

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 #: 0353	NQF Project: Surgery Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION	
De.1 Measure Title: Failure to Rescue 30-Day Mortality (risk adjusted)	
De.2 Brief description of measure: Percentage of patients who died with a complication within 30 days from admission.	
1.1-2 Type of Measure: Outcome	
De.3 If included in a composite or paired with another measure, please identify composite or paired measure Failure to Rescue In-Hospital Mortality (risk adjusted)	
De.4 National Priority Partners Priority Area: Safety	
De.5 IOM Quality Domain:	
De.6 Consumer Care Need: Getting better	

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	NQF Staff
<p>A. The measure is in the public domain or an intellectual property (measure steward agreement) 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></p> <p>A.1 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)? Yes</p> <p>A.2 Indicate if Proprietary Measure (as defined in measure steward agreement): Proprietary measure</p> <p>A.3 Measure Steward Agreement: Agreement will be signed and submitted prior to or at the time of measure submission</p> <p>A.4 Measure Steward Agreement attached:</p>	<p>A</p> <p>Y <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and	B

update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	Y <input type="checkbox"/> N <input type="checkbox"/>
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ► Purpose: Public Reporting, Quality Improvement (Internal to the specific organization)	C Y <input type="checkbox"/> N <input type="checkbox"/>
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: Yes, fully developed and tested D.2 Have NQF-endorsed measures been reviewed to identify if there are similar or related measures? Yes	D Y <input type="checkbox"/> N <input type="checkbox"/>
(for NQF staff use) Have all conditions for consideration been met? Staff Notes to Steward (if submission returned):	Met Y <input type="checkbox"/> N <input type="checkbox"/>
Staff Notes to Reviewers (issues or questions regarding any criteria):	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
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. Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria) 1a. High Impact	Eval Rating
(for NQF staff use) Specific NPP goal:	
1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, Frequently performed procedure, Severity of illness 1a.2 1a.3 Summary of Evidence of High Impact: Failure to Rescue measure has a very high impact because it is applicable to the majority of surgical procedures performed at acute care hospitals. Failure to Rescue affects large number of patients and applies to frequently performed procedures. Failure to Rescue, predicts death after an adverse event which accounts for severity of illness to properly adjust the death rate. The measure is less sensitive to errors in severity adjustment (because all patients in the analysis have complications) and more dependent on hospital characteristics relative to patient characteristics than the mortality rate, while having equivalent reliability. FTR has intuitive appeal as a quality marker, attempting to measure a hospital’s ability to manage complications, while being less likely to confuse worse severity of illness with worse quality of care. 1a.4 Citations for Evidence of High Impact: 1. Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. Med Care. 1992;30:615-629. 2. Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: Comparing definitions to measure quality of care. Med Care. 2007;45:918-925. 3. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. JAMA. 1995;274:317-323. 4. Silber JH, Rosenbaum PR, Williams SV, et al. The relationship between choice of outcome measure and	1a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>

