NATIONAL QUALITY FORUM

Memo

TO: Surgery Standing Committee

FR: Kathryn Streeter, Senior Project Manager and Christy Skipper, Project Manager

RE: Post-Comment Call to Discuss Public and Member Comments

DA: November 2, 2016

Purpose of the Call

The Surgery Standing Committee will meet via conference call on Monday, November 7, 2016 from 3:00-5:00pm ET. The purpose of this call is to:

- Review and discuss comments received during the post-evaluation public and member comment period.
- Re-vote on two measures that did not reach consensus on a recommendation by the Committee.
- Provide input on proposed responses to the post-evaluation comments.
- Determine whether reconsideration of any measures or other courses of action is warranted.

NQF staff has drafted responses to the comments. Committee members should review all comments and draft responses prior to the call.

Standing Committee Actions

1. Review this briefing memo and Draft Report.
2. Be prepared to reconsider and decide whether to revote on measure *#0351 Death Rate Among Surgical Inpatients with Serious Treatable Complications (PSI 04)* as requested by the developer. The complete measure worksheet is provided in Appendix A.
3. Review and re-discuss the two measures where consensus was not achieved in order to reach consensus.
4. Review and consider the full text of all comments received and the proposed responses to the post-evaluation comments (see Comment Table).
5. Be prepared to provide feedback and input on proposed post-evaluation comment responses.

Conference Call Information

Please use the following information to access the conference call line and webinar:
Speaker dial-in #: 866.599.6630
Public dial-in #: 855.599.0737
Public Web Link: http://nqf.commpartners.com/se/Rd/Mt.aspx?563280
Committee Web Link: Please use the link sent to your email by CommPartners

***In order to vote, Committee members should use their individual webinar links sent via email.**

Background

The Surgery portfolio is one of NQF's largest measure portfolio with over 100 endorsed measures. This NQF project aimed to evaluate additional performance measures that will help guide cardiac, vascular, orthopedic, urologic, and gynecologic surgeries that include adult and pediatric population. The 25-member Surgery Standing Committee met for a 2-day in-person meeting to evaluate 24 measures: 10 new measures and 14 measures undergoing maintenance review. The Committee recommended 14 measures for endorsement; eight were not recommended; and consensus was not reached for two measures.

Comments Received

NQF solicits comments on measures undergoing review in various ways and at various times throughout the evaluation process. First, NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). Second, NQF solicits member and public comments prior to the evaluation of the measures via an online tool located on the project webpage. Third, NQF opens a 30-day comment period to both members and the public after measures have been evaluated by the full committee and once a report of the proceedings has been drafted.

Pre-evaluation comments

The pre-evaluation comment period was open from June 1 to June 6, 2016 for all 24 measures under review. One pre-evaluation comment was received and was in favor of endorsement of a measure submitted to the portfolio. The pre-evaluation comment was provided to the Committee prior to their initial deliberations.

Post-evaluation comments

The Draft Report went out for the 30-day Public and Member commenting period on September 22 – October 21, 2016. During this commenting period, NQF received 71 comments from 6 member organizations and 11 comments from the public. Comments were generally supportive of the Committee's recommendations.

In order to facilitate discussion, the majority of the post-evaluation comments have been categorized into major topic areas or themes. Where possible, NQF staff has proposed draft responses for the Committee to consider. Although all comments and proposed responses are subject to discussion, we will not necessarily discuss each comment and response on the post-comment call. Instead, we will spend the majority of the time considering the major topics and/or those measures with the most significant issues that arose from the comments. Note that the organization of the comments into major topic areas is not an attempt to limit Committee discussion.

We have included all of the comments that we received (both pre- and post-evaluation) in the comment table. This comment table contains the commenter's name, comment, associated measure, topic (if applicable), and—for the post-evaluation comments—draft responses for the Committee's consideration. Please refer to this comment table to view and consider the individual comments received and the proposed responses to each.

One comment was submitted after the commenting period closed on measure #3024 Carotid Endarterectomy Evaluation of Vital Status and NIH Stroke Scale at Follow Up:

The NIH Stroke Scale (NSS) is not only a validated outcome measure for stroke [PMID:26359360], but also allows an understanding of the clinical impact of each stroke rather

than artificially treating each stroke as identical. For this reason the American Association of Neurological Surgeons (AANS) has always championed the use of the NSS as a tool for assessing stroke severity and believes that it should be included in upcoming National Quality Forum consensus development processes for carotid endarterectomy. That being said, we agree with the NQF critique of the proposed measure that stated it "would be stronger if [it] was using the NIH stroke scale to measure an actual outcome within 30 or 60 days post discharge as opposed to the process of administering the tool." Therefore, we support measure #3024 while allowing that it has room for improvement.

Reconsideration Request

#0351: Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04)

The developer has requested that this measure be reconsidered by the Committee.

During the in-person meeting, the Committee did not pass the measure on Validity. In regard to Validity, the Committee was concerned that the specification that excludes patients who are transferred from the denominator could provide a method to game the measure, particularly if the patient's condition worsens. The Committee noted that transferring a patient to a higher level of care is often appropriate, however, when a patient is transferred to a higher level of care, the receiving hospital becomes responsible for the treatment provided by the transferring hospital. The Committee did not believe risk adjustment fully accounted for the fact that the receiving hospital would become responsible for the patient's outcome. The Committee also questioned whether transfer issues were addressed adequately to understand threats to validity and that handling of transfers make it difficult to validate if the appropriate efforts were made to save a patient. (See Appendix A).

Developer Comment: We are writing to request that the National Quality Forum (NQF) Surgery Standing Committee reconsider the decision to remove endorsement of Death Rate Among Surgical Inpatients with Serious Treatable Complications (PSI 04), (NQF 0351). This long-standing Patient Safety Indicator (PSI) has been endorsed by NQF since 2008. Our request for reconsideration is based on concern that NQF's standard review process was not applied properly during the in-person meeting on August 16, 2016, particularly with respect to the following:

1) Appropriate review and evaluation of the measure for Criteria 2. Scientific Acceptability Sub-criteria 2a. Validity

2) Discussion of the use case of the measure prior to full discussion of the scientific acceptability for the measure

3) Consistent evaluation of related (not competing) measures across NQF standing committees

First, according to the NQF's Guidance for Evaluating Validity and as noted by Dr. Karen Johnson during the review, measure developers need only submit validity testing with respect to computed performance measures scores, not data element validity. AHRQ submitted

information about construct validity, which should have been the focus of the validity discussion, not the detailed discuss of claims data and data element validity.

Second, although AHRQ acknowledges the difficultly of conducting reviews that are use-agnostic, the reviewers brought up concerns about the use of the measure by CMS during scientific acceptability discussions. It is AHRQ's understanding the NQF seeks to endorse measures that are deemed scientifically rigorous and suitable for not just quality improvement but also general accountability purposes (not specific accountability purposes). The NQF review process is intended to be use-agnostic. Specific use cases of the measure, particularly the appropriate use of the measures in CMS programs, are to be discussed during NQF's Measure Application Project committee meetings.

Third, while acknowledged in the introduction of the measure, NQF's re-endorsement of a related measure by the Patient Safety Standing Committee was not emphasized during the review discussions. In particular, in the course of that re-endorsement discussion for NQF 0352 (Failure to Rescue In-Hospital Mortality, risk adjusted), which was developed and is stewarded by the Children's Hospital of Philadelphia, the Patient Safety Committee carefully evaluated the design of "failure to rescue" measures. This Committee discussed and accepted the developer's evidence-based arguments in favor of including patients who had reported complications present on admission in the measure denominator. When different NQF Standing Committees fail to evaluate similar measures, with similar design features, in a consistent manner, the consequences include confusion across the stakeholder community and mixed messages to measure developers, stewards, and users.

In addition, as noted in the NQF-Endorsed Measures for Surgical Procedures 2015-2017: Draft Report for Comment (September 22, 2016), reviewers wanted additional information about transfers, risk adjustment and use of claims data to measure complications.

AHRQ respectfully requests that NQF ask that the Committee exercise the option to re-vote on the validity of the measure during the post-comment call to preserve the integrity of the NQF process, and consider the additional information being submitted by AHRQ.

> **Action Item:** Based on the information provided by the developer, would the Committee like to reconsider this measure?
>
> **Action Item:** If the Committee chooses to reconsider this measure, the Committee will re-vote on Validity. If the measure passes the Validity subcriterion, the Committee will vote on Feasibility, Usability and Use, and on overall suitability for endorsement.

"Consensus Not Reached" Measures

1543: Postoperative Stroke or Death in Asymptomatic Patients undergoing Carotid Artery Stenting (CAS)

During the in-person meeting, the Committee noted that there were no published guidelines for carotid artery stenting and that this procedure is not recommended by the major medical

societies. Committee members also questioned whether the measure should be considered an appropriate use measure due to the increased risk of stroke or death, compared to the risk of stroke or death by surgery. Other Committee members stated that despite indication, the procedure is still being performed and therefore it would be important to measure the outcome. The Committee should review the comments that were received, and then re-discuss the measure. During the in-person meeting, consensus was not reached on the Evidence (Y-12; N-10) and Validity (H-0; M-13; L-9; I-0) subcriteria

Thirteen comments were submitted for this measure; all were supportive comments, noting that the measure will aid in the appropriate selection of patients who receive carotid stenting. One comment stated that "review of the medical literature that was discussed…omitted significant amounts of data and that it was inaccurate in summarizing the clinical guidelines on carotid stenting." See Appendix B or the Comment table (ID 6370) for the full comment.

> **Developer Response:** In response to Dr. Powell's comments: As Chair of the SVS Quality Performance and Measures Committee and one of the presenters of this measure that day we are not against CAS. As I stated that day CEA and CAS have similar outcomes with experience operators. We are advocating this measure be continued so we can monitored outcomes and compared them with CEA. There have been studies that suggest some subsets of patients may have different outcomes yet ongoing research will define this . I strongly encouraged the NQF Committee that day to continue this measure. Brad Johnson, MD SVS Chairman of Quality Performance and Measures Committee
>
> **Proposed Committee Response:** Response pending Committee discussion.
>
> **Action Item:** The Committee will re-vote on the Evidence and Validity subcriteria, and, if the measure passes, it will vote on overall suitability for endorsement.

