

# NATIONAL QUALITY FORUM

## Measure Evaluation 4.1 December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the [evaluation criteria](#) are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

**TAP/Workgroup** (if utilized): Complete all **yellow highlighted** areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: If there is no TAP or workgroup, the SC also evaluates the subcriteria (**yellow highlighted areas**).

**Steering Committee:** Complete all **pink** highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met

C = Completely (unquestionably demonstrated to meet the criterion)

P = Partially (demonstrated to partially meet the criterion)

M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)

N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)

NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) NQF Review #: 0200	NQF Project: Surgery Endorsement Maintenance 2010
<b>MEASURE DESCRIPTIVE INFORMATION</b>	
<b>De.1 Measure Title:</b> <a href="#">Death among surgical inpatients with treatable serious complications (failure to rescue)</a>	
<b>De.2 Brief description of measure:</b> <a href="#">Percentage of surgical inpatients with complications of care whose status is death</a> <a href="#">Percentage of surgical inpatients with complications of care whose status is death</a>	
<b>1.1-2 Type of Measure:</b> <a href="#">Outcome</a>	
<b>De.3</b> If included in a composite or paired with another measure, please identify composite or paired measure	
<b>De.4 National Priority Partners Priority Area:</b> <a href="#">Safety</a>	
<b>De.5 IOM Quality Domain:</b>	
<b>De.6 Consumer Care Need:</b>	

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	<b>NQF Staff</b>
<b>A.</b> The measure is in the public domain or an intellectual property ( <a href="#">measure steward agreement</a> ) is signed. <i>Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.</i> <b>A.1</b> Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)? <b>A.2</b> Indicate if Proprietary Measure (as defined in measure steward agreement): <b>A.3</b> Measure Steward Agreement: <b>A.4</b> Measure Steward Agreement attached:	<b>A</b> <b>Y</b> <input checked="" type="radio"/> <b>N</b> <input type="radio"/> <b>N</b> <input type="radio"/> <b>O</b> <input type="radio"/>
<b>B.</b> The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least	<b>B</b> <b>Y</b> <input checked="" type="radio"/> <b>O</b> <input type="radio"/>

every 3 years.	NO
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ► <b>Actual/Planned Use:</b>	C YO NO
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1 Testing: D.2 Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?	D YO NO
(for NQF staff use) Have all conditions for consideration been met? Staff Notes to Steward (if submission returned):	Met YO NO
Staff Notes to Reviewers (issues or questions regarding any criteria):	
Staff Reviewer Name(s):	

**Comment [KP]:** 1a. The measure focus addresses:

- a specific national health goal/priority identified by NQF's National Priorities Partners; OR
- healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
<b>1. IMPORTANCE TO MEASURE AND REPORT</b>	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. <i>Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)</i> (1a. High Impact)	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 Demonstrated High Impact Aspect of Healthcare: 1a.2  1a.3 Summary of Evidence of High Impact: PSI 4 A higher risk-adjusted mortality rate for death among surgical inpatients with serious treatable complications is associated with significantly higher costs. The AHRQ QIs have the advantage of taking the multidimensional nature of hospital quality into account. As the coefficients on the AHRQ QIs show, measures of hospital quality can have conflicting effects on hospital costs. A single measure that combines these effects into one variable offers less insight into hospital performance than the outcomes for each measure. [1]  Patient Safety Events Are Common at U.S. Hospitals: Between 2005 and 2007 there were 913,215 total patient safety events among Medicare beneficiaries. Common Patient Safety Events are Very Costly: Between 2005 and 2007 these patient safety events were associated with over \$6.9 billion of wasted healthcare cost. Less Improvement Seen Among Most Common Events: Eight patient safety indicators showed improvement while seven indicators worsened in 2007 compared to 2005. Some of the most common and most serious indicators worsened, including decubitus ulcer (bed sores), sepsis, respiratory failure, deep vein thrombosis (blood clots in the legs), and pulmonary embolism (potentially fatal blood clots forming in the lungs). Approximately One-in-Ten Medicare Patients with Patient Safety Events Died: Between 2005 and 2007 there were 97,755 actual in-hospital deaths that occurred among patients who experienced one or more of the 15 patient safety events. [2] Silber and colleagues have published a series of studies establishing the construct validity of failure-to-rescue rates through their associations with hospital characteristics and other measures of hospital performance. Among patients admitted for cholecystectomy and transurethral prostatectomy, failure to	1a CO PO MO NO

rescue was independent of severity of illness at admission, but was significantly associated with the presence of surgical house staff and a lower percentage of board-certified anesthesiologists. [3] The adverse occurrence rate was independent of this hospital characteristic. In a larger sample of patients who underwent general surgical procedures, lower failure-to-rescue rates were found at hospitals with high ratios of registered nurses to beds. [4] Failure rates were strongly associated with risk-adjusted mortality rates, as expected, but not with complication rates. [5]

More recently, Needleman and Buerhaus confirmed that higher registered nurse staffing (RN hours/adjusted patient day) and better nursing skill mix (RN hours/licensed nurse hours) were consistently associated with lower failure-to-rescue rates, even using administrative data to define complications. [6]

**1a.4 Citations for Evidence of High Impact:** [1] Laditka JN, Laditka SB, Cornman CB. Evaluating hospital care for individuals with Alzheimer's disease using inpatient quality indicators. Am J Alzheimers Dis Other Demen. 2005 Jan-Feb;20(1):27-36. PMID: 15751451.

[2] HealthGrades. Every 1.7 Minutes a Medicare Beneficiary Experiences a Patient Safety Event. Business Wire. Available on-line: <http://www.allbusiness.com/government/government-bodies-offices/12279340-1.html>. Accessed 1/11/2011.

[3] Silber JH, Williams SV, Krakauer H, Schwartz JS. Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue. Med Care 1992;30(7):615-29.

[4] Silber J, Rosenbaum P, Ross R. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? J Am Stat Assoc 1995;90:7-18.

[5] Silber JH, Rosenbaum PR, Williams SV, Ross RN, Schwartz JS. The relationship between choice of outcome measure and hospital rank in general surgical procedures: Implications for quality assessment. Int J Qual Health Care 1997;9(3):193-200.

[6] Needleman J, Buerhaus PI, Mattke S, Stewart M, Zelevinsky K. Nurse Staffing and Patient Outcomes in Hospitals. Boston MA: Health Resources and Services Administration; 2001 February 28. Report No.:230-99-0021.

#### 1b. Opportunity for Improvement

**1b.1 Benefits (improvements in quality) envisioned by use of this measure:** This indicator is intended to identify patients who die following the development of a complication. The underlying assumption is that good hospitals identify these complications quickly and treat them aggressively.