<p>hospital rank in general surgical procedures: Implications for quality assessment. <i>Int J Qual Health Care</i>. 1997;9:193-200.</p> <p>5. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist direction and patient outcomes. <i>Anesthesiology</i>. 2000;93:152-163.</p> <p>6. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist board certification and patient outcomes. <i>Anesthesiology</i>. 2002;96:1044-1052.</p> <p>7. Aiken LH, Clarke SP, Sloane DM, et al. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. <i>JAMA</i>. 2002;288:1987-1993.</p> <p>8. Aiken LH, Clarke SP, Cheung RB, et al. Educational levels of hospital nurses and surgical patient mortality. <i>JAMA</i>. 2003;290:1617-1623.</p> <p>9. Silber JH, Rosenbaum PR, Ross RN. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? <i>J Am Stat Assoc</i>. 1995;90:7-18.</p> <p>10. Silber JH, Rosenbaum PR, Romano PS, Rosen AK, Wang Y, Teng Y, Halenar MJ, Even-Shoshan O, Volpp KG. Hospital Teaching Intensity, Patient Race, and Surgical Outcomes. <i>Arch Surg</i>. 2009;144:113-120.</p> <p>11. Friese CR, Earle CC, Silber JH, Aiken LH. Hospital characteristics, clinical severity, and outcomes for surgical oncology patients. <i>Surgery</i> 2010; 147:602-609.</p> <p>12. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in Hospital Mortality Associated with Inpatient Surgery. <i>N Engl J Med</i> 2009; 361:1368-75.</p>	
<p>1b. Opportunity for Improvement</p> <p>1b.1 Benefits (improvements in quality) envisioned by use of this measure: The use of Failure to rescue, predicting death after an adverse occurrence, hospitals would be able to improve their quality of care. Hospitals and health care providers benefit from knowing not only their institution’s mortality rate, but also their institution’s ability to rescue patients after an adverse occurrence. Using failure to rescue measure is especially important if hospital resources needed for prevention were different from those needed for rescue. From a research and policy perspective knowing the failure to rescue rate in addition to the mortality rate will improve our understanding of mortality statistics.</p> <p>1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers: In Aiken et al. shows if the proportion of BSN nurses in all hospitals was 60% rather than 20% 14.2 fewer deaths per 1000 patients with complications (failure to rescue) would be expected. Moreover failure to rescue rates would be decidedly lower if both the workloads of nurses were lighter and the workforce were composed of higher percent-ages of BSN-prepared nurses. (see table 4 in Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational Levels of Hospital Nurses and Surgical Patient Mortality)</p> <p>1b.3 Citations for data on performance gap: In Silber JH et al Hospital Teaching Intensity, Patient Race, Cross-sectional analyses of outcomes data for 232,342 general, orthopedic, and vascular surgery patients discharged from 168 non-federal adult general Pennsylvania hospitals between April 1, 1998, and November 30, 1999, linked to administrative and survey data providing information on educational composition, staffing, and other characteristics.</p> <p>1b.4 Summary of Data on disparities by population group: In Silber JH et al Hospital Teaching Intensity, Patient Race, and Surgical Outcomes. <i>Arch Surg</i>. 2009, shows failure-to rescue rates were consistently lower in hospitals with higher resident-to-bed ratios. Hospitals of high teaching intensity (resident-to-bed ratio=0.6) compared with non-teaching hospitals (resident-to-bed ratio=0) were associated with 14%(95% CI, 12%-15%) lower odds of failure to rescue for combined surgery, with similar finding for subgroup analysis. (see table 3 in paper)</p> <p>1b.5 Citations for data on Disparities: For information reported in 1b4 the data sample was 2,021,214 patients with medicare claims on general, orthopedic, and vascular surgery admissions in the United States for 2000-2005.</p>	<p>1b</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>1c. Outcome or Evidence to Support Measure Focus</p> <p>1c.1 Relationship to Outcomes (For non-outcome measures, briefly describe the relationship to desired</p>	<p>1c</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p>

outcome. For outcomes, describe why it is relevant to the target population): Failure-to-rescue is defined as the probability of death following a complication. The measure will help improve both the management of the hospital and our understanding of hospital mortality rates.

M
N

1c.2-3. Type of Evidence: Cohort study

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

Failure to rescue is influenced by hospital characteristics. Rates differ based on different hospital characteristics such as number of hospital beds, anesthesiologists who are board certified, surgeons who are board certified, presence of house staff, and high technology hospitals, etc. Failure to rescue is an indicator of hospital quality of care. Patients in the age range of 18-90 are analyzed because patients under the age of 18 are considered a pediatric population and have a different set of complications. We use 90 years as a cut-point because of our concern regarding the increased use of do-not-resuscitate at higher ages [Wenger et al. Epidemiology of Do-Not Resuscitate Orders. Disparity by Age, Diagnosis, Gender, Race, and Functional Impairment. Arch Intern Med. 1995; 155(19):2056-62, Hakim et al. Factors Associated with Do-Not-Resuscitate Orders: Patients', Preferences, Prognoses, and Physicians Judgments. Ann Intern Med.1996; 125:284-293.]. While we do adjust for admission severity when reporting FTR, and this includes age, we still thought it prudent to use an upper bound on age, since DNR status prior to the procedure is not well defined at hospitals [Tabak YP, Johannes RS, Silber JH, Kurtz SG, Gibber EM. Should do-not-resuscitate status be included as a mortality risk adjustor? The impact of DNR variations on performance reporting. Med Care 2005; 43:658-666]

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

Silber JH, Williams SV, Krakauer H, et al. Hospital and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. Med Care. 1992;30:615-629. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. JAMA. 1995;274:317-323 Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: Comparing definitions to measure quality of care. Med Care. 2007;45:918-925

1c.6 Method for rating evidence: In Silber et al JAMA 1995, refers to the "power" of a measure as the ability of that measure to detect differences between hospitals or groups of hospitals, with respect to the outcome measure in question. Should the difference between two hospital failure rates achieve statistical significance, while the difference between those same hospitals' death rates not achieve statistical significance, then we would consider the failure rate to be more powerful than the death rate. It can be shown that for equivalent adverse occurrence rates, the power to distinguish between two hospitals using the failure rate is always greater than or equal to the power using the death rate. Although somewhat counterintuitive, this result occurs because, although the failure rate and the death rate use the number of deaths as their numerators, the denominator of the failure rate is the number of patients with adverse occurrences, while the denominator of the death rate is the total number of patients. When adverse occurrence rates are not equal across hospitals, the power of the failure rate statistic may be greater than, equal to, or less than that of the death rate. When comparing two hospitals with failure rates F1 and F2 death rates D1 and D2 and adverse occurrence rates A1 and A2 it can be shown that whenever $F1 \geq F2$, $D1 \geq D2$ and $A1 \leq A2$ then the power in distinguishing such hospitals using the failure rate is greater than or equal to the power when using the death rate. For situations where $F1 \geq F2$ and $D1 < D2$ the sufficient conditions for superior power using the failure rate instead of the death rate is given in the Appendix. Finally, these results are unchanged if one considers either hospital 1 or 2 in the above arguments to be a group of hospitals or the average of all hospitals (so that hospital 1 or 2 represents a very large sample size). In summary, failure rate was a function of anesthesia board certification and the presence of surgical housestaff (hospital characteristics) but not a function of admission severity of illness score (patient characteristics). Since the death rate appears to be composed of two distinct rates, quality of care measurement may be improved if all three rates are reported instead of relying on the adjusted mortality rate alone. In so doing, we may better understand the reasons for variation in hospital mortality rates.