3020: PBM-04 Initial Transfusion Threshold

This new eMeasure is being considered for trial approval status. During the in-person meeting, Committee members expressed several concerns over the specifications of this measure. Members noted that there are other indications for a transfusion besides a hemoglobin measurement, such as hemorrhagic shock, bleeding, and current active bleeding, which are not reported as part of the measure. Committee members also suggested expanding the numerator to include a category for patients whose hemoglobin levels were not measured prior to a transfusion, include pediatric patients, and exclude pregnant patients undergoing postpartum hemorrhage. The Committee should review the comments that were received, and then re-discuss the measure. During the in-person meeting, consensus was not reached on the Scientific Acceptability: eMeasure Trial Measure Specifications (H-1; M-7; L-9; I-2) or feasibility (H-3; M-6; L-6; I-2) criteria.

The measure received 2 comments. One comment was not in support of the measure and one comment was submitted by the developer (see below).

Developer Comment: The Joint Commission respectfully disagrees with the draft report's findings that consensus was not reached on the eMeasure specifications for this measure. In fact, the eMeasure specifications were never discussed. NQF criteria for Approval for Trial Use indicate that a candidate eMeasure must meet "all criteria under Importance to Measure and Report (clinical evidence and opportunity for improvement/ performance gap)", "the eMeasure feasibility assessment must be completed", and that "results from testing with a simulated (or test) data set demonstrate that the QDM and HQMF are used appropriately and that the measure logic performs as expected". These criteria were clearly met by this measure, although the latter two were not discussed during the Standing Committee's deliberations. The objections of the committee and subsequent failure to reach consensus were based on perceived issues with measure validity, a component of Scientific Acceptability review, and therefore should have been outside the scope of this eMeasure review. The Joint Commission strongly recommends that NQF re-examine its processes for evaluating eMeasures submitted for Approval for Trial Use, and clarifies for Standing Committees and staff, as well as measure developers, the appropriate scope of this review.

NQF Response: The Approval for Trial Use program was designed by NQF to facilitate the development of innovative quality eMeasures that could fill existing gaps in clinical care. The NQF requirements for endorsement with respect to an eMeasure require testing in at least two separate electronic health record (EHR) systems. This is in addition to the measures being specified according to the Health Quality Measures Format (HQMF) and aligning with the Quality Data Model (QDM) as well as having value sets published within the Value Set Authority Center (VSAC). NQF recognizes that for some measures, these requirements, particularly in identifying two EHRs to test in, may be challenging. However, NQF does not want to impede the progress of needed measures and thus the Trial Use program allows for the measure to be implemented into the field in which data can be collected and evaluated. Once enough data has been gathered, the measure can then be properly assessed and submitted to a committee for endorsement consideration.

However, a measure for Trial Use consideration is evaluated in the same way as a measure being considered for endorsement. The measure must be scientifically acceptable, and must have a strong evidence base for consideration. The only difference is in the testing itself, in that a measure for Trial Use consideration only has to submit BONNIE results to demonstrate that the measure logic works as intended and that the metric produced by the measure match its objective. A committee that is evaluating a Trial Use measure will still consider its scientific acceptability and importance to measure. If the measure passes those criteria, and the BONNIE testing indicates that the measure functions as it should, then it would be considered as part of the Trial Use program. However, if the committee does not feel that the measure demonstrates importance to measure and collect; and/or does not meet the scientific acceptability criteria, then it may be rejected, as any other measure would. A measure for Trial Use is evaluated in the same manner as a measure for endorsement, with the exception being on the testing of the measure and, if the committee accepts the

measure, it is placed into the Trial Use program instead of being endorsed. A eMeausure for Trial Use consideration is not evaluated solely on the basis of its technical specifications.

Action Item: The Committee will re-vote on the eMeasure Trial Measure Specifications sub-criterion, and, if it passes this subcriterion, it will vote on a recommendation for trial approval.

Comments and their Disposition

Two major themes were identified in the post-evaluation comments, as follows:

1. Comments in support of Committee's recommendations
2. Evaluation and discussion of the Sociodemographic Status trial period

Theme 1 – Comments in support of Committee's recommendations

A vast majority of the comments (68) were in support of the Committee's recommendations on all measures for which consensus was reached. Of these, 13 comments emphasized the importance of *#1534 In-hospital mortality following elective EVAR of AAAs*.

Action Item: None

Theme 2 – Evaluation and discussion of the Sociodemographic Status trial period

Two comments expressed concern related to the "lack of rigor and robustness of the risk adjustment reviews" and suggested that other SDS factors must be considered "to understand the potential impact on a hospital's performance".

NQF Response: The Committee has reviewed your comment and appreciates your input. The SDS trial period is a temporary change to NQF's policy. During this 2-year trial period, NQF is gathering information about the feasibility, limitations and challenges of including SDS factors in the risk-adjustment approach. Consideration of sociodemographic factors in risk adjustment models is a critical issue in measurement science. The Committee was charged with evaluating the measure specifications and testing submitted on the measure as developed by the measure developer. Given the constraints on the current data elements available, the Committee relied on the methods used by the developer to test the conceptual and empirical relationship between SDS factors and readmissions and complications. The Committee recognizes that risk adjustment for SDS factors is a rapidly progressing area and that more work is needed to appreciate the effects of social risk, understand the most relevant patient and community level factors, collect data on these risk factors, and determine the best methods to incorporate these risk factors into performance measures. The Committee looks forward to continued deliberations on these issues and to reexamining these measures as better data emerges.

In their deliberation of these measures, the developer stated there was a conceptual relationship between the SDS variables. Specifically, the developer reported socially disadvantaged patients had a higher disease burden or receive worse or disparate care, and that there is a source unrelated to hospital quality of care which could hinder patient adherence to things like post-discharge instructions. Using decomposition analysis to measure effects at the patient level and at the hospital level, the developers reported that the hospital effect dominated with SDS variables. These results were in contrast to the clinical data elements where the patient effect tended to dominate. Due to the dominant hospital effect, the developers reported that they risk adjust away a component of hospital quality when the variables are included in the model. The developer reported that the three SDS factors were statistically significant in the model and inclusion of the variables in the model did not change the c-statistic. Ultimately, the Committee agreed that they would not recommend risk adjustment for SDS since finding disparities among groups on these measures is something that should be reported and followed.

The Disparities Standing Committee has highlighted the ongoing challenges of risk adjustment for SDS factors. The Committee recently reviewed the newly released National Academy of Medicine report, "Accounting for Social Risk Factors in Medicare Payment: Data" that examined the availability of data on social risk factors. The report found that there are few factors currently available for use (e.g., dual eligibility, nativity, urbanicity/rurality) while other factors needed additional research for improved use or are not sufficiently available now. The availability of social risk factor data will continue to evolve and warrants ongoing monitoring. Please note that NQF has maintained a non-prescriptive approach to the selection and testing of variables included in risk adjustment models. NQF has not required that certain SDS variables be tested and does not set requirements around the inclusion of any specific variables. Similarly, NQF does not set "cut-points" for the statistical testing of a risk adjustment model. The evaluation of the model is left to the Standing Committee reviewing the measure. This approach applies to both clinical and SDS variables.

Finally, the report will be edited to reflect the Committee's full discussion on risk adjustment of these measures

Action Item: None.

Appendix A

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0351
De.2. **Measure Title:** Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04)
Co.1.1. **Measure Steward:** Agency for Healthcare Research and Quality
De.3. **Brief Description of Measure:** In-hospital deaths per 1,000 surgical discharges, among patients ages 18 through 89 years or obstetric patients, with serious treatable complications (shock/cardiac arrest, sepsis, pneumonia, deep vein thrombosis/ pulmonary embolism or gastrointestinal hemorrhage/acute ulcer). Includes metrics for the number of discharges for each type of complication. Excludes cases transferred to an acute care facility. A risk-adjusted rate is available. The risk-adjusted rate of PSI 04 relies on stratum-specific risk models. The stratum-specific models are combined to calculate an overall risk-adjusted rate.
1b.1. **Developer Rationale:** This indicator targets patients who are admitted for surgery who die following the development of a serious but treatable complication of care. Examples of such complications include: 1) shock or cardiac arrest, 2) sepsis, 3) pneumonia, 4) deep vein thrombosis or pulmonary embolism, and 5) gastrointestinal hemorrhage or acute ulcer. This indicator is fundamentally different than other PSIs, as it reflects the effectiveness of the hospital in rescuing a patient from complications versus preventing the underlying complications.