**1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:**

Adjusted rates by patient/hospital characteristics, 2007

Estimate	Standard error	Age: for conditions affecting any age
70.969	0.516	18-44
106.977	0.376	45-64
124.040	0.469	65 and over

Estimate	Standard error	Gender
112.064	0.363	Male
99.768	0.366	Female

Estimate	Standard error	Median income of patient's ZIP code
107.685	0.446	First quartile (lowest income)
106.520	0.514	Second quartile
103.842	0.541	Third quartile
103.204	0.583	Fourth quartile (highest income)

Estimate	Standard error	Location of patient residence (NCHS)
107.015	0.442	Large central metropolitan

1b  
C  
P  
M  
N

**Comment [KP]:** 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

**Comment [k]:** 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

100.585	0.531	Large fringe metropolitan
107.118	0.593	Medium metropolitan
106.053	0.860	Small metropolitan
110.033	0.833	Micropolitan
105.732	1.069	Not metropolitan or micropolitan
Estimate	Standard error	Expected payment source
101.823	0.497	Private insurance
103.325	0.362	Medicare
110.349	0.684	Medicaid
114.903	1.368	Other insurance
126.797	1.093	Uninsured / self-pay / no charge
Estimate	Standard error	Hospital Ownership/control
102.652	0.302	Private, not-for-profit
113.708	0.712	Private, for-profit
113.224	0.666	Public
Estimate	Standard error	Teaching status
104.802	0.412	Teaching
106.210	0.328	Nonteaching
Estimate	Standard error	Location of hospital
105.513	0.399	Large central metropolitan
102.973	0.594	Large fringe metropolitan
106.869	0.570	Medium metropolitan
107.649	0.792	Small metropolitan
107.499	0.934	Micropolitan
104.114	1.972	Not metropolitan or micropolitan
Estimate	Standard error	Bed size of hospital
101.319	0.969	Less than 100
108.257	0.428	100 - 299
104.302	0.447	300 - 499
104.935	0.524	500 or more

**1b.3 Citations for data on performance gap:**

See the following report for a complete treatment of the methodology: "Methods: Applying AHRQ Quality Indicators to Healthcare Cost and Utilization Project (HCUP) Data for the National Healthcare Quality Report" [URL: <http://hcupnet.ahrq.gov/QI%20Methods.pdf?JS=Y>]

**1b.4 Summary of Data on disparities by population group:**

After adjusting for age, gender, race, diabetes, CVD, and cancer, compared with those without CKD, hospitalized patients with CKD were showed no difference in death among surgical inpatients with serious treatable complications (aRR = 0.95, 95% CI = 0.85 to 1.05, p = 0.600). [1]

A total of 1.35 million trauma patients were identified, with 19,338 patients (1.43%) experiencing at least one of the applicable PSIs. On multivariate analysis, controlling for injury severity and disease comorbidity, the adjusted odds ratios (ORs) for occurrence of at least 1 applicable PSI were noted to increase for patients who are 1) above age 35, 2) male gender (OR 1.25, 95% CI 1.19-1.31), and 3) black (OR 1.20 vs. whites, 95% CI 1.10-1.30) but not for any other racial groups. These results did not change significantly on sensitivity analysis. Patients who are above age 35, male gender, and black are associated with increased

likelihood of experiencing a patient safety event in trauma care. When all else is equal, black patients are approximately 20% more likely than any other racial groups to experience a patient safety event, even after controlling for injury severity and disease comorbidity. [2]

Patients age 65 and older experienced significantly higher rates than younger patients for in death among surgical inpatients with serious treatable complications. [3]

#### References

- [1] Seliger SL; Zhan M; Hsu VanD; Walker LD; Fink JC. Chronic kidney disease adversely influences patient safety. J Am Soc Nephrol. 2008 December; 19(12): 2414-2419. doi: 10.1681/ASN.2008010022.  
 [2] Chang DC, Handly N, Abdullah F, Efron DT, Haut ER, Haider AH, Pronovost PJ, Cornwell EE. The occurrence of potential patient safety events among trauma patients: are they random? Ann Surg. 2008 Feb;247(2):327-34. PMID: 18216541  
 [3] Thornlow DK. Increased risk for patient safety incidents in hospitalized older adults. MedSurg Nursing, 18, 5, 287(5).

#### 1b.5 Citations for data on Disparities:

Data for patients hospitalized in the Veteran's Health Administration during 2004 to 2005 was analyzed to conduct a cross-sectional study of Chronic Kidney Disease (CKD) and adverse safety events. We identified 315,213 Veterans Health Administration (VHA) patients with at least one acute hospitalization within the study period, CKD was present among 29% (n = 71,666) of the study population, and these patients were older; slightly less likely to be black; and more likely to have diabetes, cardiovascular disease (CVD), cancer, and length of stay (LOS) >3 d than those without CKD. [1]

Retrospective analysis of a nationally representative dataset using Nationwide Inpatient Sample (representative 20% sample from 37 states) for 5 years (2000 through 2004). Outcome = occurrence of at least one of the applicable PSIs on multiple logistic regression analysis, with confirmation by sensitivity analysis. [2]

HCUPnet generated statistics using data from the 2004 Nationwide Inpatient Sample (NIS), which contains all payer data on hospital inpatient stays from states participating in HCUP and is designed to approximate a 20% sample of U.S. community hospitals. As testimony to its size, the 2004 NIS contains data on approximately 8 million inpatient hospital discharge records. Statistical methods not specified. [3]

#### 1c. Outcome or Evidence to Support Measure Focus

**1c.1 Relationship to Outcomes** (For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population): This indicator is intended to identify patients who die following the development of a complication. The underlying assumption is that good hospitals identify these complications quickly and treat them aggressively. Failure to Rescue may be fundamentally different than other indicators reviewed in this report, as it may reflect different aspects of quality of care (effectiveness in rescuing a patient from a complication versus preventing a complication). This indicator includes pediatric patients. It is important to note that children beyond the neonatal period inherently recover better from physiological stress and thus may have a higher rescue rate.

**1c.2-3. Type of Evidence:** Expert opinion, Systematic synthesis of research

**1c.4 Summary of Evidence** (as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome):

PSI 4 A higher risk-adjusted mortality rate for death among surgical inpatients with serious treatable complications is associated with significantly higher costs. The AHRQ QIs have the advantage of taking the multidimensional nature of hospital quality into account. As the coefficients on the AHRQ QIs show, measures of hospital quality can have conflicting effects on hospital costs. A single measure that combines these effects into one variable offers less insight into hospital performance than the outcomes for each measure. [1]

Patient Safety Events Are Common at U.S. Hospitals: Between 2005 and 2007 there were 913,215 total

**Comment [k]: 1c.** The measure focus is:

- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;

OR

- if an intermediate outcome, process, structure, etc., there is evidence that supports the specific measure focus as follows:
  - o **Intermediate outcome** - evidence that the measured intermediate outcome (e.g., blood pressure, HbA1c) leads to improved health/avoidance of harm or cost/benefit.
  - o **Process** - evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
  - o **Structure** - evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.

**Comment [k]: 4** Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status - patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

1c  
CO  
PO  
MO  
NO

patient safety events among Medicare beneficiaries. Common Patient Safety Events are Very Costly: Between 2005 and 2007 these patient safety events were associated with over \$6.9 billion of wasted healthcare cost. Less Improvement Seen Among Most Common Events: Eight patient safety indicators showed improvement while seven indicators worsened in 2007 compared to 2005. Some of the most common and most serious indicators worsened, including decubitus ulcer (bed sores), sepsis, respiratory failure, deep vein thrombosis (blood clots in the legs), and pulmonary embolism (potentially fatal blood clots forming in the lungs). Approximately One-in-Ten Medicare Patients with Patient Safety Events Died: Between 2005 and 2007 there were 97,755 actual in-hospital deaths that occurred among patients who experienced one or more of the 15 patient safety events. [2] Silber and colleagues have published a series of studies establishing the construct validity of failure-to-rescue rates through their associations with hospital characteristics and other measures of hospital performance. Among patients admitted for cholecystectomy and transurethral prostatectomy, failure to rescue was independent of severity of illness at admission, but was significantly associated with the presence of surgical house staff and a lower percentage of board-certified anesthesiologists. [3] The adverse occurrence rate was independent of this hospital characteristic. In a larger sample of patients who underwent general surgical procedures, lower failure-to-rescue rates were found at hospitals with high ratios of registered nurses to beds. [4] Failure rates were strongly associated with risk-adjusted mortality rates, as expected, but not with complication rates. [5] More recently, Needleman and Buerhaus confirmed that higher registered nurse staffing (RN hours/adjusted patient day) and better nursing skill mix (RN hours/licensed nurse hours) were consistently associated with lower failure-to-rescue rates, even using administrative data to define complications. [6]