1c.7 Summary of Controversy/Contradictory Evidence: N/A

1c.8 Citations for Evidence *(other than guidelines)*: 1. Silber JH, Williams SV, Krakauer H, et al. Hospital

and patient characteristics associated with death after surgery: A study of adverse occurrence and failure-to-rescue. *Med Care*. 1992;30:615-629.

2. Silber JH, Romano PS, Rosen AK, et al. Failure-to-rescue: Comparing definitions to measure quality of care. *Med Care*. 2007;45:918-925.
3. Silber JH, Rosenbaum PR, Schwartz JS, et al. Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. *JAMA*. 1995;274:317-323.
4. Silber JH, Rosenbaum PR, Williams SV, et al. The relationship between choice of outcome measure and hospital rank in general surgical procedures: Implications for quality assessment. *Int J Qual Health Care*. 1997;9:193-200.
5. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist direction and patient outcomes. *Anesthesiology*. 2000;93:152-163.
6. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist board certification and patient outcomes. *Anesthesiology*. 2002;96:1044-1052.
7. Aiken LH, Clarke SP, Sloane DM, et al. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*. 2002;288:1987-1993.
8. Aiken LH, Clarke SP, Cheung RB, et al. Educational levels of hospital nurses and surgical patient mortality. *JAMA*. 2003;290:1617-1623.
9. Silber JH, Rosenbaum PR, Ross RN. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? *J Am Stat Assoc*. 1995;90:7-18.
10. Silber JH, Rosenbaum PR, Romano PS, Rosen AK, Wang Y, Teng Y, Halenar MJ, Even-Shoshan O, Volpp KG. Hospital Teaching Intensity, Patient Race, and Surgical Outcomes. *Arch Surg*. 2009;144:113-120.
11. Friese CR, Earle CC, Silber JH, Aiken LH. Hospital characteristics, clinical severity, and outcomes for surgical oncology patients. *Surgery* 2010; 147:602-609.
12. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in Hospital Mortality Associated with Inpatient Surgery. *N Engl J Med* 2009; 361:1368-75.

1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number):
 N/A

1c.10 Clinical Practice Guideline Citation: N/A

1c.11 National Guideline Clearinghouse or other URL: N/A

1c.12 Rating of strength of recommendation (also provide narrative description of the rating and by whom):
 N/A

1c.13 Method for rating strength of recommendation (If different from [USPSTF system](#), also describe rating and how it relates to USPSTF):
 N/A

1c.14 Rationale for using this guideline over others:

The motivation behind the development of traditional FTR was based on 2 questions. The first was an empirical question—suppose hospitals were ranked by adjusted mortality and adjusted complication rates. Would these rankings be highly correlated? The answer is rather surprising—there is generally poor correlation or no correlation in most analyses. Second, suppose 2 hospitals had identical adjusted mortality rates but different adjusted complication rates. Would one prefer care at the hospital with the higher or lower complication rate? If one believes that complications are predominantly driven by patient characteristics, then one may decide to choose the hospital with the higher complication rate, as it achieved an equivalent mortality rate with a sicker population of patients. So there is an empirical question to ask—are adjusted complication rates more related to hospital or patient factors? This has been looked at in a number of ways—and the evidence to date suggests that complication measures are less sensitive to hospital characteristics, after adjusting for severity of illness, than mortality based measures. This is an underlying assumption of FTR theory—complications are undesirable outcome measures because they reflect underlying patient severity and diagnosis coding more than they reflect hospital care. Instead, a hospital’s quality is put to the test when a patient develops a complication, and whether a patient is salvaged after a complication will be a function of the care delivered by the hospital and its knowledge base, depth, and facilities. Thus, “good” hospitals will rescue patients by identifying complications quickly and treating them