S.4. **Numerator Statement:** Number of deaths (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator.
S.7. **Denominator Statement:** Surgical discharges, for patients ages 18 through 89 years or MDC 14 (pregnancy, childbirth, and puerperium), with all of the following:
• any-listed ICD-9-CM or ICD-10-PCS procedure codes for an operating room procedure; and
• the principal procedure occurring within 2 days of admission or an admission type of elective (ATYPE=3); and
• meet the inclusion and exclusion criteria for STRATUM_SHOCK (shock or cardiac arrest), STRATUM_SEPSIS (sepsis), STRATUM_PNEUMONIA (pneumonia), STRATUM_DVT (deep vein thrombosis or pulmonary embolism), or STRATUM_GI_HEM (gastrointestinal hemorrhage or acute ulcer)
STRATUM_SHOCK (shock or cardiac arrest)
• any secondary ICD-9-CM or ICD-10-CM diagnosis codes or any-listed ICD-9-CM or ICD-10-PCS procedure codes for shock or cardiac arrest
STRATUM_SEPSIS (sepsis)
• any secondary ICD-9-CM or ICD-10-CM diagnosis codes for sepsis.
STRATUM_PNEUMONIA (pneumonia)
• any secondary ICD-9-CM or ICD-10-CM diagnosis codes for pneumonia or pneumonitis.
STRATUM_DVT (deep vein thrombosis or pulmonary embolism)
• any secondary ICD-9-CM or ICD-10-CM diagnosis codes for deep vein thrombosis or pulmonary embolism.
STRATUM_GI_HEM (gastrointestinal hemorrhage or acute ulcer)
• any secondary ICD-9-CM or ICD-10-CM diagnosis codes for gastrointestinal hemorrhage or acute ulcer. Surgical discharges are defined by specific MS-DRG codes and ICD-9-CM/ICD-10-PCS codes indicating "major operating room procedures."
S.10. **Denominator Exclusions:** Exclude cases:
• transferred to an acute care facility (DISP = 2)
• with missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing)

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Administrative claims
S.26. **Level of Analysis:** Facility

Original Endorsement Date: May 15, 2008 **Most Recent Endorsement Date:** Jan 31, 2012

IF this measure is included in a composite, NQF Composite#/title:
IF this measure is paired/grouped, NQF#/title:
De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- This maintenance measure, last reviewed in 2011, measures in-hospital deaths per 1,000 surgical discharges, among patients ages 18 through 89 years or obstetric patients, with serious treatable complications (shock/cardiac arrest, sepsis, pneumonia, deep vein thrombosis/ pulmonary embolism or gastrointestinal hemorrhage/acute ulcer).
- The developer provided rationale to support the relationship between swift diagnosis and intervention by experienced, skilled nurses and a reduced probability of death from the specific complications stating that better quality nurses and higher nurse staffing levels have been shown to improve PSI 04 rates and that these rates are strongly associated with risk-adjusted mortality rates but not with complication rates.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☒ **The developer provided updated evidence for this measure:**

Updates:

- Although not required, the developer provided clinical practice guidelines that focus on early recognition and aggressive treatment of sepsis, antithrombotic therapy for venous thromboembolism, rapid diagnosis and stabilization of patients with gastrointestinal hemorrhage, and perioperative cardiovascular risk assessment and management.
- The developer also conducted an environmental scan of the literature on failure to rescue (FTR) in February 2016, and provided an updated summary of the body of evidence found through this recent search.

Exception to evidence: N/A

Questions for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*
- *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee believe there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm Health outcome (Box 1) →relationship between outcome and at least one healthcare action identified/supported by stated rationale (Box 2) → Pass

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- In this submission, the developer notes a study that shows that due to improvements in care, between 1998 and 2007, FTR decreased by 6.05% per year (p<0.0001). However, the developer also notes that PSI 04 still captures approximately 43,000 deaths each year in the 34 states in the all-payer reference population. (Downey JR, Hernandez-Boussard T, Banka G, Morton JM. Is patient safety improving? National trends in patient safety indicators: 1998-2007. Health Serv Res. 2012;47(1 Pt 2):414-430.)
- The developer conducted all analyses using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) from calendar years 2011-2013.
- The observed rate for death rate has declined from 118.0419 in 2011 to 116.3869 in 2013. Distribution of hospital performance for death rate in 2-year pooled data showed a median of 102.61 in 2011-2012 and 99.92 in 2012-2013

Table 1. Risk adjusted death rate per 1,000 surgical discharge, 2013

Patient/hospital characteristic	Risk-adjusted estimate
Total U.S.	**116.387**
Age 18-39	74.420
Age 40-64	100.445
Age 65 and over	133.431
Male	120.100
Female	111.841
First quartile (lowest income)	121.085
Fourth quartile (highest income)	112.644
Medicare	116.361
Medicaid	120.407
Uninsured/self-pay/no-charge	136.964
Northeast	122.821
Midwest	110.965
South	118.210
West	115.587

Disparities

- The developer provided data (see Table 1 above) among 182,512 surgical patients with serious treatable conditions, older patients, men, those from lower income communities, those with Medicare, Medicaid or uninsured, and those treated in the Northeast were at greater risk of death (after controlling for a variety of clinical risk factors) than younger patients, women, those from higher income communities, those with private insurance and those treated in the Midwest or West.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *What does the data tell us about disparities for this condition?*

Preliminary rating for opportunity for improvement: ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a.

- Why exclude patients transferred to acute care? (can't track them in HCUP)
- Only reliable in larger hospitals (shrinkage adjustments used)
- consistent correlations with teaching status, advanced nursing staff, nursing ratio, high-tech (? what this means)
- Complex risk-adjustment per complication.
- this is an outcome measure of the number of deaths in surgical patients who develop serious treatable conditions postoperatively (shock, pneumonia, DVT, hemorrhage). It is linked to the quality of nursing care in evidence provided by the developer.
- The measure meets this criterion. In addition to the original supporting evidence, the developer provided clinical practice guidelines and an environmental scan of literature supporting early recognition and treatment and failure to rescue, both related to the outcome tracked by this measure.

1b.

- there still is a significant gap- 43,000 deaths per year in 34 measured states in an all payor dataset.
- older men and those with less well-paying or no insurance have higher mortality than the median rate for this measure.
- Although performance appears to be improving, the remaining gap is large enough to warrant a national performance measure. The results to appear to vary by some of the population subgroups provided, although I found myself desiring more context to understand this amount of variation vs. performance variation between/among facilities.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Administrative claims

- All analyses were completed using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), 2011-2013.

Specifications:

- The measure is specified as a facility-level measure for the hospital/acute care setting.
- The numerator includes the number of deaths (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator.
- The denominator includes the total number of surgical discharges, defined by specific MS-DRG codes and ICD-9-CM/ICD-10-PCS codes indicating "major operating room procedures," for patients ages 18 through 89 years or MDC 14 (pregnancy, childbirth, and puerperium), with all of the following:
  - any-listed ICD-9-CM or ICD-10-PCS procedure codes for an operating room procedure; and
  - principal procedure occurring within 2 days of admission or an admission type of elective (ATYPE=3); and
  - meet the inclusion and exclusion criteria for STRATUM_SHOCK (shock or cardiac arrest), STRATUM_SEPSIS (sepsis), STRATUM_PNEUMONIA (pneumonia), STRATUM_DVT (deep vein thrombosis or pulmonary embolism), or STRATUM_GI_HEM (gastrointestinal hemorrhage or acute ulcer)
- This outcome measure is risk adjusted, using a statistical risk model for each stratum.
- ICD-9 and ICD-10 codes, and a conversion table are provided.
- Denominator exclusions are cases that were transferred to an acute care facility (DISP = 2) and cases with missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing)
- The developer notes that annual updates are completed for all measures according to standard protocol. In addition, approximately every two years, AHRQ updates the risk-adjustment parameter estimates based on the

most recent year of data (i.e., the most current reference population possible). Therefore, since the last update, there have been some changes to the specifications in forthcoming 2016 version 6.0 specifications since the previously endorsed version (v4.4).

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Are 2016 specification changes reasonable for this measure?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- In the previous review, the developer conducted empirical analysis using AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million discharges, which the Committee found to be acceptable.

Describe any updates to testing

- Updates have been made to the reliability testing since the last review. In this submission, additional performance measure score reliability testing has been conducted using 2-year pooled data (2012-2013) from the HCUP State Inpatient Databases (SID). These data were obtained from 3,199 hospitals.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- The developer completed *performance measure score* testing using a signal-to-noise analysis, appropriate for this measure, which assesses differences in performance between hospitals ("the signal") to stability within hospitals (random measurement error or "the noise").
- Hospital size is used as a weight when calculating an overall reliability estimate. Weighting reduces the influence of hospitals that have less reliable rates due to a smaller number of patients at risk (small denominators).

Results of reliability testing

- Performance score testing results are provided by deciles, based on hospital size, for 319 and 320 hospitals from 2012-2013 HCUP SID data.
  - Overall signal-to-noise ratio of 0.60 for 3,199 hospitals with an overall average of 115 discharges per year,
    - a range from 0.4738 to 0.7765 for 1,280 hospitals with > 82 average discharges per year, and
    - a range from 0.0579 to 0.3954 for 1,919 hospitals with < 82 average discharges per year.
- The signal-to-noise method results in a reliability statistic that ranges from 0 to 1 for each facility. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences between hospital performance. A value of 0.7 is often regarded as a minimum acceptable reliability value. From the data presented only hospitals in the decile representing the largest hospitals (437.5 average discharges) have reliability above 0.7.
- The developer notes that " the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8."

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1) → empiric reliability testing (Box 2) →Testing of the measure score (Box 4) → appropriate method (Box 5) → testing results (Box 6) →6b moderate certainty

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- As a part of the original endorsement process in 2007, the developer conducted empirical validity testing based on Medicare inpatient fee-for-service claims for general surgical admissions from July 1,1999 through June 30, 2000 using data including 1,467 hospitals and 403,679 Medicare beneficiaries between 65 and 90 years of age. The developer assessed construct validity by estimating logic models using detailed patient characteristics and 5 hospital characteristics shown to be associated with better quality of care in previous studies.
- Omega statistics were also evaluated in these analyses. The omega statistic represents the ratio of the squared sum of the log odds for patient characteristics at the discharge-level variables divided by the corresponding quantity for hospital-level variables. Lower omega ratios may be indicative of a more desirable quality indicator. The omega ratio summarizing the contribution of patient characteristics at the discharge-level versus hospital-level variables was 57, compared to a ratio of 189 for the overall risk adjusted surgical mortality rate and 128 for NQF #0352.
- The developer also evaluated face validity using a structured panel review process that was based on the RAND appropriateness method, a modified Delphi process known as nominal group technique; the panel was convened in 2002 and was comprised of 7 multispecialty members.

Describe any updates to validity testing

- Empirical validity testing has been updated since the last review, using HCUP SID reference data from 2012-2013. The analyses use a broader definition of teaching status. The developer states that teaching hospitals had higher unadjusted PSI 04 rates, but lower adjusted PSI 04 rates relative to nonteaching hospitals but that this effect was less pronounced with the more inclusive definition of teaching hospitals (and all-payer data instead of Medicare data).