## References:

- [1] Laditka JN, Laditka SB, Cornman CB. Evaluating hospital care for individuals with Alzheimer's disease using inpatient quality indicators. *Am J Alzheimers Dis Other Demen*. 2005 Jan-Feb;20(1):27-36. PMID: 15751451.
- [2] HealthGrades. Every 1.7 Minutes a Medicare Beneficiary Experiences a Patient Safety Event. *Business Wire*. Available on-line: <http://www.allbusiness.com/government/government-bodies-offices/12279340-1.html>. Accessed 1/11/2011.
- [3] Silber JH, Williams SV, Krakauer H, Schwartz JS. Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue. *Med Care* 1992;30(7):615-29.
- [4] Silber J, Rosenbaum P, Ross R. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? *J Am Stat Assoc* 1995;90:7-18.
- [5] Silber JH, Rosenbaum PR, Williams SV, Ross RN, Schwartz JS. The relationship between choice of outcome measure and hospital rank in general surgical procedures: Implications for quality assessment. *Int J Qual Health Care* 1997;9(3):193-200.
- [6] Needleman J, Buerhaus PI, Mattke S, Stewart M, Zelevinsky K. Nurse Staffing and Patient Outcomes in Hospitals. Boston MA: Health Resources and Services Administration; 2001 February 28. Report No.:230-99-0021.

**1c.5 Rating of (strength/quality of evidence) (also provide narrative description of the rating and by whom):**

Not Applicable.

**1c.6 Method for rating evidence:** Not Applicable.

**1c.7 Summary of Controversy/Contradictory Evidence:** Not Applicable.

**1c.8 Citations for Evidence (other than guidelines):** Not Applicable.

**1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number):**  
Not Applicable.

**1c.10 Clinical Practice Guideline Citation:** Not Applicable.

**1c.11 National Guideline Clearinghouse or other URL:** Not Applicable.

**Comment [k]:** 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system <http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm>). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.



1c.12 <b>Rating of strength of recommendation</b> (also provide narrative description of the rating and by whom): Not Applicable.	
1c.13 <b>Method for rating strength of recommendation</b> (If different from <a href="#">USPSTF system</a> , also describe rating and how it relates to USPSTF): Not Applicable.	
1c.14 <b>Rationale for using this guideline over others:</b> Not Applicable.	
<b>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Importance to Measure and Report?</b>	1
<b>Steering Committee: Was the threshold criterion, Importance to Measure and Report, met?</b> Rationale:	1 Y O N
<b>2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES</b>	
Extent to which the measure, <u>as specified</u> , produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ( <a href="#">evaluation criteria</a> )	<a href="#">Eval</a> <a href="#">Rating</a>
<b>2a. MEASURE SPECIFICATIONS</b>	
S.1 Do you have a web page where current detailed measure specifications can be obtained? S.2 If yes, provide web page URL:	2a- specs CO PO MO NO
<b>2a. Precisely Specified</b>	
2a.1 <b>Numerator Statement</b> (Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome): <a href="#">Surgical inpatients with complications of care whose discharge status is death</a>	
2a.2 <b>Numerator Time Window</b> (The time period in which cases are eligible for inclusion in the numerator):	
2a.3 <b>Numerator Details</b> (All information required to collect/calculate the numerator, including all codes, logic, and definitions):	
2a.4 <b>Denominator Statement</b> (Brief, text description of the denominator - target population being measured): <a href="#">Major surgical discharges with complications of care: sepsis (ICD-9-CM codes 038, 790.7), pneumonia (ICD-9-CM codes 507.0, 997.3, 514, 482, 485, 486), GI bleeding (ICD-9-CM codes 531.00-531.31, 531.9, 532.00-532.31, 532.9, 533.00-533.31, 533.9, 534.00-534.31, 534.9, 535.01, 535.4, 578.9, 530.82), shock/cardiac arrest (ICD-9-CM codes 427.5, 785.5, 785.50, 785.51, 785.59, 799.1, 93.93, 99.60, 99.63), DVT/PE (ICD-9-CM codes 451.81, 451.11, 451.19, 415.11, 415.1, 453.8)</a>	
2a.5 <b>Target population gender:</b> 2a.6 <b>Target population age range:</b>	
2a.7 <b>Denominator Time Window</b> (The time period in which cases are eligible for inclusion in the denominator):	
2a.8 <b>Denominator Details</b> (All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions):	

**Comment [k]:** USPSTF grading system <http://www.ahrq.gov/clinic/uspstf/grade.htm>: A - The USPSTF recommends the service. There is high certainty that the net benefit is substantial. B - The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. C - The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. D - The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. I - The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

**Comment [KP]:** 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP).

<p><b>2a.9 Denominator Exclusions</b> (Brief text description of exclusions from the target population): <b>Exclusions</b> as noted for each complication of care :</p> <p>Sepsis excluding sepsis as a primary diagnosis, AIDS and immunocompromised states, LOS &lt; 3 days, infection related admission</p> <p>Pneumonia excluding any primary diagnosis of pneumonia, any secondary diagnosis of ICD-9-CM 480, 481, 483, 484, 487, MDC4 (respiratory system), AIDS and immunocompromised states</p> <p>GI Bleeding excluding primary GI bleed as a primary diagnosis, MDC 6 (digestive tract), MDC 7 (hepatobiliary tract and pancreas), primary diagnosis of anemia due to blood loss, trauma, burn, alcoholism</p> <p>Shock/cardiac arrest excluding shock as primary diagnosis, MDC 4 (respiratory), MDC 5 (cardiac), hemorrhage or trauma as a primary diagnosis</p> <p>DVT/PE excluding DVT/PE as a primary diagnosis and pregnancy related PE</p> <p><b>2a.10 Denominator Exclusion Details</b> (All information required to collect exclusions to the denominator, including all codes, logic, and definitions):</p> <p><b>2a.11 Stratification Details/Variables</b> (All information required to stratify the measure including the stratification variables, all codes, logic, and definitions):</p> <p><b>2a.12-13 Risk Adjustment Type:</b> Case-mix adjustment</p> <p><b>2a.14 Risk Adjustment Methodology/Variables</b> (List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method):</p> <p>Gender, age, DRG, co-morbidity categories - The AHRQ PSI software includes risk-adjustment variables generated using regression coefficients from a baseline file representing a large proportion of the hospitalized U.S. population.</p> <p>The predicted value for each case is computed using standard logistic regression and covariates for gender, age (in 5-year age groups), modified CMS DRG, and the AHRQ Comorbidity category. The reference population used in the regression is the universe of discharges for states that participate in the HCUP State Inpatient Data (SID) for the years 2001-2003 (combined), a database consisting of 38 states and approximately 90 million discharges. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., hospital, state, and region). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.</p> <p><b>2a.15-17 Detailed risk model available Web page URL or attachment:</b></p> <p><b>2a.18-19 Type of Score:</b></p> <p><b>2a.20 Interpretation of Score:</b></p> <p><b>2a.21 Calculation Algorithm</b> (Describe the calculation of the measure as a flowchart or series of steps):</p> <p><b>2a.22 Describe the method for discriminating performance (e.g., significance testing):</b></p> <p><b>2a.23 Sampling (Survey) Methodology</b> If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):</p> <p><b>2a.24 Data Source</b> (Check the source(s) for which the measure is specified and tested)</p> <p>Electronic administrative data/claims</p> <p><b>2a.25 Data source/data collection instrument</b> (Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.):</p> <p><b>2a.26-28 Data source/data collection instrument reference web page URL or attachment:</b></p>	
---	--

**Comment [k]:** 11 Risk factors that influence outcomes should not be specified as exclusions.  
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.



