20(5):239-52.
11. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998;36(1):8-27.
12. AHRQ Quality Indicators - Guidance for Using the AHRQ Quality Indicators for Hospital-level Public Reporting or Payment. In Rockville, MD, AHRQ, 2004.
13. United States Department of Health and Human Services: Hospital Compare. Available at: <http://www.hospitalcompare.hhs.gov/>
14. Jha AK, Li Z, Orav EJ, Epstein AM. Care in U.S. hospitals-the Hospital Quality Alliance program. N Engl J Med. 2005;353(3):265-74.
15. Kahn CN 3rd, Ault T, Isenstein H, Potetz L, Gelder SV. Snapshot of Hospital Quality Reporting and Pay-For-Performance Under Medicare. Health Aff (Millwood). 2006;25(1):149-62.
16. Inpatient Quality Indicators - 3M™ APR™ DRG Limited License Grouper. In
17. AHRQ Quality Indicators - Guide to Inpatient Quality Indicators: Quality of Care in Hospitals - Volume, Mortality, and Utilization. Rockville, MD, AHRQ, 02-RO204; 2002.
18. McFarlane E, Murphy J, Olmsted MG, Drozd EM, Hill C. America's Best Hospitals: 2006 Methodology.
19. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. Qual Saf Health Care. 2003;12(Suppl 2):ii58-ii63.
20. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. Med Care. 2000;38(8):796-806.
21. Lawthers AG, McCarthy EP, Davis RB, Peterson LE, Palmer RH, Iezzoni LI. Identification of in-hospital complications from claims data. Is it valid? Med Care. 2000;38(8):785-95.
22. Iezzoni LI. Finally present on admission but needs attention. Med Care. 2007;45(4):280-2.
23. McDonald KM, Davies SM, Geppert J, Romano PS. Why rescue the administrative data version of the "failure to rescue" quality indicator. Med Care. 2007;45(4):277-9.
24. Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP. Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. N Engl J Med. 1988;318(6):352-5.
25. Williams SC, Schmaltz SP, Morton DJ, Koss RG, Loeb JM. Quality of care in US\$ hospitals as reflected by standardized measures, 2002-2004. N Engl J Med. 2005;353(3):255-64.

APPENDIX

Table 5. Hospital Characteristics

	All US Gen Med Hospitals with at Least One Calculable Medical PSI (n=4,414)	US Gen Med Hospitals with No Calculable Medical PSI (n=90)
Bed size		
● 0-99	1997 (45)	83 (92)
● 100-399	1996 (45)	5 (6)
● >400	421 (10)	2 (3)
Region		
● Northeast	606 (14)	5 (5)
● West	817 (19)	37 (41)
● Midwest	1306 (30)	24 (27)
● South	1685 (38)	24 (27)
For profit	656 (15)	8 (9)
Members of	277 (6)	0 (0)
COTH		
Urban	2373 (54)	17 (19)
Medical ICU	2946 (75)	7 (11)
Cardiac ICU	1510 (38)	1 (1)
Percent	47.9±14.4	41.7±25.1
Medicare		
Percent	15.9±10.3	15.0±13.9
Medicaid		