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:

- The additional empirical validity testing was conducted by correlating the measure score to hospital characteristics, including hospital teaching status. The developer used all-payer data from 34 states in 2012-2013 to confirm the association between hospital teaching status and PSI 04.

Validity testing results:

- Risk adjusted death rates are reported as overall reference population rates and as a distribution of performance for the measure.
- Empirical validity testing generally demonstrated favorable correlations between the measure score and hospital teaching status. As hypothesized, teaching hospitals and high volume PCI hospitals have lower PSI 04 rates. The following table shows odds ratio and p value for structural measures of quality when all patient characteristics are included in the model with only one hospital characteristic at a time:

Table 2.

Hospital characteristic	Marginal analysis Odds ratio (p-value)	Partial analysis Odds ratio (p-value)	Interpretation
Teaching hospital	0.807 (p<0.0001)	0.852 (p<0.0001)	Protective
High tech hospital	0.924 (p<0.005)	1.049 (NS)	Protective
Large hospital	0.872 (p<0.0001)	0.917 (p<0.01)	Protective
Less well-staffed hospital	1.108 (p<0.005)	1.044 (NS)	Not protective
Better nursing skill mix	0.832 (p<0.0001)	0.870 (p<0.0001)	Protective

- The developer notes that as summarized in the Evidence Form, numerous studies have linked failure to rescue measures, including PSI 04, to structure and process measures. Multiple studies have found lower FTR rates in hospitals with higher nurse-to-bed ratios, better nurse skill mix ratios, and better US-trained nurse ratios. Higher hospital volume was associated with lower FTR rates in at least 6 studies. In addition, studies have found that hospitals with the highest patient satisfaction scores and hospitals with better compliance with NQF Safe Practices had lower risk adjusted odds of FTR.
- The developer notes that although they did not conduct new face validity, they anticipate the results of face validity to be similar if a new panel was convened because the characteristics of these events, treatment and prevention of approaches have not changed substantially since 2002.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer conducted an empirical evaluation of exclusions using the 2013 data from 34 states to examine the percent of potential denominator cases excluded.
- The stratum-specific exclusions are meant to exclude cases for which the "complication" was actually the principal reason for admission or the primary indication for surgery. Some exclusions (e.g. immunocompromised state) are intended to exclude patients for whom death may be the expected outcome (i.e., less preventable). The table below summarizes the highly frequent exclusions.

	Denominator		
	Exclusion Count	After Exclusions	% Change
No Exclusion		329,827	
Transfers to acute care facility	9,889	319,827	3.0%
Pneumonia: Dx of immunocompromised	8,696	321,020	2.6%
Pneumonia: MDC 4	9,006	320,710	2.7%

Sepsis: Principal dx of Septicemia	45,202	284,514	13.7%
Sepsis: Principal dx of Infection	14,982	314,734	4.5%
Shock/Cardiac Arrest: MDC 4 or 5	27,819	301,897	8.4%
GI hemorrhage/Acute ulcer: MDC 6 0r 7	14,134	315,582	4.3%

- All patients transferred to other hospitals must be excluded from the analysis because the relevant outcome of these patients (i.e., dead or alive at the time of discharge from the acute inpatient setting) cannot be ascertained without social security numbers or other data elements to support linkage.
- Based on Needleman et al.'s analysis, AHRQ does NOT exclude patients with complications that were present on admission (i.e., upon transfer from another hospital, emergency department, or ambulatory surgery center).

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Risk adjustment summary

- The developer used a statistical risk model with 362 risk factors.
- Factors in the risk adjustment model were considered as a standard set of covariates grouped into four categories: demographics, severity of illness (Major Diagnostic Categories [MDCs]), Modified Diagnostic Related Groups [MDRGs]), comorbidities (AHRQ Comorbidity Software categories), and transfer-in status.
- The risk-adjusted rate for the overall PSI 04 is calculated as the observed to expected ratio multiplied by the reference population rate (the observed and expected values are summed across five categories of PSI 04 risk).
- Only those covariates present in at least 30 records for that PSI 04 strata are retained. A parsimonious model was identified using backward stepwise selection with bootstrapping.
- The analysis evaluates performance of the risk adjustment model(s) with respect to in-hospital death. The developer used c-statistic to measure how well the risk adjustment model distinguishes events from non-events.
- There are five distinct risk models (one for each type of complication or stratum).These five models currently have c-statistics ranging from 0.726 to 0.860. The c-statistic for the overall PSI 04 model is 0.829 in the 2013 HCUP data and it represents the overall performance of all five models.
- The developer notes that SDS variables were not included in the risk adjustment because there was no evidence or causal model to suggest that socioeconomic factors are associated with in-hospital death following serious surgical complications independent of quality of care, or are mediated by pre-hospital care (which may not fall within the proper realm of hospital accountability).

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the appropriate risk adjustment variables being used?*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in the risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- According to the developer, over all hospitals, using smoothed rates, this indicator has limited discrimination for identifying low or high performing hospitals; 16% of hospitals can be classified as better or worse than the threshold and 27% as better or worse than the benchmark, based on conventional statistical criteria. As hospital size increases, the discrimination also increases.
- For this analysis, the developer notes that "benchmark" refers to the smoothed indicator rate based on the 20th percentile of the reference population (i.e., 20% of hospitals have a lower mortality rate or better performance). "Threshold" refers to the indicator rate based on the 80th percentile (i.e., 80% have lower mortality or better performance).
- Assuming an underlying Gamma distribution for the smoothed rates of the measure, the benchmark and

threshold values are identified using population reference rates and signal variances computed from the entire AHRQ QI POA Reference Population.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- PSI 04 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all hospital discharge records.
- For these variables, frequencies of missing data are typically less than 1% of the state database. The developer notes that it is unlikely that bias would occur from such a low frequency of missing data.

Guidance from the Validity Algorithm: Specifications consistent with the evidence (Box 1) → Potential threats to validity addressed (Box 2) →Empirical validity testing performed (Box 3) →testing of measure score (Box 6) →Method was appropriate (Box 7) → Box 8 (high certainty) – High
Note that the body of literature has validated this measure.
Preliminary rating for validity: ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a.

- very clear documentation of data elements, defined using administrative and billing codes (ICD-9 and 10).
- It appears to me that this measure is only intended to be used in the HCUP SID dataset, which is publically available. It can't be used to identify specific hospitals for quality improvement efforts, though, I believe.
- In our organization's experience with the AHRQ QI's, consistent implementation is greatly enhanced by the fact that AHRQ providers the code/software, limiting the number of judgement calls that need to be made.

2a2.

- the developer tested the 'signal to noise' ratio for the measure, and this was acceptable (>0.7) for only the largest hospitals (>436 discharges). The developer feels that 0.4-0.8 is acceptable, which would include all hospitals in the dataset.
- I agree with the algorithm results on the measure worksheet. The testing was performed on a large data set and the results demonstrate moderate reliability.

2b1.

- I think the specifications align with the evidence presented.

2b2.

- This has been done in several ways. In prior applications, the developers conducted face validity with a panel of experts using the Delphi method. They also looked at performance of the measure using Medicare data to look at 5 other elements of hospital structure associated with better outcomes. They also generated an 'omega statistic', comparing the ratio of discharge to hospital-level odds ratios, which was acceptable compared to other measures. They recently updated validity testing by examining how teaching hospitals fared with and without risk adjustment and found reasonable results.
- Testing was performed with a large and diverse data set.

2b3.

- exclusions are reasonable and the rate of exclusions is presented. risk adjustment is intensive; goodness of fit statistics are high for risk adjusted models testing for event vs non event.
- SES is specifically not adjusted for.
- the ability to discriminate (meaningful differences) is moderate. only a quarter of hospitals can be distinguished as above or below the benchmark. This is improved when looking at larger hospitals only."
- I have no concerns with the validity of this measure

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer noted the following:

- All data elements are defined fields in electronic claims. The data are coded by someone other than the individual obtaining the original information. (e.g., DRG, ICD-9 codes on claims)
- The indicator is based on readily available administrative billing and claims data and U.S. Census data, thus is very feasible.
- The AHRQ QI software has been publicly available at no cost since 2001.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

- The developer states that all data elements are generated during the claims generation process, making this a feasible measure to report. It appears that AHRQ software is required to classify the claims diagnoses into the data fields used to construct the measure. This is an extra step but is not terribly burdensome.
- We calculate this measure and know several other organizations that do as well and all agree it's very straightforward to use the AHRQ QI software to generate measure results.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

See section 4.1 for list of programs.

Improvement results

- See Table 1 in response to question 1b.2 (also included in supplemental materials)
- The developer noted that the observed that PSI 04 rates have been relatively stable from 2011-2013 in the AHRQ QI POA Reference Population data (116-118 deaths per 1,000 patients with perioperative or postoperative complications).
- An earlier study of administrative data showed a decrease by 6.05% per year from the years 1998 to 2007. ($p<0.0001$) (Downey et al., 2012).

Unexpected findings (positive or negative) during implementation

- No evidence has been identified suggesting unintended consequences or findings for this measure.

Potential harms:

- No evidence has been identified suggesting unintended consequences for this measure.