<p>aggressively, resulting in lower FTR. Although many “failures,” just like deaths, are often not preventable, we have argued that FTR may be a better measure for comparing hospital quality because of better severity adjustment properties, and because of its focus on hospital actions. By studying a population of patients who, by definition, have already developed a complication, the specifics of severity of illness adjustment becomes less important in failure rate analyses, because all patients have experienced complications and thus are more uniformly ill.</p>	
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Importance to Measure and Report</i>?</p>	<p>1</p>
<p>Steering Committee: Was the threshold criterion, <i>Importance to Measure and Report</i>, met? Rationale:</p>	<p>1 Y <input type="checkbox"/> N <input type="checkbox"/></p>
<p style="text-align: center;">2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES</p>	
<p>Extent to which the measure, <u>as specified</u>, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)</p>	<p>Eval Rating</p>
<p style="text-align: center;">2a. MEASURE SPECIFICATIONS</p>	
<p>S.1 Do you have a web page where current detailed measure specifications can be obtained? S.2 If yes, provide web page URL:</p> <p>2a. Precisely Specified</p>	
<p>2a.1 Numerator Statement (<i>Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome</i>): Patients who died with a complication plus patients who died without documented complications. Death is defined as death within 30 days from admission.</p> <p>All patients in an FTR analysis have developed a complication (by definition).</p> <p>Complicated patient has at least one of the complications defined in Appendix B(see website http://www.research.chop.edu/programs/cor/outcomes.php). Complications are defined using the secondary ICD9 diagnosis and procedure codes and the DRG code of the current admission.</p> <p>Comorbidities are defined in Appendix C(see website http://www.research.chop.edu/programs/cor/outcomes.php) using secondary ICD9 diagnosis codes of the current admission and primary or secondary ICD9 diagnosis codes of previous admission within 90 days of the admission date of the current admission.</p> <p>*When physician part B is available, the definition of complications and comorbidities are augmented to include CPT codes.</p> <p>2a.2 Numerator Time Window (<i>The time period in which cases are eligible for inclusion in the numerator</i>): Within 30 days from admission.</p> <p>2a.3 Numerator Details (<i>All information required to collect/calculate the numerator, including all codes, logic, and definitions</i>): Patients who died with complication and patients who died without documented complications. Death is defined as death within 30 days from admission.</p>	
<p>2a.4 Denominator Statement (<i>Brief, text description of the denominator - target population being measured</i>): General Surgery, Orthopedic and Vascular patients in specific DRGs with complications plus patients who died in the hospital without complications.</p> <p>Inclusions: adult patients admitted for one of the procedures in the General Surgery, Orthopedic or Vascular DRGs (see appendix A http://www.research.chop.edu/programs/cor/outcomes.php)</p>	<p>2a-specs C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

<p>Inclusions: adult patients admitted for one of the procedures in the General Surgery, Orthopedic or Vascular DRGs (see appendix A)</p> <p>2a.5 Target population gender: Female, Male</p> <p>2a.6 Target population age range: 18-90</p> <p>2a.7 Denominator Time Window (<i>The time period in which cases are eligible for inclusion in the denominator</i>): Within 30 days from admission</p> <p>2a.8 Denominator Details (<i>All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions</i>): Adult patients admitted for one of the procedures in the General Surgery, Orthopedic or Vascular DRGs (see Appendix A http://www.research.chop.edu/programs/cor/outcomes.php) who developed an in hospital complication and those who died without a complication.</p>
<p>2a.9 Denominator Exclusions (<i>Brief text description of exclusions from the target population</i>): Patients over age 90, under age 18.</p> <p>2a.10 Denominator Exclusion Details (<i>All information required to collect exclusions to the denominator, including all codes, logic, and definitions</i>): N/A</p>
<p>2a.11 Stratification Details/Variables (<i>All information required to stratify the measure including the stratification variables, all codes, logic, and definitions</i>): Complicated patient has at least one of the complications defined in Appendix B (http://www.research.chop.edu/programs/cor/outcomes.php) Complications are defined using the secondary ICD9 diagnosis and procedure codes and the DRG code of the current admission. When Physician Part B file is available, the definition of complications and comorbidities are augmented to include CPT codes.</p>
<p>2a.12-13 Risk Adjustment Type: Risk-adjustment devised specifically for this measure/condition</p> <p>2a.14 Risk Adjustment Methodology/Variables (<i>List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method</i>): Risk Adjustment: Model was developed using logistic regression analysis.</p> <p>Associated data elements: age in years, sex, race, comorbidities, DRGs (combined with and without complications) and procedure codes within DRGs, transfer status.</p> <p>Failure to rescue is adjusted using a logistic regression model where y is a failure and the total N is composed of patients who develop a complication and patients who died without a complication.</p> <p>According to developer: The model adjustment variables can vary. We have found that FTR results are fairly stable, even with little adjustment, since all patients in an FTR analysis have developed a complication (by definition), they are a more homogeneous group of patients than the entire population. Hence severity adjustment plays somewhat less of a role than in other outcome measures.</p>
<p>2a.15-17 Detailed risk model available Web page URL or attachment: URL http://www.research.chop.edu/programs/cor/outcomes.php</p>
<p>2a.18-19 Type of Score: Rate/proportion</p> <p>2a.20 Interpretation of Score: Better quality = Lower score</p> <p>2a.21 Calculation Algorithm (<i>Describe the calculation of the measure as a flowchart or series of steps</i>): Refer to website (http://www.research.chop.edu/programs/cor/outcomes.php)</p>
<p>2a.22 Describe the method for discriminating performance (e.g., significance testing): T-test for comparing rates</p>
<p>2a.23 Sampling (Survey) Methodology <i>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):</i></p>