<p><b>2a.29-31 Data dictionary/code table web page URL or attachment:</b></p> <p><b>2a.32-35 Level of Measurement/Analysis</b> (Check the level(s) for which the measure is specified and tested)  <a href="#">Facility/Agency</a></p> <p><b>2a.36-37 Care Settings</b> (Check the setting(s) for which the measure is specified and tested)  <a href="#">Hospital</a></p> <p><b>2a.38-41 Clinical Services</b> (Healthcare services being measured, check all that apply)</p>	
<b>TESTING/ANALYSIS</b>	
<p><b>2b. Reliability testing</b></p> <p><b>2b.1 Data/sample</b> (description of data/sample and size): We restricted our analysis to 20 states (4) for which HCUP State Inpatient Databases (SID) were available. There were 1,601 nonfederal, urban, general hospitals in those 20 states. Over 300 hospitals were eliminated from the sample because of key missing variables in the American Hospital Association (AHA) Annual Survey of Hospital data, which was also used for this study, or because they had missing observations for some of the OIs that we used. Thus, our sample consisted of 1,290 urban, acute-care hospitals for which complete data were available for 2001. [1]</p> <p>For the sixth consecutive year, HealthGrades has analyzed patient safety among Medicare patients in all of the nearly 5,000 U.S. non-federal hospitals based on 15 indicators of patient safety developed by the federal government's Agency for Healthcare Research and Quality (AHRQ). [2]</p> <p>The AHRQ PSI Version 2.1 has been translated into ICD-10-AM (Australian Modification), and PSI Version 3.0a has been independently translated into ICD-10-GM (German Modification). We converted these two country-specific coding algorithms into ICD-10-WHO (World Health Organization version) and combined them to form one master list. [3]</p> <p>The AHRQ-PSI indicators were refined for use with the condition onset flag, resulting in the AusPSIs. [4]</p> <p>The Agency for Healthcare Research and Quality Patient Safety Indicators (PSIs) were used to identify 14 PSIs among 161,004 surgeries. [5]</p> <p><b>2b.2 Analytic Method</b> (type of reliability &amp; rationale, method for testing):  A likelihood ratio test of the hypothesis that the coefficients on all of these variables were equal to 0 (<math>\lambda</math>) = 35.3, <math>p &lt; .01</math>). [1]</p> <p>AHRQ PSI's applied to 5,000 non-federal hospitals. [2]</p> <p>A modified Delphi method was employed to generate a final ICD-10 WHO coding list. Members of an international expert panel—including physicians, professional medical coders, disease classification specialists, health services researchers, epidemiologists, and users of the PSI—independently evaluated this master list and rated each code as either “include,” “exclude,” or “uncertain,” following the AHRQ PSI definitions. After summarizing the independent rating results, we held a face-to-face meeting to discuss codes for which there was no unanimous consensus and newly proposed codes. [3]</p> <p>These indicators are currently in the International Classification of Diseases Ninth Revision, Clinical Modification (ICD-9-CM) whereas Australia has coded its data in ICD-10-Australian Modification (ICD-10-AM) since 1998. We describe the process recently undertaken to translate and revise the patient safety indicators (PSIs) so they can be of use with ICD-10-AM. [4]</p> <p>We used propensity score matching and multivariate regression analyses to predict expenditures and outcomes attributable to the 14 PSIs. [5]</p>	<p>2b</p> <p>C</p> <p>P</p> <p>M</p> <p>N</p>

**Comment [KP]:** 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

**Comment [k]:** 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

**2b.3 Testing Results** (*reliability statistics, assessment of adequacy in the context of norms for the test conducted*):

PSI 4 A higher risk-adjusted mortality rate for death among surgical inpatients with serious treatable complications is associated with significantly higher costs. The AHRQ QIs have the advantage of taking the multidimensional nature of hospital quality into account. As the coefficients on the AHRQ QIs show, measures of hospital quality can have conflicting effects on hospital costs. A single measure that combines these effects into one variable offers less insight into hospital performance than the outcomes for each measure.

[1]

Patient Safety Events Are Common at U.S. Hospitals: Between 2005 and 2007 there were 913,215 total patient safety events among Medicare beneficiaries. Common Patient Safety Events are Very Costly: Between 2005 and 2007 these patient safety events were associated with over \$6.9 billion of wasted healthcare cost. Less Improvement Seen Among Most Common Events: Eight patient safety indicators showed improvement while seven indicators worsened in 2007 compared to 2005. Some of the most common and most serious indicators worsened, including decubitus ulcer (bed sores), sepsis, respiratory failure, deep vein thrombosis (blood clots in the legs), and pulmonary embolism (potentially fatal blood clots forming in the lungs). Approximately One-in-Ten Medicare Patients with Patient Safety Events Died: Between 2005 and 2007 there were 97,755 actual in-hospital deaths that occurred among patients who experienced one or more of the 15 patient safety events. [2]

PSI 4: death among surgical inpatients with serious treatable complications was not included because many procedure codes are required. [3]

The initial translation (electronic mapping, review and revision by expert coder, programming of codes and testing on data from 1996-1998 [ICD 9-CM] to 1998-2006 [ICD-10-AM, through 4 editions]) found that differences between ICD-9-CM and ICD-10-AM datasets presented some challenges. After this phase, which was faithful to AHRQ's case definitions, the indicators were refined for use with the condition onset flag, resulting in the AusPSIs. [4]

Principal Findings. Excess 90-day expenditures likely attributable to PSIs ranged from \$646 for technical problems (accidental laceration, pneumothorax, etc.) to \$28,218 for acute respiratory failure, with up to 20 percent of these costs incurred postdischarge. With a third of all 90-day deaths occurring postdischarge, the excess death rate associated with PSIs ranged from 0 to 7 percent. The excess 90-day readmission rate associated with PSIs ranged from 0 to 8 percent. Overall, 11 percent of all deaths, 2 percent of readmissions, and 2 percent of expenditures were likely due to these 14 PSIs. Conclusions. The effects of medical errors continue long after the patient leaves the hospital. Medical error studies that focus only on the inpatient stay can underestimate the impact of patient safety events by up to 20-30 percent. [5]

**References**

[1] Laditka JN, Laditka SB, Cornman CB. Evaluating hospital care for individuals with Alzheimer's disease using inpatient quality indicators. *Am J Alzheimers Dis Other Dement.* 2005 Jan-Feb;20(1):27-36. PMID: 15751451.

[2] HealthGrades. Every 1.7 Minutes a Medicare Beneficiary Experiences a Patient Safety Event. *Business Wire.* Available on-line: <http://www.allbusiness.com/government/government-bodies-offices/12279340-1.html>. Accessed 1/11/2011.

[3] Hude Quan, MD, PhD; Saskia Dröslér, MD; Vijaya Sundararajan, et al. Adaptation of AHRQ Patient Safety Indicators for Use in ICD-10 Administrative Data by an International Consortium. In *Advances in Patient Safety: New Directions and Alternative Approaches* (Vol. 1: Assessment). Henriksen K, Battles JB, Keyes MA, et al., editors. Rockville (MD): Agency for Healthcare Research and Quality; 2008 Aug. Bookshelf ID: NBK43634.