Feedback : N/A

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

- This measure has been used in dozens of programs listed by the developer. It was associated with a 6.5% annual rate improvement initially, though the rate has stabilized in recent years. I can think of no unintended consequences.
- The measure is widely used.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0352 : Failure to Rescue In-Hospital Mortality (risk adjusted)
- 0353 : Failure to Rescue 30-Day Mortality (risk adjusted)

Harmonization

- AHRQ response to harmonization in current submission is linked here.
- During the previous evaluation of this measure the Committee discussed that these measures, while conceptually similar, have different aims, i.e., capture of avoidable complications vs. failure to rescue (NQF #0352 and #0353). Overall the Committee agreed that these measures have different objectives and are complementary.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0351

Measure Title: Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 5/31/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. 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 multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: In-hospital mortality

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

This indicator targets patients who are admitted for surgery (defined by specific MS-DRGs and an ICD-9-CM/ICD-10-PCS code for an operating room procedure, and a principal procedure within 2 days of admission OR admission type of elective) who die following the development of a serious, but treatable complication of care. Complications included in PSI 04 are as follows: 1) shock or cardiac arrest, 2) sepsis, 3) pneumonia or pneumonitis, 4) deep vein thrombosis or pulmonary embolism, and 5) gastrointestinal hemorrhage/acute ulcer. Evidence supports the concept that swift diagnosis and intervention by experienced, skilled nurses lowers the probability of death from these specific complications. This concept has been described as "failure to rescue (FTR)" to reflect the benefits of recognizing and responding skillfully to early signs of patient deterioration. Both better quality nurses and higher nurse staffing levels have been shown to improve PSI 04 rates. PSI 04 rates have been shown to be strongly associated with risk-adjusted mortality rates, as expected, but not with complication rates.[1]

Please note: Relevant text regarding evidence of the health outcome for previous submissions to NQF is included below in black.

Mortality is a frequent outcome among patients with serious treatable complications

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 housestaff and a lower percentage of board-certified anesthesiologists. The adverse occurrence rate was independent of this hospital characteristic. In a larger sample of 74,647 patients who underwent general surgical procedures in 1991-92, lower failure to rescue rates were found at hospitals with high ratios of registered nurses to beds. Failure rates were strongly associated with risk adjusted mortality rates, as expected, but not with complication rates. Finally, among 16,673 patients admitted for coronary artery bypass surgery, failure rates were lower (whereas complication rates were higher) at hospitals with magnetic resonance imaging facilities, bone marrow transplantation units, or approved residency training programs.

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 among major surgery patients from 799 hospitals in 11 states in 1997, even using administrative data to define complications. An increase from the 25th to the 75th percentile on these two measures of staffing was associated with 5.9% (95% CI, 1.5% to 10.2%) and 3.9% (95% CI, -1.1% to 8.8%) decreases, respectively, in the rate of failure-to-rescue among major surgery patients. These associations were inconsistent among medical patients, in that nursing skill mix was associated with the failure-to-rescue rate (rate ratio 0.81, 95% CI 0.66-1.00) but aggregate registered nurse staffing was not (rate ratio 1.00, 95% CI 0.99-1.01). An increase from the 25th to the 75th percentile on nursing skill mix was associated with a 2.5% (95% CI, 0.0% to 5.0%) decrease in the failure-to-rescue rate among medical patients.

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.

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.

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.

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.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

Please note that this is an outcome measure, so a systematic review of the body of evidence that supports the performance measure is not required. However, information is provided in 1a.4.1 and 1a.8 below, to provide additional context and support for the measure. Specifically, there are several high-quality systematic reviews and clinical practice guidelines that focus on early recognition and aggressive treatment of sepsis, antithrombotic therapy for venous thromboembolism, rapid diagnosis and stabilization of patients with gastrointestinal hemorrhage, and perioperative cardiovascular risk assessment and management. These practice guidelines reflect the process-outcome pathways that are facilitated by highly skilled, well-functioning multidisciplinary teams.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1. Dellinger RP, Levy MM, et al. Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. Crit Care Med. 2013 Feb;41(2):580-637. doi: 10.1097/CCM.0b013e31827e83af.
2. Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014; 130:2215.
3. Kearon C, Akl EA, Ornelas J, et al Antithrombotic therapy for vte disease: chest guideline and expert panel report. *Chest* 2016;149(2):315-352. doi:10.1016/j.chest.2015.11.026.

4. Konstantinides SV, Torbicki A, et. al. Task Force for the Diagnosis and Management of Acute Pulmonary Embolism of the European Society of Cardiology (ESC).2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism Eur Heart J. 2014 Nov 14;35(43):3033-69, 3069a-3069k. doi: 10.1093/eurheartj/ehu283. Epub 2014 Aug 29.
5. Laine L, Jensen DM. Management of patients with ulcer bleeding. Am J Gastroenterol. 2012 Mar;107(3):345-60. [133 references] http://www.guideline.gov/content.aspx?id=38023&search=gi+hemorrhage, accessed February 23, 2016.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Not applicable

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Not applicable

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Not applicable

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

Not applicable

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

Not applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

Not applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

Not applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

Not applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Not applicable

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Not applicable

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Not applicable

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

Not applicable

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Not applicable

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Not applicable

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Not applicable

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: Click here to enter date range

Not applicable

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Not applicable

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Not applicable

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Not applicable

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Not applicable

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

Formal environmental scans of the literature, including routine PubMed searches are performed to continually update evidence. The current evidence review results presented below constitute the most recent update, conducted in February 2016. Search terms included relevant MeSH terms (serious treatable complications, PSI 4 or PSI 04). The search was limited to English language publications. For completeness we also tested more inclusive search strings. Below we have provided a summary of the most up-to-date evidence.

1a.8.2. Provide the citation and summary for each piece of evidence.

Summary

The concept of failure to rescue (FTR), initially developed by Silber, has been widely studied.[2,3] Silber et al. published a series of studies linking risk adjusted mortality rates following complications to various physician factors, including residents and surgical housestaff, board-certified anesthesiologists, board certified surgeons, registered nurse staffing ratios.[2-4] Needleman and Buerhaus adapted Silber's concept of FTR for use with administrative data, targeting nurse sensitive measures by narrowing the number of complications included.[5,6] The Buerhaus measure was further modified and operationalized by AHRQ into the current PSI 04 specifications (first endorsed by NQF in 2008).[6] Ongoing research continues to confirm the association with nurse staffing, nurse skill mix, and other structural measures and processes of care (as outlined as follows) with PSI 04 and the similar FTR measure.[4,5,7-19] Johnston and colleagues recently published a systematic review of 42 articles to "identify the factors that contribute to high FTR rates and delayed escalation of care, and… summarize outcomes of interventions aimed at decreasing the rates of FTR and improving escalation of care."[3]

Structural measures: Nursing

FTR rates are associated with many hospital level measures of high-quality care. One of the most studied areas is the association of nursing characteristics and PSI 04 (including the variant measure for FTR).[4,5,8-19] As stated above, PSI 04 has special relevance to nursing, as nurses are on the front line of patient care, with key roles in prevention, surveillance, early detection, and intervention. Multiple studies have found lower FTR rates in hospitals with higher nurse-to-bed ratios, better nurse skill mix ratios, and higher US-trained nurse ratios.[5,9,11-17,19-24]

The most recent studies include those by Unruh et al, Park et al., and Blegen et al.[22,24,25] Using latent growth curve models, Unruh et al. found that increases in RN full time equivalents were associated with decreases in PSI 04 in 124 Florida hospitals over 9 years.[25] In addition to nurse staffing, Park et al. examined patient turnover rates in non-ICUs (56.1%) and ICUs (45.4%) in 42 teaching hospitals in 2005.[24] In general, they found that more RN hours per patient day were associated with lower rates of PSI 04, controlling for non-RN staffing and hospital characteristics. Patient turnover was not related to PSI 04 rates in either non-ICUs or ICUs; however, the association between RN staffing and PSI 04 in non-ICU settings differed significantly depending on the level of patient turnover. Specifically, when patient turnover rates increased from 48.6% (25th percentile) to 60.7% (75th percentile), the effect of RN staffing on PSI 04 was reduced by 11.5%.[24] Using 21 of the same teaching hospitals, which agreed to provide detailed data on nurse education over 4 quarters, Blegen et al. found that the proportion of baccalaureate-educated nurses was inversely associated with PSI 04 ($r = -0.399$; $p < 0.05$).[22]

Magnet designation by the American Nurses Credentialing Center is a related concept that captures nursing empowerment and excellence. Mills and Gillespie matched 80 Magnet hospitals with 80 non-Magnet hospitals in the 2001-2005 Nationwide Inpatient Samples on 12 hospital characteristics and reported no significant difference in PSI 04 rates between matched hospitals.[26] McHugh et al. focused on four states (CA, FL, NJ, PA) in 2006-2007, and linked state hospital discharge data with detailed data on nurse

characteristics and work environments, from surveys of over 100,000 registered nurses.[27] A composite measure of nursing, estimated as the likelihood of a hospital being Magnet credentialed as a function of nursing factors, was significantly associated with lower odds of failure-to-rescue (OR 0.48, 95% CI 0.37-0.63). The Practice Environment Scale of the Nursing Work Index was the most important component of that composite. When an indicator for Magnet status was added to this model, the composite nursing measure was still significantly associated with failure-to-rescue (OR 0.57, 95% CI 0.41-0.79) and the Magnet effect approached significance (OR 0.88, 95% CI 0.77-1.01), indicating that most but not all of the Magnet effect can be explained by other aspects of the nursing work environment. Kutney-Lee et al. extended these findings by showing that between 1999 and 2007, 11 newly Magnet recognized hospitals in Pennsylvania reduced their FTR rates by 6.1 per 1000 patients (P=0.02) compared with 125 non-Magnet comparison hospitals.[28] Finally, Friese et al. used Medicare data on patients hospitalized for coronary artery bypass graft surgery, colectomy, or lower extremity bypass in 1998–2010 to show that patients treated in Magnet hospitals were 8.6% (95% CI: 0.88, 0.95) less likely to die from FTR than patients treated in non-Magnet hospitals.[29]

One recent study found mixed effects with respect to the impact of regulations designed to improve nurse staffing on PSI 04 rates in California. In this study, Mark et al. divided CA hospitals into quartiles based on their pre-regulation nurse staffing levels for medical-surgical and pediatric services (Quartile 1 = lower RN staffing).[30] Using difference-in-differences Poisson fixed effects models to compare California with 12 other states, they found that PSI 04 decreased significantly more in California Quartile 1 hospitals than in comparison state hospitals in both the immediate pre-regulation and post-regulation periods (by 37.1% and 30.7%, respectively, $p < .05$), while RN staffing improved by 27.3% and 35.0%, respectively. However, PSI 04 also decreased significantly more in California Quartile 4 hospitals than in comparison state hospitals in the post-regulation period (by 32.9%, $p < .05$), even though these hospitals only improved their RN staffing by 15.1%.