<p>Measure not based on sample, all surgical patients between the ages of 18 and 90 admitted to an acute care hospital.</p> <p>2a.24 Data Source (Check the source(s) for which the measure is specified and tested) Administrative claims</p> <p>2a.25 Data source/data collection instrument (Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.): Linked patients hospitalizations claims records, augmented with Outpatient and Part B records; can also use unlinked data if linked files are not available to identify comorbidities and develop definitions of severity and other risk measure.</p> <p>2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL http://www.resdac.org/</p> <p>2a.29-31 Data dictionary/code table web page URL or attachment: URL http://www.research.chop.edu/programs/cor/outcomes.php</p> <p>2a.32-35 Level of Measurement/Analysis (Check the level(s) for which the measure is specified and tested) Facility, Health Plan, Integrated Delivery System, Population : County or City, Population : National, Population : Regional, Population : State</p> <p>2a.36-37 Care Settings (Check the setting(s) for which the measure is specified and tested) Hospital/Acute Care Facility</p> <p>2a.38-41 Clinical Services (Healthcare services being measured, check all that apply) Clinicians: Physicians (MD/DO)</p>	
TESTING/ANALYSIS	
<p>2b. Reliability testing</p> <p>2b.1 Data/sample (description of data/sample and size): Medicare inpatient claims for general surgical admissions for the period July 1, 1999 to June 30, 2000. There were a total of 1467 hospitals and 403,679 patients. We included patients between 65 and 90 years of age.</p> <p>2b.2 Analytic Method (type of reliability & rationale, method for testing): We defined reliability as described by Lord and Novick using split sample methodology. (Lord FM, Novick MR. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley; 1968)</p> <p>2b.3 Testing Results (reliability statistics, assessment of adequacy in the context of norms for the test conducted): Using Spearman-Brown half split half sample reliability had a correlation of 0.32 and the upper bound on validity was 0.56.</p>	<p>2b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>2c. Validity testing</p> <p>2c.1 Data/sample (description of data/sample and size): Medicare inpatient claims for general surgical admissions for the period July 1, 1999 to June 30, 2000. There were a total of 1467 hospitals and 403,679 patients. We included patients between 65 and 90 years of age.</p> <p>2c.2 Analytic Method (type of validity & rationale, method for testing): a) Rank correlation between various hospital outcomes (Death, Failure to Rescue, Complications, other measures of Failure to Rescue, Failure to Rescue Complement measures) b) Marginal and partial coefficients in logit models using detailed patient characteristics and hospital characteristics shown to be associated with better outcomes in previous studies.^{2, 7} The marginal results use one hospital characteristic at a time along with all patient characteristics. “Partial” regression results, using all hospital and patient variables simultaneously have the disadvantage that correlation between</p>	<p>2c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