[4] McConchie S, Shephard J, Waters S, McMillan AJ, Sundararajan V. The AusPSIs: the Australian version of the Agency of Healthcare Research and Quality patient safety indicators. *Aust Health Rev.* 2009 May;33(2):334-41. PMID: 19563325.

[5] Encinosa WE, Hellinger FJ. The impact of medical errors on ninety-day costs and outcomes: an examination of surgical patients. *Health Serv Res.* 2008 Dec;43(6):2067-85. Epub 2008 Jul 25. PMID:

18662169; DOI: 10.1111/j.1475-6773.2008.00882.x

**2c. Validity testing**

**2c.1 Data/sample** (*description of data/sample and size*): Acute care hospitals in the USA, UK, Sweden, Spain, Germany, Canada and Australia in 2004 and 2005/2006. [1]

We carried out a retrospective cross-sectional study on all hospital inpatients discharged in 2005 (including deaths) from the three Mayo Clinic Rochester hospitals (n = 60 599) to assess adverse events. [2]

The University of Michigan Health System (UMHS) is an academic medical center and includes 3 hospitals: University Hospital, Mott Children's Hospital, and Women's Hospital which have a total of 802 beds and more than 44,000 annual discharges. During a 4-week pilot, 66 PSI cases for 56 patients were identified. [3]

Data on healthcare resources were obtained from the Specialized Care Information System and from MBDS records for 2006 from 20 public and state-assisted hospitals within the Madrid Health System. [4]

We undertook a retrospective chart review of 60 denominator cases of death among surgical inpatients with serious treatable complications (FTR) identified by the algorithm at each of 40 University HealthSystem Consortium institutions. [5]

Nine years (1997-2005) of all Veterans Health Administration (VA) administrative hospital discharge data. [6]

The PSIs were applied to all acute inpatient hospitalizations at Veterans Health Administration (VA) facilities in fiscal 2001. [7]

We used the 2003 MedPAR dataset to examine the performance of 4,504 acute-care hospitals on four medical PSIs among Medicare enrollees. [8]

Hospital discharges from Mayo Clinic Rochester hospitals in 2005 (N = 60,599). All hospital inpatients including surgical, medical, pediatric, maternity, psychiatric, and rehabilitation patients. About 33% of patients traveled more than 120 miles for care. [9]

**2c.2 Analytic Method** (*type of validity & rationale, method for testing*):

Routine hospitalization-related administrative data from seven countries were analyzed. Using algorithms adapted to the diagnosis and procedure coding systems in place in each country, authorities in each of the participating countries reported summaries of the distribution of hospital-level and overall (national) rates for each AHRQ Patient Safety Indicator to the OECD project secretariat. [1]

Adverse events were identified through multiple methods: (i) Agency for Healthcare Research and Quality-defined patient safety indicators (PSIs) using ICD-9 diagnosis codes from administrative discharge abstracts, (ii) provider-reported events, and (iii) Institute for Healthcare Improvement Global Trigger Tool with physician confirmation. PSIs were adjusted to exclude patient conditions present at admission. [2]

The patient safety indicator (PSI)-case review system (CRS) is based on the Agency for Healthcare Research and Quality PSIs and uses a diagnosis timing variable to ensure that only inpatients with a condition that was acquired after hospital admission are identified and referred to clinicians for investigation. Voluntary reports of patient safety events are compared to PSI identified safety events (includes Adult PSI: Accidental puncture/laceration, Decubitus ulcer, Death among Surgical Inpatients with Serious Treatable Complications, Iatrogenic pneumothorax, Obstetric trauma: vaginal delivery, Postoperative hemorrhage/hematoma, Postoperative DVT/PE, Postoperative physiological/metabolic, Postoperative sepsis, Selected infections due to medical care and Pediatric PSI: Accidental puncture/laceration, Foreign body left in during procedure, Postoperative sepsis, Selected infections due to medical care). [3]

Correlational analysis. [4]

The primary outcome was the overall accuracy of the algorithm compared with chart review. We also

2c  
CO  
PO  
MO  
NO

**Comment [KP]:** 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

**Comment [k]:** 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

assessed accuracy by complication type, patient characteristics, institution, service assignment, and mortality. [5]

Retrospective analysis using diagnoses and procedures to derive annual rates and standard errors for 13 PSIs. For either hospitals or hospital networks (Veterans Integrated Service Networks [VISNs]), we calculated the percentages whose PSI rates were consistently high or low across years, as well as 1-year lagged correlations, for each PSI. We related our findings to the average annual number of adverse events that each PSI represents. We also assessed time trends for the entire VA, by VISN, and by hospital. [6]

Two methods-regression analysis and multivariable case matching- were used independently to control for patient and facility characteristics while predicting the effect of the PSI on each outcome. [7]

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. [8]

Hospital discharges from Mayo Clinic Rochester hospitals in 2005 (N = 60,599). All hospital inpatients including surgical, medical, pediatric, maternity, psychiatric, and rehabilitation patients. About 33% of patients traveled more than 120 miles for care. [9]

**2c.3 Testing Results** (*statistical results, assessment of adequacy in the context of norms for the test conducted*):

Each country's vector of national indicator rates and the vector of American patient safety indicators rates published by AHRQ (and re-estimated as part of this study) were highly correlated (0.821-0.966). However, there was substantial systematic variation in rates across countries. This pilot study reveals that AHRQ Patient Safety Indicators can be applied to international hospital data. However, the analyses suggest that certain indicators (e.g. 'birth trauma', 'complications of anesthesia') may be too unreliable for international comparisons. Data quality varies across countries; undercoding may be a systematic problem in some countries. Efforts at international harmonization of hospital discharge data sets as well as improved accuracy of documentation should facilitate future comparative analyses of routine databases. [1]

About 4% (2401) of hospital discharges had an adverse event identified by at least one method. Around 38% (922) of identified events were provider-reported events. Nearly 43% of provider-reported adverse events were skin integrity events, 23% medication events, 21% falls, 1.8% equipment events and 37% miscellaneous events. Patients with adverse events identified by one method were not usually identified using another method. Only 97 (6.2%) of hospitalizations with a PSI also had a provider-reported event and only 10.5% of provider-reported events had a PSI. Different detection methods identified different adverse events. 4 Discharges with PSI: death among surgical inpatients with serious treatable complications = 161; Discharges with corresponding provider-reported adverse event = 25 (15.5%) [2]

Sixty-four cases (96%) were unrelated to adverse events recorded in the voluntary patient safety reporting system. Clinicians reviewed 50 cases, and 43 (86%) had a confirmed event. Nineteen (44%) were deemed potentially preventable and were associated with selected infections due to medical care, postoperative sepsis, postoperative pulmonary embolism or deep vein thrombosis, decubitus ulcer, and accidental puncture or laceration (See Table 3). [3]

We found correlations above 0.78 between death among surgical inpatients with serious treatable complications and average length of hospital stay. [4]

Of 2354 cases, 1193 (50.7%) were accurately identified by the algorithm as having had at least one of the FTR-qualifying complications during hospitalization. Of the 3073 complications identified in these patients, 1497 (48.7%) were correctly flagged by the algorithm, 907 (29.5%) were present on admission, 419 (13.6%) were not confirmed by chart review, and 250 (8.1%) met a predefined complication-specific criterion for exclusion. The case accuracy rate varied significantly by institution (mean, 50.7%; range, 18.3-100%;  $P < 0.001$ ), service assignment (surgical service, 62.9% vs. nonsurgical service, 42.9%;  $P < 0.001$ ), and mortality (alive, 43.9% vs. dead, 67.5%;  $P < 0.001$ ) but was not affected by patients' age, gender, race, or insurance status. As currently calculated from administrative data, the FTR algorithm misidentifies half of the cases

on average, is least accurate for nonsurgical cases, and is widely variable across institutions. This indicator may be useful internally to flag possible cases of quality failure but has limitations for external institutional comparisons. Improvements in coding quality and consistency across institutions are needed. [5]