Structural Measures: Physician staffing

As with nursing, higher resident-to-bed ratios and higher case volume are associated with lower FTR rates. Confirmatory findings have been reported in general surgery, orthopedics, vascular surgery, and cardiovascular surgery.[11,20,31-33] The use of resident housestaff is hypothesized to increase ability to identify and act upon complications early.

Most recently, a 2013 retrospective cohort analysis of the National Surgical Quality Improvement Program (NSQIP) data (2005-2009) found that resident trainee participation in complex, oncologic surgery was associated with significantly higher rates of 30-day postoperative complications in NSQIP-participating hospitals, but lower 30-day mortality and lower FTR among patients suffering complications (5.9 vs. 7.6%, AOR 0.79, 95% CI 0.68-0.90).[34] Ferraris et al (2014) analyzed NSQIP data from 200 hospitals and found that cases with surgical resident involvement had increased operative morbidity (11.4% vs 7.8% with attending only; $P < .001$) and prolonged operative time (127 minutes vs 93 minutes for attending only; $P < .001$), but more favorable FTR (9.4% vs 12.4% for attending alone; $P < .001$).[35] The most serious complications occurred 5 to 10 days before death, suggesting that there is a window for intervention to rescue patients with early aggressive treatments for the sentinel complication. Among 1,056,865 CABG patients in the 1998-2007 Nationwide Inpatient Sample, teaching hospitals had 14% lower FTR rates (OR 0.86, 95% CI 0.79-0.93) than non-teaching hospitals, despite higher complication rates (OR 1.023, 95% CI 1.00-1.05), in the first quarter of the academic cycle.[36] Finally, Navathe et al. (2013) found no differences in FTR in teaching hospitals when comparing rates prior to and after ACGME duty hour reform.[37]

Patient clinical and sociodemographic factors

Risk for FTR varies by procedure and complication type.[35,38] In the NSQIP database, timing of complication also impacted mortality rate, with early cardiac arrest and unplanned intubation associated with lower risk-adjusted mortality (HR 0.59, CI 0.51-0.68; HR 0.38, CI 0.33-0.43, respectively).[39] Late pneumonia was associated with higher observed mortality but not adjusted mortality. Ferris et al. reported that more than two-thirds of patients with failure to rescue have multiple complications.[35] Silber et al. found that blacks had higher failure-to-rescue rates (6.1% vs. 5.1%, P<0.001) than whites, matched on age, sex year, state and procedure in Medicare data. However, when preoperative medical risk factors were added to this matching algorithm, there was no significant racial difference in FTR. [40] Similarly, Black patients had higher odds of observed FTR than white patients (1.14, 95% CI 1.07-1.23) in four states, although this risk was not significant when adjusted by SES, and the white patients had higher risk than black patients when adding other patient characteristics to the model.[41] The study identified a modest but significant interaction between race and nurse staffing for FTR.[40] Each additional patient in a nurse's workload raised the odds of FTR for black patients by a factor of 1.10 (95% CI 1.02-1.18) and 1.04 for white patients (95% CI 1.01-1.06).[41] Among Medicare patients undergoing cancer surgery, patients in the lowest SES quintile had the highest FTR rates (26.7% vs 23.2%, P = .007), and hospitals treating larger portions of SES also had higher rates. However, these disparities did not remain after adjustment for hospital effects.[42]

Other hospital level factors

Higher hospital volume was associated with lower FTR rates in at least 6 studies.[5,31,33,43-46] Across three high complexity

cancer procedures, volume was more strongly associated with FTR rates than complication rates (lowest volume quintile odds ratio 1.17, 95% CI 1.02-1.33 for complications vs. 2.89, 95% CI 2.40-3.48 for FTR, compared to the highest volume quintile).[31] Among 119,434 Medicare cardiovascular surgery patients, FTR was consistently lower among hospitals with higher procedural volume across procedure types.[33] Similar results have been found for hepatic and hepato-pancreatico-biliary procedures, ovarian cancer resection, and surgical oncology patients.[43-46] In one study, NCI cancer centers had lower FTR rates than other hospitals among surgical oncology patients (OR 0.68, 95% CI 047-0.97).[47] Among cardiac surgery patients at 17 hospitals, hospitals with lower FTR rates had longer postoperative and intensive care unit stays after the index operation (2 to 3 days; p<0.001).[48] Wakeam et al. (2014) found that higher burden, as classified by the Safety Net hospital categories, was associated with higher odds of adjusted FTR for High Burden Hospitals (OR, 1.35; 95% CI, 1.19-1.53; P < .001) and Moderate Burden Hospitals (OR, 1.15; 95% CI, 1.05-1.27; P = .005) compared with Low Burden Hospitals.[49]

Hospital processes of care and outcomes measures

FTR has been studied in relationship to other quality measures. Sacks et al (2015) linked Medicare claims with NSQIP and Hospital Compare data on reported patient satisfaction. Hospitals in the highest quartile of Hospital Consumer Assessment of Healthcare Providers and Systems scores had significantly lower risk-adjusted odds of FTR (OR = 0.82, 95% CI, 0.70-0.96).[50] Supporting the preventability of FTR events, Brooke, et al. (2012) found that hospitals that complied fully with the 27 National Quality Forum (NQF) safe practices had an increased likelihood of diagnosing a complication after any of six high-risk operations (odds ratio [OR], 1.13; 95% confidence interval [CI], 1.03-1.25), but had a decreased likelihood of failure to rescue (OR, 0.82; 95% CI, 0.71-0.96), and a decreased odds of mortality (OR, 0.80; 95% CI, 0.71-0.91).[51]

Several studies have examined interventions to reduce FTR. In a single site study, an intervention including crew resource management and checklist implementation was associated with a reduction in FTR from 25% to 12% (P=0.03).[52] Two studies demonstrate lower FTR rates with more available technology.[4,53] In surgical oncology, NCI cancer center designation has also been found to be associated with reduced FTR rates following cancer surgery [47,54] Sheetz and colleagues evaluated whether increased hospital care intensity (HCI) is associated with improved outcomes following seven major cardiovascular, orthopedic, or general surgical operations in the Medicare population. High-HCI hospitals had greater rates of major complications than low-HCI centers (risk ratio, 1.04; 95% CI, 1.03-1.05) and there was a decrease in failure to rescue at high compared with low-HCI hospitals (risk ratio, 0.95; 95% CI, 0.94-0.97). Using multilevel-models, HCI reduced the variation in failure-to-rescue rates between hospitals by 2.7% after accounting for patient comorbidities and hospital resources.[55]

Hyder et al. estimated the improvement in mortality that may be realized through a reduction in FTR using the Nationwide Inpatient Sample (NIS, 2007-2011). High-mortality hospitals had higher FTR rates than low-mortality hospitals (22.4% vs 20.2%, p = 0.0020). Using Monte Carlo models they estimated that reducing the FTR gap by nearly 75% (2.73%; 95% CI 2.61 to 2.87) would potentially result in a 50% absolute reduction in baseline mortality (5.22%) for five target subpopulations.[56]

Indicator specifications

Several variations of the FTR measure are currently available, including the NQF endorsed measure from Children's Hospital of Philadelphia (NQF 0352) and PSI 04 (NQF 0351). In a comparison study, Silber at al showed the split sample reliability of the broader Silber FTR was higher than PSI 04 (0.32 vs. 0.18 vs. 0.18).[21] Silber's FTR is more highly correlated with 30 day mortality rate than PSI 04 (0.83 vs. 0.43). However, these results are expected given the broad scope of the Silber measure to capture most in-hospital deaths. This study may also suggest that the AHRQ measure captures a different aspect of quality care than broad mortality measures and a reduction in reliability may be acceptable given the focus on nursing sensitive complications.[57] Horwitz assessed the validity of V2.1 Rev 1 of PSI 04 using chart review data from the University Hospital Consortium. [58] In this study, using chart data as the gold study, they determined that 13.6% of the complications did not occur (15.8% of surgical cases), while an additional 8.1% met an AHRQ exclusion criterion but were incorrectly included. Since this study, the PSI 04 algorithm has been continuously improved as a result of learnings from studies of the PSI algorithms and user feedback.

Needleman et al., using PSI 04 (v3.1), examined whether the accuracy of PSI 04 (v3.1) could be improved by testing three present-on-admission (POA)-based exclusion rules using California HCUP Data.[16] Although these exclusion rules improved

the C-statistic of the failure-to-rescue measure, they did not affect the strong association between PSI 04 and higher nurse staffing and a greater percentage of registered nurses. The mortality rate was 22% among patients with hospital-acquired complications compared to 13% for patients with POA complications. Patients with hospital-acquired complications also had longer lengths of stay than patients with POA complications (14.1 versus 8.4 days; risk-adjusted difference = 4.3 days). The authors conclude that "failure-to-rescue is found in this analysis to be a robust measure and generates similar results regardless of how it is parameterized."[16]

References:

1. Silber J, Rosenbaum P, Williams S, Ross R, Schwartz J. 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.
2. 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):615-629.
3. Johnston MJ, Arora S, King D, et al. A systematic review to identify the factors that affect failure to rescue and escalation of care in surgery. *Surgery.* 2015;157(4):752-763.
4. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist direction and patient outcomes. *Anesthesiology.* 2000;93(1):152-163.
5. Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. *N Engl J Med.* 2002;346(22):1715-1722.
6. Rosen AK, Rivard P, Zhao S, et al. Evaluating the patient safety indicators: How well do they perform on Veterans Health Administration data? . *Med Care.* 2005;43(9):873-884.
7. Needleman J, Buerhaus PI, Mattke S, Stewart M, Zelevinsky K. *Nurse Staffing and Patient Outcomes in Hospitals.* Boston, MA: Health Resources Services Administration; February 28, 2001 2001. 230-99-0021.
8. Manojlovich M, Talsma A. Identifying nursing processes to reduce failure to rescue. *J Nurs Adm.* 2007;37(11):504-509.
9. Schmid A, Hoffman L, Happ MB, Wolf GA, DeVita M. Failure to rescue: a literature review. *J Nurs Adm* 2007;37(4):188-198.
10. Silber JH, Rosenbaum PR, Ross RN. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? . *Journal of the American Statistical Association.* 1995;90(429):7-18.
11. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA.* 2002;288(16):1987-1993.
12. Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational levels of hospital nurses and surgical patient mortality. *JAMA.* 2003;290(12):1617-1623.
13. Kendall-Gallagher D, Aiken LH, Sloane DM, Cimiotti JP. Nurse Specialty Certification, Inpatient Mortality, and Failure to Rescue. *J Nurs Scholarship.* 2011;43(2):188-194.
14. Friese CR, Aiken LH. Failure to rescue in the surgical oncology population: Implications for nursing and quality improvement. *Oncol Nurs Forum.* 2008;35(5):779-785.
15. Ghaferi AA, Osborne NH, Birkmeyer JD, Dimick JB. Hospital Characteristics Associated with Failure to Rescue from Complications after Pancreatectomy. *J Am Coll Surgeons.* 2010;211(3):325-330.
16. Needleman J, Buerhaus PI, Vanderboom C, Harris M. Using present-on-admission coding to improve exclusion rules for quality metrics: the case of failure-to-rescue. *Med Care.* 2013;51(8):722-730.
17. Seago JA, Williamson A, Atwood C. Longitudinal analyses of nurse staffing and patient outcomes - More about failure to rescue. *J Nurs Admin.* 2006;36(1):13-21.
18. Boyle SM. Nursing unit characteristics and patient outcomes. *Nurs Econ.* 2004;22(3):111-+.
19. Clarke SP, Aiken LH. Failure to rescue. *Am J Nurs.* 2003;103(1):42-47.
20. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist board certification and patient outcomes. *Anesthesiology.* 2002;96(5):1044-1052.
21. Silber JH, Romano PS, Rosen AK, Wang Y, Even-Shoshan O, Volpp KG. Failure-to-rescue: comparing definitions to measure quality of care. *Med Care.* 2007;45(10):918-925.
22. Blegen MA, Goode CJ, Park SH, Vaughn T, Spetz J. Baccalaureate education in nursing and patient outcomes. *J Nurs Adm.* 2013;43(2):89-94.
23. Neff DF, Cimiotti J, Sloane DM, Aiken LH. Utilization of non-US educated nurses in US hospitals: implications for hospital mortality. *Int J Qual Health Care.* 2013;25(4):366-372.
24. Park SH, Blegen MA, Spetz J, Chapman SA, De Groot H. Patient turnover and the relationship between nurse staffing and patient outcomes. *Res Nurs Health.* 2012;35(3):277-288.
25. Unruh LY, Zhang NJ. Nurse staffing and patient safety in hospitals: new variable and longitudinal approaches. *Nurs Res.* 2012;61(1):3-12.
26. Mills AC, Gillespie KN. Effect of Magnet hospital recognition on 2 patient outcomes. *J Nurs Care Qual.* 2013;28(1):17-23.

27. McHugh MD, Kelly LA, Smith HL, Wu ES, Vanak JM, Aiken LH. Lower mortality in magnet hospitals. *Med Care.* 2013;51(5):382-388.
28. Kutney-Lee A, Stimpfel AW, Sloane DM, Cimiotti JP, Quinn LW, Aiken LH. Changes in patient and nurse outcomes associated with magnet hospital recognition. *Med Care.* 2015;53(6):550-557.
29. Friese CR, Xia R, Ghaferi A, Birkmeyer JD, Banerjee M. Hospitals In 'Magnet' Program Show Better Patient Outcomes On Mortality Measures Compared To Non-'Magnet' Hospitals. *Health Affair.* 2015;34(6):986-992.
30. Mark BA, Harless DW, Spetz J, Reiter KL, Pink GH. California's minimum nurse staffing legislation: results from a natural experiment. *Health Serv Res.* 2013;48(2 Pt 1):435-454.
31. Ghaferi AA, Birkmeyer JD, Dimick JB. Hospital volume and failure to rescue with high-risk surgery. *Med Care.* 2011;49(12):1076-1081.
32. Mell MW, Kind A, Bartels CM, Smith MA. Failure to rescue and mortality after reoperation for abdominal aortic aneurysm repair. *J Vasc Surg.* 2011;54(2):346-351; discussion 351-342.
33. Gonzalez AA, Dimick JB, Birkmeyer JD, Ghaferi AA. Understanding the volume-outcome effect in cardiovascular surgery: the role of failure to rescue. *JAMA Surg.* 2014;149(2):119-123.
34. Castleberry AW, Clary BM, Migaly J, et al. Resident education in the era of patient safety: a nationwide analysis of outcomes and complications in resident-assisted oncologic surgery. *Ann Surg Oncol.* 2013;20(12):3715-3724.
35. Ferraris VA, Bolanos M, Martin JT, Mahan A, Saha SP. Identification of patients with postoperative complications who are at risk for failure to rescue. *JAMA Surg.* 2014;149(11):1103-1108.
36. Gopaldas RR, Overbey DM, Dao TK, Markley JG. The impact of academic calendar cycle on coronary artery bypass outcomes: a comparison of teaching and non-teaching hospitals. *J Cardiothorac Surg.* 2013;8:191.
37. Navathe AS, Silber JH, Small DS, et al. Teaching hospital financial status and patient outcomes following ACGME duty hour reform. *Health Serv Res.* 2013;48(2 Pt 1):476-498.
38. Ghaferi AA, Birkmeyer JD, Dimick JB. Complications, failure to rescue, and mortality with major inpatient surgery in medicare patients. *Ann Surg.* 2009;250(6):1029-1034.
39. Wakeam E, Hyder JA, Tsai TC, Lipsitz SR, Orgill DP, Finlayson SR. Complication timing and association with mortality in the American College of Surgeons' National Surgical Quality Improvement Program database. *J Surg Res.* 2015;193(1):77-87.
40. Silber JH, Rosenbaum PR, Kelz RR, et al. Examining Causes of Racial Disparities in General Surgical Mortality: Hospital Quality Versus Patient Risk. *Med Care.* 2015;53(7):619-629.
41. Carthon JM, Kutney-Lee A, Jarrin O, Sloane D, Aiken LH. Nurse staffing and postsurgical outcomes in black adults. *J Am Geriatr Soc.* 2012;60(6):1078-1084.
42. Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. *JAMA Surg.* 2014;149(5):475-481.
43. Schneider EB, Ejaz A, Spolverato G, et al. Hospital volume and patient outcomes in hepato-pancreatico-biliary surgery: is assessing differences in mortality enough? *J Gastrointest Surg.* 2014;18(12):2105-2115.
44. Spolverato G, Ejaz A, Hyder O, Kim Y, Pawlik TM. Failure to rescue as a source of variation in hospital mortality after hepatic surgery. *Br J Surg.* 2014;101(7):836-846.
45. Wright JD, Herzog TJ, Siddiq Z, et al. Failure to rescue as a source of variation in hospital mortality for ovarian cancer. *J Clin Oncol.* 2012;30(32):3976-3982.
46. Yasunaga H, Hashimoto H, Horiguchi H, Miyata H, Matsuda S. Variation in cancer surgical outcomes associated with physician and nurse staffing: a retrospective observational study using the Japanese Diagnosis Procedure Combination Database. *BMC Health Serv Res.* 2012;12:129.
47. Friese CR, Earle CC, Silber JH, Aiken LH. Hospital characteristics, clinical severity, and outcomes for surgical oncology patients. *Surgery.* 2010;147(5):602-609.
48. LaPar DJ, Ghanta RK, Kern JA, et al. Hospital variation in mortality from cardiac arrest after cardiac surgery: an opportunity for improvement? *Ann Thorac Surg.* 2014;98(2):534-539; discussion 539-540.
49. Wakeam E, Hevelone ND, Maine R, et al. Failure to rescue in safety-net hospitals: availability of hospital resources and differences in performance. *JAMA Surg.* 2014;149(3):229-235.
50. Sacks GD, Lawson EH, Dawes AJ, et al. Relationship Between Hospital Performance on a Patient Satisfaction Survey and Surgical Quality. *JAMA Surg.* 2015;150(9):858-864.
51. Brooke BS, Dominici F, Pronovost PJ, Makary MA, Schneider E, Pawlik TM. Variations in surgical outcomes associated with hospital compliance with safety practices. *Surgery.* 2012;151(5):651-659.
52. Young-Xu YN, Fore AM, Metcalf A, Payne K, Neily J, Sculli GL. Using Crew Resource Management and a 'Read-and-Do Checklist' to Reduce Failure-to-Rescue Events on a Step-Down Unit. *Am J Nurs.* 2013;113(9):51-57.
53. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med.* 2009;361(14):1368-1375.

54. Friese C, Lake E, Aiken LH, Silber JH, Sochalski J. Hospital nurse practice environments and outcomes for surgical oncology patients. *Health Serv Res.* 2008;43(4):1145-1163.
55. Sheetz KH, Dimick JB, Ghaferi AA. The association between hospital care intensity and surgical outcomes in medicare patients. *JAMA Surg.* 2014;149(12):1254-1259.
56. Hyder JA, Wakeam E, Adler JT, DeBord Smith A, Lipsitz SR, Nguyen LL. Comparing Preoperative Targets to Failure-to-Rescue for Surgical Mortality Improvement. *J Am Coll Surg.* 2015;220(6):1096-1106.
57. Needleman J, Buerhaus PI. Failure-to-rescue: Comparing definitions to measure quality of care. *Medical Care.* 2007;45(10):913-915.
58. Horwitz LI, Cuny JF, Cerese J, Krumholz HM. Failure to rescue - Validation of an algorithm using administrative data. *Medical Care.* 2007;45(4):283-287.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
PSI04_Measure_Evidence_Form_160531_v2.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
This indicator targets patients who are admitted for surgery who die following the development of a serious but treatable complication of care. Examples of such complications include: 1) shock or cardiac arrest, 2) sepsis, 3) pneumonia, 4) deep vein thrombosis or pulmonary embolism, and 5) gastrointestinal hemorrhage or acute ulcer. This indicator is fundamentally different than other PSIs, as it reflects the effectiveness of the hospital in rescuing a patient from complications versus preventing the underlying complications.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
These tables (1a. and 1b.) are also included in the supplemental files.