<p>hospital characteristics can cause difficulty in interpreting the effects of individual hospital variables. Hospital characteristics associated with better outcomes (1) teaching hospital status (member of the American Council of Teaching Hospitals); (2) high technology status (does the hospital perform open heart surgery or perform organ transplantation); (3) hospital size greater than 200 beds; (4) bed-to-nurse ratio (where nurses are the sum of RN plus LPN FTE positions); and (5) nursing skill mix (the ratio of RN/(RN+LPN)).2-8</p> <p>c) The relative contribution of patient-to-hospital characteristics that predicted each outcome of interest, as provided by the omega statistic.^{2, 9} The omega statistic computes a ratio of the squared sum of the log odds for model patient variables divided by a similar quantity calculated for the model hospital variables. All else being equal, outcome measures that have lower omega ratios may be more desirable quality indicators, since the lower the omega, the greater the hospital's impact on outcome relative to the patient's impact. This is especially important if modeling patient severity is difficult (as with claims data) so that the lower the omega suggests the higher relative influence of hospital characteristics as compared to patient.</p> <p>2c.3 Testing Results (<i>statistical results, assessment of adequacy in the context of norms for the test conducted</i>): FTR itself is highly correlated with death, with a Kendall's tau equal to 0.83, representing a probability of concordance equal to 0.91.</p>	
<p>2d. Exclusions Justified</p> <p>2d.1 Summary of Evidence supporting exclusion(s): Patients younger than 18 are excluded because they are considered in the pediatric population and have a different set of complications. We use 90 years as a cut-point because of our concern regarding the increased use of do-not-resuscitate at higher ages [Wenger et al. Epidemiology of Do-Not Resuscitate Orders. Disparity by Age, Diagnosis, Gender, Race, and Functional Impairment. Arch Intern Med. 1995; 155(19):2056-62, Hakim et al. Factors Associated with Do-Not-Resuscitate Orders: Patients', Preferences, Prognoses, and Physicians Judgments. Ann Intern Med.1996; 125:284-293.]. While we do adjust for admission severity when reporting FTR, and this includes age, we still thought it prudent to use an upper bound on age, since DNR status prior to the procedure is not well defined at hospitals [Tabak YP, Johannes RS, Silber JH, Kurtz SG, Gibber EM. Should do-not-resuscitate status be included as a mortality risk adjustor? The impact of DNR variations on performance reporting. Med Care 2005; 43:658-666]</p> <p>2d.2 Citations for Evidence: 1. Wenger NS, Pearson ML, Desmond KA, Harrison ER, Rubenstein LV, Rogers WH, Kahn KL. Epidemiology of Do-Not Resuscitate Orders. Disparity by Age, Diagnosis, Gender, Race, and Functional Impairment. Arch Intern Med. 1995; 155(19):2056-62 2. Hakim RB, Teno JM, Harrell Jr. FE, Knaus WA, Wenger N, Phillips RS, Layde P, Califf R, Connors Jr. AF, Lynn J. Factors Associated with Do-Not-Resuscitate Orders: Patients', Preferences, Prognoses, and Physicians Judgments. Ann Intern Med. 1996; 125:284-293. 3. Tabak YP, Johannes RS, Silber JH, Kurtz SG, Gibber EM. Should do-not-resuscitate status be included as a mortality risk adjustor? The impact of DNR variations on performance reporting. Med Care 2005; 43:658-666</p> <p>2d.3 Data/sample (<i>description of data/sample and size</i>): N/A</p> <p>2d.4 Analytic Method (<i>type analysis & rationale</i>): N/A</p> <p>2d.5 Testing Results (<i>e.g., frequency, variability, sensitivity analyses</i>): N/A</p>	<p>2d</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>2e. Risk Adjustment for Outcomes/ Resource Use Measures</p> <p>2e.1 Data/sample (<i>description of data/sample and size</i>): Two different data samples were used to analyze risk adjustment. 1.) 5,972 Medicare patients undergoing elective cholecystectomy or transurethral prostatectomy (Silber et al. Hospital and Patient Characteristics Associated with Death After Surgery A study of Adverse Occrueenece and Failure to Rescue Med Care 1992). 2.) 2,021,214 patients with medicare claims on general, orthopedic, and vascular surgery admissions in the</p>	<p>2e</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>

<p>United States for 2000-2005. (Silber et al. Hospital Teaching Intensity, Patient Race, and Surgical Outcomes Arch Surg 2009)</p> <p>2e.2 Analytic Method (<i>type of risk adjustment, analysis, & rationale</i>): Risk Adjustment: Model was developed using logistic regression analysis, where y is a failure and the total N is composed of patients who develop a complication and patients who died without a complication.</p> <p>2e.3 Testing Results (<i>risk model performance metrics</i>): In earlier work we did report calibration as tested with the Hosmer-Lemeshow statistic, however the research community found that this calibration test fails its asymptotics, it overcalls with large sample size, we do not recommend its use. It is well known that the Hosmer-Lemeshow test is misleading with large data sets, and therefore we have not thought this to be a valid approach. C-statistic ranges 0.70 for the FTR 30 day risk adjustment model (Silber et. al Med Care 1992) to 0.792 (Silber et al. Arch Surg 2009). However c-statistics are also misleading when comparing across populations. Since FTR is a subset of the mortality and complication data set, one cannot compare, in a meaningful way, the c-statistic from FTR to that of mortality or complication.</p> <p>2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: N/A</p>	
<p>2f. Identification of Meaningful Differences in Performance</p> <p>2f.1 Data/sample from Testing or Current Use (<i>description of data/sample and size</i>): Medicare inpatient claims for general surgical admissions for the period July 1, 1999 to June 30, 2000. There were a total of 1467 hospitals and 403,679 patients. We included patients between 65 and 90 years of age.</p> <p>2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (<i>type of analysis & rationale</i>): T-test for comparing rates.</p> <p>2f.3 Provide Measure Scores from Testing or Current Use (<i>description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance</i>): 75% Q3 = 0.16, Median= 0.12, 25% Q1 =0.09, Mean= 0.13, Std Deviation =0.05.</p>	<p>2f C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>2g. Comparability of Multiple Data Sources/Methods</p> <p>2g.1 Data/sample (<i>description of data/sample and size</i>): FTR was developed using standardized hospital discharge records, which are widely collected by states agencies and which hospitals are mandated to report to CMS. One of the big advantages of adopting FTR is that the data on which it is based is uniformly reported, checked for errors and edited. This is administrative data available for the entire population over 65 and for all patients admitted to acute care hospitals.</p> <p>2g.2 Analytic Method (<i>type of analysis & rationale</i>): N/A</p> <p>2g.3 Testing Results (<i>e.g., correlation statistics, comparison of rankings</i>): N/A</p>	<p>2g C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>2h. Disparities in Care</p> <p>2h.1 If measure is stratified, provide stratified results (<i>scores by stratified categories/cohorts</i>): Disparities in care are shown in Silber et al Arch Surg 2009 where the results show white patients displayed a reduction in failure-to-rescue rates in the teaching intensive hospitals vs non-teaching hospitals (OR, 0.94; 95% CI, 0.92-0.97), black patients displayed an increased failure-to-rescue rate (OR, 1.06; 95% CI, 1.00-1.12)</p> <p>2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: Failure to Rescue can be used to detect disparities in health outcomes across providers, shown in Silber et al. Arch Surg 2009.</p>	<p>2h C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Scientific</p>	<p>2</p>