PSI rates are more stable for VISNs than for individual hospitals, but only for those PSIs that reflect the most frequent adverse events. Only the most frequent PSIs yield significant time trends, and only for larger systems. Because they are so rare, PSIs are not reliable performance measures to compare individual hospitals. The most frequent PSIs are more stable when applied to hospital networks, but needing large patient samples nullifies their potential value to managers seeking to improve quality locally or to patients seeking optimal care. [6]

The authors found statistically significant ( $p < .0001$ ) excess mortality, LOS, and cost in all groups with PSIs. The magnitude of the excess varied considerably across the PSIs. These VA findings are similar to those from a previously published study of nonfederal hospitals, despite differences between VA and non-VA systems. This study contributes to the literature measuring outcomes of medical errors and provides evidence that AHRQ PSIs may be useful indicators for comparison across delivery systems. [7]

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. 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. [8]

**RESULTS:** Over 90% of all diagnoses were present at admission whereas 27.1% of all inpatients had a secondary diagnosis coded in-hospital. About one-third of discharges with a safety indicator were flagged because of a diagnosis already present at admission, more likely among referral patients. In contrast, 54% of foreign bodies left in wounds were coded as in-hospital conditions. Severity changes during hospitalization were observed in less than 8% of discharges. Slightly over 3% of discharges were assigned to higher weight diagnosis-related groups based on an in-hospital complication. **CONCLUSIONS:** In general, many patient safety indicators do not reliably identify adverse hospital events, especially when applied to academic referral centers. Conditions recorded after admission have minimal impact on comorbidity and severity measures or on Medicare reimbursement. [9]

#### References

- [1] Drösler SE, Klazinga NS, Romano PS, Tancredi DJ, Gogorcena Aoiz MA, Hewitt MC, Scobie S, Soop M, Wen E, Quan H, Ghali WA, Mattke S, Kelley E. Application of patient safety indicators internationally: a pilot study among seven countries. *Int J Qual Health Care*. 2009 Aug;21(4):272-8. Epub 2009 Apr 24. PMID: 19395469.
- [2] Naessens JM; Campbell CR; Huddleston JM; Berg PB; Lefante JJ; Williams AR; and Culbertson RA. A Comparison of Hospital Adverse Events Identified by Three Widely Used Detection Methods. *International Journal for Quality in Health Care*. 2009;21(4):301-307. PMID: 19617381.
- [3] Bahl V.; Thompson MA.; Commisky EL.; Anderson S.; Campbell Jr. DA. Developing an adverse event reporting system using administrative data. *J Patient Saf* 2008;4:31-37. EMCARE No: 351498817.
- [4] Merchante J.M.; Vega A.M.; Hernandez A.P.; Lizarraga C.A. Analysis of the agency for healthcare research and quality patient safety indicators in public hospitals of the autonomous region of Madrid (Spain). *Medicina Clinica ( Med. Clin. ) (Spain)* July 1, 2010, 135/SUPPL. (3-11). DOI: 10.1016/S0025-7753(10)70014-2.
- [5] Horwitz LI, Cuny JF, Ceresse J, Krumholz HM. Failure to rescue: validation of an algorithm using administrative data. *Med Care*. 2007 Apr;45(4):283-7. PMID: 17496710
- [6] West AN, Weeks WB, Bagian JP. Rare adverse medical events in VA inpatient care: reliability limits to using patient safety indicators as performance measures. *Health Serv Res*. 2008 Feb;43(1 Pt 1):249-66. PMID: 18211528.

<p>[7] Rivard PE, Luther SL, Christiansen CL, Shibe Zhao, Loveland S, Elixhauser A, Romano PS, Rosen AK. Using patient safety indicators to estimate the impact of potential adverse events on outcomes. <i>Med Care Res Rev.</i> 2008 Feb;65(1):67-87. PMID: 18184870; doi: 10.1177/1077558707309611</p> <p>[8] Isaac T, Jha AK. Are patient safety indicators related to widely used measures of hospital quality? <i>J Gen Intern Med.</i> 2008 Sep;23(9):1373-8. Epub 2008 Jun 24. PMID: 18574640.</p> <p>[9] Naessens JM, Campbell CR, Berg B, Williams AR, Culbertson R. Impact of diagnosis-timing indicators on measures of safety, comorbidity, and case mix groupings from administrative data sources. <i>Med Care.</i> 2007 Aug;45(8):781-8. PMID: 1766731</p>	
<b>2d. Exclusions Justified</b>	
2d.1 Summary of Evidence supporting exclusion(s):	
2d.2 Citations for Evidence:	
2d.3 Data/sample (description of data/sample and size):	
2d.4 Analytic Method (type analysis & rationale):	2d
2d.5 Testing Results (e.g., frequency, variability, sensitivity analyses):	CO
<b>2e. Risk Adjustment for Outcomes/ Resource Use Measures</b>	PO
2e.1 Data/sample (description of data/sample and size):	MO
2e.2 Analytic Method (type of risk adjustment, analysis, & rationale):	NO
2e.3 Testing Results (risk model performance metrics):	NAO
2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:	
<b>2f. Identification of Meaningful Differences in Performance</b>	
2f.1 Data/sample from Testing or Current Use (description of data/sample and size):	
<p>2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (type of analysis &amp; rationale):</p> <p>Secondary data on hospital patient safety indicators (PSIs), critical access hospital status (CHA), patient case-mix, and market variables, for 89 Iowa rural hospitals during 1997-2004. PSIs were computed from Iowa State Inpatient Databases (SIDs) using Agency for Healthcare Research and Quality indicators software. Hospital CAH status was extracted from Iowa Hospital Association. Patient case-mix variables were extracted from Iowa SIDs. Market variables came from Area Resource File (ARF). We employed quasi-experimental designs that use both control groups and pretests. The hospital-year was the unit of analysis. We used generalized estimating equations logit and random-effects Tobit models to assess the effects of CAH conversion on hospital patient safety. The models were adjusted for patient case-mix and market variables. Sensitivity analyses, which varied by sample and statistical model, were used to examine the robustness of our findings. [1]</p>	
<p>We compared hospitals with online access to UpToDate with other acute care hospitals included in the Thomson 100 Top Hospitals Database (Thomson database). Metrics used in the Thomson database differentiate hospitals on a variety of performance dimensions such as quality and efficiency. Prespecified outcomes were risk-adjusted mortality, complications, the Agency of Healthcare Research and Quality</p>	2f CO PO MO NO

**Comment [KP]:** 2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
- AND
- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

**Comment [k]:** 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

**Comment [KP]:** 2e. For outcome measures and other measures (e.g., resource use) when indicated:

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;

**Comment [k]:** 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by

**Comment [KP]:** 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

**Comment [k]:** 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful;



Patient Safety Indicators, and hospital length of stay among Medicare beneficiaries. Linear regression models were developed that included adjustment for hospital region, teaching status, and discharge volume. [2]