Acceptability of Measure Properties?	
Steering Committee: Overall, to what extent was the criterion, <i>Scientific Acceptability of Measure Properties</i> , met? Rationale:	2 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
3. USABILITY	
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
3a. Meaningful, Understandable, and Useful Information	
<p>3a.1 Current Use: In use</p> <p>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). <u>If not publicly reported</u>, state the plans to achieve public reporting within 3 years): FTR information is online for the public to access (http://stokes.chop.edu/programs/cor/outcomes.php). Consumers can access FTR results through the multiple research publications on the measure. In the future FTR could be reported on a wider scale, the same way that mortality rates are reported.</p> <p>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). <u>If not used for QI</u>, state the plans to achieve use for QI within 3 years): Currently used to assess the impact of the change in the resident work hours regulations on patient outcomes in a recently NHLBI funded study (1R01HL094593-01) entitled "Work Hour Regulation for Physician Trainees: Educational and Clinical Outcomes"</p> <p>Testing of Interpretability (Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement)</p> <p>3a.4 Data/sample (description of data/sample and size): In Ghaferi et al "Variation in Hospital Mortality Associated with Inpatient Surgery" studied 84,730 patients who had undergone inpatient general and vascular surgery from 2005-2007 using data from the American College of Surgeons National Surgical Quality Improvement Program.</p> <p>3a.5 Methods (e.g., focus group, survey, QI project): Ranked ranked hospitals according to their risk-adjusted overall rate of death and divided them into five groups. For hospitals in each overall mortality quintile, we then assessed the incidence of overall and major complications and the rate of death among patients with major complications (failure to rescue rate).</p> <p>3a.6 Results (qualitative and/or quantitative results and conclusions): Rates of death varied widely across hospital quintiles, from 3.5% in very-low-mortality hospitals to 6.9% in very-high-mortality hospitals. Hospitals with either very high mortality or very low mortality had similar rates of overall complications (24.6% and 26.9%, respectively) and of major complications (18.2% and 16.2%, respectively). Rates of individual complications did not vary significantly across hospital mortality quintiles. In contrast, mortality in patients with major complications was almost twice as high in hospitals with very high overall mortality as in those with very low overall mortality (21.4% vs. 12.5%, P<0.001). Differences in rates of death among patients with major complications were also the primary determinant of variation in overall mortality with individual operations. In addition to efforts aimed at avoiding complications in the first place, reducing mortality associated with inpatient surgery will require greater attention to the timely recognition and management of complications once they occur.</p>	
3b/3c. Relation to other NQF-endorsed measures	
<p>3b.1 NQF # and Title of similar or related measures: 0200 Death among surgical inpatients with treatable serious complications (failure to rescue)</p> <p>(for NQF staff use) Notes on similar/related endorsed or submitted measures:</p>	
3b. Harmonization	3b