The Nationwide Inpatient Sample and American Hospital Association annual survey data were used for analyses. To increase comparability, the study sample was restricted to hospitals with fewer than 100 beds. Out of 292 hospitals in the sample, 185 were rural hospitals and 107 were urban hospitals. The unit of analysis was the patient. Associations between hospital location and patient and hospital characteristics were determined using 1-way analysis of variance (ANOVA) and Pearson chi-square test. Generalized estimating equation (GEE) regression models were used for multivariable analysis to adjust for clustering of patients within hospitals. Small rural hospitals were used as the reference group. Correlations among the variables were examined to identify possible multicollinearity and used to select variables that were relatively independent (Pearson correlation coefficient < 0.5) for inclusion in multivariable models. [3]

Input and output data from 1,377 urban hospitals were taken from the American Hospital Association Annual Survey and the Medicare Cost Reports. Nurse-sensitive measures of quality came from the application of the Patient Safety Indicator module of the Agency for Healthcare Research and Quality Quality Indicator software (Measures: death among surgical inpatients with serious treatable complications, infection due to medical care, postoperative respiratory failure, and postoperative sepsis to State Inpatient Databases (SID) provided by the Healthcare Cost and Utilization Project (HCUP). In the first step of the study, hospitals' relative output-based efficiency was determined in order to obtain a measure of congestion (i.e., data envelopment analysis (DEA): the productivity loss due to the occurrence of patient safety events). The outputs were adjusted to account for this productivity loss, and a second DEA was performed to obtain input slack values. Differences in slack values between unadjusted and adjusted outputs were used to measure either relative inefficiency or a need for quality improvement. [4]

Secondary data were used to examine relationships between patient-safety-related accreditation standards and patient outcomes in U.S. acute care hospitals. Accreditation performance areas were reduced into subscores to represent patient safety practices. Outcome rates were calculated using the Agency for Healthcare Research and Quality Patient Safety Indicator software. Multivariate regression was performed to determine the significance of the relationships. [5]

AHRQ PSI's applied to 5,000 non-federal hospitals. [6]

We used descriptive statistics to examine the distribution of each variable. Next, we calculated bivariate correlations with each of the variables of interest. To examine the relationship between IT adoption and the PSIs, we used linear regression analyses using each IT measure as an independent variable to predict one of the PSIs as a dependent variable. Given the findings from our descriptive statistics, several PSI variables were log-transformed to facilitate analysis. [7]

**2f.3 Provide Measure Scores from Testing or Current Use** (*description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningful differences in performance*):

CAH conversion in Iowa rural hospitals had no significant impact on the observed rates of death in low-mortality diagnosis-related groups (DRGs). CAH conversion is associated with enhanced performance of certain PSIs. [1]

Hospitals with access to UpToDate (n=424) were associated with significantly better performance than other hospitals in the Thomson database (n=3091) on risk-adjusted measures of patient safety (P=0.0163) and complications (P=0.0012) and had significantly shorter length of stay (by on average 0.167 days per discharge, 95% confidence interval 0.081-0.252 days, P<0.0001). All of these associations correlated significantly with how much UpToDate was used at each hospital. Mortality was not significantly different between UpToDate and non-UpToDate hospitals. **CONCLUSIONS:** An electronic clinical knowledge support system (UpToDate) was associated with improved health outcomes and shorter length of stay among Medicare beneficiaries in acute care hospitals in the United States. Additional studies are needed to clarify whether use of UpToDate is a marker for the better performance, an independent cause of it, or a synergistic part of other quality improvement characteristics at better-performing hospitals. [2]

The bivariate analyses done to examine the statistical relationships between the PSIs and hospital and patient characteristics show that risk of occurrence of PSI event increased with age, higher Charlson comorbidity index, and number of procedures.

The results from multivariable analyses show that the number of procedures was significantly related to death among surgical inpatients with serious treatable complications, indicating that patients with more procedures were at higher risk for PSI occurrences. Major procedural intensity was associated with death among surgical inpatients with serious treatable complications. [3]

Overall, the hospitals in our sample could increase the total amount of outputs produced by an average of 26 percent by eliminating inefficiency. About 3 percent of this inefficiency can be attributed to congestion. Analysis of subsamples showed that teaching hospitals experienced no congestion loss. We found that quality of care could be improved by increasing the number of labor inputs in low-quality hospitals, whereas high-quality hospitals tended to have slack on personnel. Results suggest that reallocation of resources could increase the relative quality among hospitals in our sample. Further, higher quality in some dimensions of care need not be achieved as a result of higher costs or through reduced access to health care. [4]

Accreditation standards reflecting patient safety practices were not related to rates of death among surgical inpatients with serious treatable complications. Death among surgical inpatients with serious treatable complications may require multifaceted strategies that are less easily translated into protocols. [5]

Patients treated at top-performing hospitals had, on average, a 43 percent lower chance of experiencing one or more medical errors compared to the poorest-performing hospitals. If all hospitals performed at the level of Patient Safety Excellence Award(TM) hospitals, approximately 211,697 patient safety events and 22,771 Medicare deaths could have been avoided while saving the U.S. approximately \$2.0 billion from 2005 through 2007. Between 2005 and 2007, 913,215 total patient safety events were recorded among Medicare beneficiaries, which represents 2.3 percent of the nearly 38 million Medicare hospitalizations. This equates to one reported patient safety event every 1.7 minutes. [6]

In univariate analyses, without risk adjustment most PSI outcome measures were not significantly correlated with any of the IT measures. However, with risk adjustment, death among surgical inpatients with serious treatable complications was weakly and inversely correlated with clinical IT adoption ( $r = -.288$ ,  $p = .008$ ). [7]

#### References

- [1] Li P, Schneider JE, Ward MM. Effect of critical access hospital conversion on patient safety. *Health Serv Res.* 2007 Dec;42(6 Pt 1):2089-108; discussion 2294-323. PMID: 17995554.
- [2] Bonis PA, Pickens GT, Rind DM, Foster DA. Association of a clinical knowledge support system with improved patient safety, reduced complications and shorter length of stay among Medicare beneficiaries in acute care hospitals in the United States. *Int J Med Inform.* 2008 Nov;77(11):745-53. Epub 2008 Jun 19. PMID: 18565788.
- [3] Vartak, S., Ward, M. M. and Vaughn, T. E. (2010), Patient Safety Outcomes in Small Urban and Small Rural Hospitals. *The Journal of Rural Health*, 26: 58-66. doi: 10.1111/j.1748-0361.2009.00266.x
- [4] Valdmanis VG, Rosko MD, and Mutter RL. Hospital Quality, Efficiency, and Input Slack Differentials. *Health Serv Res.* 2008 October; 43(5p2): 1830-1848. doi: 10.1111/j.1475-6773.2008.00893.x.
- [5] Thornlow DK, Merwin E. Managing to improve quality: the relationship between accreditation standards, safety practices, and patient outcomes. *Health Care Manage Rev.* 2009 Jul-Sep;34(3):262-72. PMID: 19625831.
- [6] HealthGrades. Every 1.7 Minutes a Medicare Beneficiary Experiences a Patient Safety Event. *Business Wire.* Available on-line: <http://www.allbusiness.com/government/government-bodies-offices/12279340-1.html>. Accessed 1/11/2011.
- [7] Menachemi, N; Saunders, C; Chukmaitov, A; Matthews, MC.; Brooks, RG. Hospital adoption of information technologies and improved patient safety: A study of 98 hospitals in Florida. *Journal of Healthcare Management*, 52, 6, 2007, 398(13). Available on-line:

<a href="http://www.entrepreneur.com/tradejournals/article/172010511.html">http://www.entrepreneur.com/tradejournals/article/172010511.html</a> Accessed 1/11/2011.	
<b>2g. Comparability of Multiple Data Sources/Methods</b>	
2g.1 Data/sample (description of data/sample and size):	2g
2g.2 Analytic Method (type of analysis & rationale):	CO PO MO NO NAO
2g.3 Testing Results (e.g., correlation statistics, comparison of rankings):	
<b>2h. Disparities in Care</b>	
2h.1 If measure is stratified, provide stratified results (scores by stratified categories/cohorts): [1] Although we did find overall disparities in care, we found that indicators for blacks, Hispanics, and Asians were not statistically worse than corresponding quality indicators for whites in the same hospital. Only a few hospitals provide lower quality of care to minorities than to whites.  [1] Darrell J. Gaskin, Christine S. Spencer, Patrick Richard, Gerard F. Anderson, Neil R. Powe and Thomas A. LaVeist. Do Hospitals Provide Lower-Quality Care To Minorities Than To Whites? Health Affairs, 27, no. 2 (2008): 518-527 doi: 10.1377/hlthaff.27.2.518	2h
2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:	CO PO MO NO NAO
<b>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Scientific Acceptability of Measure Properties?</b>	2
<b>Steering Committee: Overall, to what extent was the criterion, Scientific Acceptability of Measure Properties, met?</b> Rationale:	2 CO PO MO NO
<b>3. USABILITY</b>	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	Eval Rating
<b>3a. Meaningful, Understandable, and Useful Information</b>	
3a.1 Current Use: In use	
3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large) (If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years): In use as a part of the AHRQ Quality Indicators. They are reported in numerous forums including: <a href="http://hcupnet.ahrq.gov/Hcupnet.jsp?Id=EB57801381F71C41&amp;Form=MAINSEL&amp;JS=Y&amp;Action=%3E%3ENext%3E%3E&amp;_MAINSEL=AHQ%20Quality%20Indicators">http://hcupnet.ahrq.gov/Hcupnet.jsp?Id=EB57801381F71C41&amp;Form=MAINSEL&amp;JS=Y&amp;Action=%3E%3ENext%3E%3E&amp;_MAINSEL=AHQ%20Quality%20Indicators</a>  This measure is used in the Monahrq system that is provide for public reporting and quality improvement throughout the United States: <a href="http://monahrq.ahrq.gov/">http://monahrq.ahrq.gov/</a>	
3a.3 If used in other programs/initiatives (If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years): This measure is used in the Monahrq system that is provide for public reporting and quality improvement throughout the United States: <a href="http://monahrq.ahrq.gov/">http://monahrq.ahrq.gov/</a>	3a CO PO MO NO

**Comment [KP]: 2g.** If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

**Comment [KP]: 2h.** If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender); OR rationale/data justifies why stratification is not necessary or not feasible.

**Comment [KP]: 3a.** Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

<b>Testing of Interpretability</b> (Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement)	
3a.4 Data/sample (description of data/sample and size): Not Applicable.	
3a.5 Methods (e.g., focus group, survey, QI project): Not Applicable.	
3a.6 Results (qualitative and/or quantitative results and conclusions): Not Applicable.	
3b/3c. Relation to other NQF-endorsed measures	
3b.1 NQF # and Title of similar or related measures:	
(for NQF staff use) Notes on similar/related endorsed or submitted measures:	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): 3b.2 Are the measure specifications harmonized? If not, why?	3b CO PO MO NO NAO
3c. Distinctive or Additive Value 3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:  5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality: No competing measures found.	3c CO PO MO NO NAO
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Usability?	3
Steering Committee: Overall, to what extent was the criterion, Usability, met? Rationale:	3 CO PO MO NO
<b>4. FEASIBILITY</b>	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes	4a CO PO MO NO
4a.1-2 How are the data elements that are needed to compute measure scores generated? Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition)	
4b. Electronic Sources	
4b.1 Are all the data elements available electronically? (elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims) Yes	4b CO PO MO NO
4b.2 If not, specify the near-term path to achieve electronic capture by most providers.	
4c. Exclusions	4c

**Comment [KP]:** 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

**Comment [k]:** 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

**Comment [KP]:** 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

**Comment [KP]:** 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

**Comment [KP]:** 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

**Comment [KP]:** 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications? Yes	C P M N NA
4c.2 If yes, provide justification. The majority of US hospitals currently capture the required data electronically as a part of their administrative data.	
<b>4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences</b>	
4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results. Panelists noted that several adverse incentives may be introduced by implementing this indicator. In particular, since some type of adjustment may be desirable, this indicator may encourage the upcoding of complications and comorbidities to inflate the denominator or manipulate risk adjustment. Others noted that this indicator could encourage irresponsible resource use and allocation, although this is likely to be a controversial idea.  It is important to note that children beyond the neonatal period inherently recover better from physiological stress and thus may have a higher rescue rate.	4d C P M N
<b>4e. Data Collection Strategy/Implementation</b>	
4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues: Not Applicable.	
4e.2 Costs to implement the measure ( <i>costs of data collection, fees associated with proprietary measures</i> ): All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: <a href="http://www.qualityindicators.ahrq.gov/software.htm">http://www.qualityindicators.ahrq.gov/software.htm</a>	
4e.3 Evidence for costs: All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: <a href="http://www.qualityindicators.ahrq.gov/software.htm">http://www.qualityindicators.ahrq.gov/software.htm</a>	
4e.4 Business case documentation: All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: <a href="http://www.qualityindicators.ahrq.gov/software.htm">http://www.qualityindicators.ahrq.gov/software.htm</a>	4e C P M N
<b>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Feasibility?</b>	4
<b>Steering Committee: Overall, to what extent was the criterion, Feasibility, met?</b> Rationale:	4 C P M N
<b>RECOMMENDATION</b>	
(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.	Time-limited O
<b>Steering Committee: Do you recommend for endorsement?</b> Comments:	Y N

**Comment [KP]:** 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

**Comment [KP]:** 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

<b>A</b>	
CONTACT INFORMATION	
<b>Co.1 Measure Steward (Intellectual Property Owner)</b> <b>Co.1 <u>Organization</u></b> Agency for Healthcare Research and Quality, 540 Gaither Road, Suite 300, Rockville, Maryland, 20850	
<b>Co.2 <u>Point of Contact</u></b> John, Bott, Contractor, AHRQ Quality Indicators Measure Expert Center for Delivery, Organization and Markets, John.Bott@ahrq.hhs.gov, 301-427-1317-	
<b>Measure Developer If different from Measure Steward</b> <b>Co.3 <u>Organization</u></b>  <b>Co.4 <u>Point of Contact</u></b>	
<b>Co.5 Submitter If different from Measure Steward POC</b> John, Bott, MSSW, MBA, john.bott@ahrq.hhs.gov, 301-427-1317-, Agency for Healthcare Research and Quality	
<b>Co.6 Additional organizations that sponsored/participated in measure development</b>	
ADDITIONAL INFORMATION	
<b>Workgroup/Expert Panel involved in measure development</b> <b>Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.</b>	
<b>Ad.2 If adapted, provide name of original measure:</b> <b>Ad.3-5 If adapted, provide original specifications URL or attachment</b>	
<b>Measure Developer/Steward Updates and Ongoing Maintenance</b> <b>Ad.6 Year the measure was first released:</b> <b>Ad.7 Month and Year of most recent revision:</b> <b>Ad.8 What is your frequency for review/update of this measure?</b> <b>Ad.9 When is the next scheduled review/update for this measure?</b>	
<b>Ad.10 Copyright statement:</b>	
<b>Ad.11 Disclaimers:</b>	
<b>Ad.12 -14 Additional Information web page URL or attachment:</b>	
<b>Date of Submission (MM/DD/YY): 01/01/0001</b>	