<p>If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source <u>or</u> different topic but same target population):</p> <p>3b.2 Are the measure specifications harmonized? If not, why?</p>	<p>C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>3c. Distinctive or Additive Value</p> <p>3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:</p> <p>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:</p> <p>Needleman et al adapted the FTR measure to “nurse sensitive complications” by selecting a limited number of complications for the FTR measure. This change in definition, which we will call FTR-N, was developed to better focus on nursing quality of care. Because only deaths after nursing sensitive complications are studied, a large number of deaths are not used in the analysis. Subsequently, AHRQ again adapted the FTR-N definition to reflect quality from a “patient safety” perspective (ie, the identification of deaths that were especially likely to be preventable). Expert panels guided both of these adaptations through consensus development panels. The National Quality Forum, through its own process of selecting National Voluntary consensus Standards for Nursing-Sensitive Care, endorsed Needleman et al’s adaptation and assigned it to AHRQ for updating and support.FTR-N includes only 6 complications (pneumonia, shock, gastrointestinal bleeding, cardiac arrest, sepsis, and deep venous thrombosis) in its denominator definition, and it excludes deaths in patients without these complications. FTR-A adds renal failure to the FTR-N list of eligible complications, and modestly alters the definition of several others Table 1C and 1D display the impact of restricting the denominator of FTR to more limited sets of complications, as in the FTR-N and FTR-A definitions, respectively. Note first that the number of patients defined as having a complication fell from 189,031 (46.8%) in Table 1A to 43,500 (10.8%) in Table 1C and 39,101 (9.7%) in Table 1D. However, this smaller complication rate comes at an important cost—of all deaths, the proportion coded as having a complication (the precedence rate) fell from 95% in Table 1A to only 51% in Table 1C, and 58.5% in Table 1D. (Refer tp Silber et al. Med Care 2007)</p>	<p>3c</p> <p>C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Usability</i>?</p>	<p>3</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Usability</i>, met?</p> <p>Rationale:</p>	<p>3</p> <p>C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>4. FEASIBILITY</p>	
<p>Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)</p>	<p>Eval Rating</p>
<p>4a. Data Generated as a Byproduct of Care Processes</p> <p>4a.1-2 How are the data elements that are needed to compute measure scores generated?</p> <p>Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)</p>	<p>4a</p> <p>C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>4b. Electronic Sources</p> <p>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)</p> <p>Yes</p> <p>4b.2 If not, specify the near-term path to achieve electronic capture by most providers.</p>	<p>4b</p> <p>C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

<p>4c. Exclusions</p> <p>4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications? No</p> <p>4c.2 If yes, provide justification.</p>	<p>4c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences</p> <p>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. FTR is given to minimal susceptibility to inaccuracies or errors since it uses data collected uniformly across all hospitals and providers. The data is carefully checked by CMS before it is being released to researchers. However there may be unobserved differences among patients due to the lack of more detailed clinical information available only through chart abstraction.</p>	<p>4d C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>4e. Data Collection Strategy/Implementation</p> <p>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: We have developed FTR measures based on restricted information, available only from the inpatient files. When possible, such as in the Medicare population, we improve the risk adjustment by using more patient level information available in the outpatient or Carrier file</p> <p>4e.2 Costs to implement the measure (<i>costs of data collection, fees associated with proprietary measures</i>): CMS data is made available to researchers through ResDac, and its cost depends on the number of records requested, the number of years, and the type of file (inpatient, outpatient, or carrier)</p> <p>4e.3 Evidence for costs: N/A</p> <p>4e.4 Business case documentation: N/A</p>	<p>4e C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Feasibility?</p>	<p>4</p>
<p>Steering Committee: Overall, to what extent was the criterion, Feasibility, met? Rationale:</p>	<p>4 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
RECOMMENDATION	
<p>(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.</p>	<p>Time-limited <input type="checkbox"/></p>
<p>Steering Committee: Do you recommend for endorsement? Comments:</p>	<p>Y <input type="checkbox"/> N <input type="checkbox"/> A <input type="checkbox"/></p>
CONTACT INFORMATION	
<p>Co.1 Measure Steward (Intellectual Property Owner) Co.1 <u>Organization</u></p>	

The Children ´ s Hospital of Philadelphia, 34th St. and Civic Center Blvd., Philadelphia, Pennsylvania, 19104
Co.2 Point of Contact Jeffrey, Silber, PhD, MD, silber@email.chop.edu , 215-590-2540-
Measure Developer If different from Measure Steward Co.3 Organization The Children ´ s Hospital of Philadelphia, 34th St. and Civic Center Blvd., Philadelphia, Pennsylvania, 19104
Co.4 Point of Contact Orit, Even-Shoshan, MS, shoshan@email.chop.edu , 215-590-2809-
Co.5 Submitter If different from Measure Steward POC Orit, Even-Shoshan, MS, shoshan@email.chop.edu , 215-590-2809-, The Children ´ s Hospital of Philadelphia
Co.6 Additional organizations that sponsored/participated in measure development N/A
ADDITIONAL INFORMATION
Workgroup/Expert Panel involved in measure development Ad.1 Provide a list of sponsoring organizations and workgroup/panel members’ names and organizations. Describe the members’ role in measure development. A group of clinicians and coding experts from the University of Pennsylvania reviewed the updated ICD, CPT, and DRG codes and updated the measure to reflect current coding.
Ad.2 If adapted, provide name of original measure: N/A Ad.3-5 If adapted, provide original specifications URL or attachment
Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 Year the measure was first released: Ad.7 Month and Year of most recent revision: Ad.8 What is your frequency for review/update of this measure? Ad.9 When is the next scheduled review/update for this measure?
Ad.10 Copyright statement/disclaimers:
Ad.11 -13 Additional Information web page URL or attachment: URL http://www.research.chop.edu/programs/cor/outcomes.php
Date of Submission (MM/DD/YY): 06/08/2011