Table 1a. Reference Population Observed Rate for Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04), 2011-2013

Overall Reference Population Rate

Year2	Number of Hospitals (Numerator)1	Outcome of Interest Population at Risk (Denominator)1	Observed Rate Per 1000 Surgical Discharges1
2013	2,783	21,242	182,512 116.3869
2012	2,860	21,897	185,872 117.8069
2011	2,748	21,403	181,317 118.0419

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)
1The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population of all hospitals included in the reference population data (denominator).

2Reference population is limited to states with present on admission data (POA). Since many states did not report POA data prior to 2011 we have not included testing prior to 2011.

Table 1b. Distribution of Hospital Performance for Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04) in 2-year Pooled Data (2011-2012, 2012-2013)1

Distribution of Hospital-level Observed Rates in Reference Population

Year3	Number of Hospitals	Rates per 1000 Surgical Discharges (p=percentile)2							
		Mean	SD2	p5	p25	Median	p75	p95	
2011-2012	3,212	103.83	85.55	0.00	58.06	102.61	140.35	217.39	
2012-2013	3,398	100.76	88.28	0.00	51.28	99.92	137.25	212.12	

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

1Consistent with the recommended minimum reporting time period, results are presented for data combining 2 years of data: 2011 and 2012, 2012 and 2013. Data from 2012 are included in both time periods reported. Limitations in present on admission data (POA) data availability (see below) do not allow for use of earlier years.

2The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals in the dataset with at least one case in the denominator, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile. Standard deviation refers to the spread in observed values in relation to the mean.

3Reference population is limited to states with present on admission data (POA). Since many states did not report POA data prior to 2011 we have not included testing prior to 2011.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Not applicable

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use*.

Table 2 shows that the risk of death for surgical inpatients with serious treatable complications varies by age, sex, community-level income, expected payer, and region. In 2013, among 182,512 surgical patients with serious treatable conditions, older patients, men, those from lower income communities, those with Medicare, Medicaid or uninsured, and those treated in the Northeast were at greater risk of death (after controlling for a variety of clinical risk factors) than younger patients, women, those from higher income communities, those with private insurance and those treated in the Midwest or West. These findings are based on 182,512 discharges from 2,783 hospitals in the 34 states in 2013 and reflect national population estimates. The findings may be different at an individual hospital-level.

Please note: Table 2, as shown below, is unformatted and may be difficult to read. A formatted Table 2 is provided in the supplemental files.

Table 2. Risk-Adjusted Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04) per 1,000 surgical discharges, by patient and hospital characteristics, 2013

Patient/hospital characteristic	Risk-adjusted Estimate1	Std Errorp-value	(Ref Grp = *)	Lower 95% CL	Upper 95% CL
Total U.S.	116.387	0.675		115.063	117.711
Patient Characteristics					
Age Groups:					
18-39	74.420	2.443	*	69.631	79.209
40-64	100.445	1.105	<.001	98.280	102.611
65 and over	133.431	0.923	<.001	131.621	135.241
Gender:					
Male2	120.100	0.909	*	118.319	121.882
Female	111.841	1.010	<.001	109.860	113.821
Patient Zip Code Median Income					

First quartile (lowest income)			121.085	2.383	<.001	116.415	125.755
Second quartile	118.659	1.532	<.001	115.656	121.661		
Third quartile	120.176	1.393	<.001	117.445	122.907		
Fourth quartile (highest income)2			112.644	0.969	*	110.746	114.543
Location of patient residence (NCHS)3:							
Rural	121.536	5.067	0.148	111.605	131.468		
Urban2	116.192	0.683	*	114.854	117.531		
Expected payment source:							
Private insurance2			109.914	1.534	*	106.906	112.921
Medicare	116.361	0.838	<.001	114.718	118.004		
Medicaid	120.407	2.300	<.001	115.899	124.915		
Uninsured / self-pay / no charge			136.964	3.203	<.001	130.687	143.241
Other insurance	116.814	4.217	0.062	108.549	125.078		
Location of Care:							
Northeast2			122.821	1.769	*	119.355	126.288
Midwest	110.965	1.440	<.001	108.143	113.788		
South	118.210	1.097	0.013	116.061	120.359		
West	115.587	1.375	<.001	112.893	118.282		

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

1Rates are adjusted using the AHRQ QI PSI POA Reference Population for 2013 as the standard population. Age and gender are removed from models for the relevant strata.

2Reference group

3NCHS - National Center for Health Statistics designation for urban-rural locations. Metropolitan areas are considered urban and micropolitan or non-core areas are considered rural.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

Not applicable

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Patient/societal consequences of poor quality

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

This indicator was originally proposed by Silber et al. as a more powerful tool than the risk-adjusted mortality rate to detect true differences in patient outcomes across hospitals.1 The underlying premise was that better hospitals are distinguished not by having fewer adverse occurrences but by more successfully averting death among (i.e., rescuing) patients who experience such complications. Silber et al's original definition was based on key clinical findings abstracted from the medical records of 2,831 cholecystectomy patients and 3,141 transurethral prostatectomy patients admitted to 531 hospitals in 1985.1 The key postoperative diagnoses that defined the denominator at risk of "failure to rescue" (FTR) included cardiac arrhythmias, congestive heart failure, cardiac arrest, pneumonia, pulmonary embolus, pneumothorax, renal dysfunction, stroke, wound infection, and unplanned return to surgery.1 More recently, Needleman and Buerhaus adapted failure to rescue to administrative data sets, with a specific focus on optimizing the sensitivity of the indicator to nurse staffing and skill mix. Their denominator definition included the ICD-9-CM codes for sepsis, pneumonia (including aspiration), acute upper gastrointestinal bleeding, shock, cardiac/respiratory arrest, deep vein thrombosis (DVT), and pulmonary embolus (PE).2 Both specifications have been linked to factors such as board-certified anesthesiologists, board certified surgeons, registered nurse staffing ratios, nursing skill mix, and other structural and processes measures of high-quality care. 1-18 Due to improvements in care, between 1998 and 2007, FTR decreased by 6.05% per year

(p<0.0001).19 However, Table 1 above shows that PSI 04 still captures approximately 43,000 deaths each year in the 34 states in the all-payer reference population.

1c.4. Citations for data demonstrating high priority provided in 1a.3

1. 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):615-629.
2. Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. N Engl J Med. 2002;346(22):1715-1722.
3. Navathe AS, Silber JH, Small DS, et al. Teaching hospital financial status and patient outcomes following ACGME duty hour reform. Health Serv Res. 2013;48(2 Pt 1):476-498.
4. Needleman J, Buerhaus PI, Mattke S, Stewart M, Zelevinsky K. Nurse Staffing and Patient Outcomes in Hospitals. Boston, MA: Health Resources Services Administration; February 28, 2001 2001. 230-99-0021.
5. Manojlovich M, Talsma A. Identifying nursing processes to reduce failure to rescue. J Nurs Adm. 2007;37(11):504-509.
6. Schmid A, Hoffman L, Happ MB, Wolf GA, DeVita M. Failure to rescue: a literature review. J Nurs Adm 2007;37(4):188-198.
7. Silber JH, Rosenbaum PR, Ross RN. Comparing the contributions of groups of predictors: Which outcomes vary with hospital rather than patient characteristics? . Journal of the American Statistical Association. 1995;90(429):7-18.
8. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. Jama-J Am Med Assoc. 2002;288(16):1987-1993.
9. Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational levels of hospital nurses and surgical patient mortality. JAMA. 2003;290(12):1617-1623.
10. Kendall-Gallagher D, Aiken LH, Sloane DM, Cimiotti JP. Nurse Specialty Certification, Inpatient Mortality, and Failure to Rescue. J Nurs Scholarship. 2011;43(2):188-194.
11. Friese CR, Aiken LH. Failure to rescue in the surgical oncology population: Implications for nursing and quality improvement. Oncol Nurs Forum. 2008;35(5):779-785.
12. Ghaferi AA, Osborne NH, Birkmeyer JD, Dimick JB. Hospital Characteristics Associated with Failure to Rescue from Complications after Pancreatectomy. J Am Coll Surgeons. 2010;211(3):325-330.
13. Needleman J, Buerhaus PI, Vanderboom C, Harris M. Using present-on-admission coding to improve exclusion rules for quality metrics: the case of failure-to-rescue. Med Care. 2013;51(8):722-730.
14. Seago JA, Williamson A, Atwood C. Longitudinal analyses of nurse staffing and patient outcomes - More about failure to rescue. J Nurs Admin. 2006;36(1):13-21.
15. Boyle SM. Nursing unit characteristics and patient outcomes. Nurs Econ. 2004;22(3):111-+.
16. Clarke SP, Aiken LH. Failure to rescue. Am J Nurs. 2003;103(1):42-47.
17. Silber JH, Kennedy SK, Even-Shoshan O, et al. Anesthesiologist direction and patient outcomes. Anesthesiology. 2000;93(1):152-163.
18. Johnston MJ, Arora S, King D, et al. A systematic review to identify the factors that affect failure to rescue and escalation of care in surgery. Surgery. 2015;157(4):752-763.
19. Downey JR, Hernandez-Boussard T, Banka G, Morton JM. Is patient safety improving? National trends in patient safety indicators: 1998-2007. Health Serv Res. 2012;47(1 Pt 2):414-430.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*

Not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Surgery

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://1.usa.gov/1TksC2k Note: The URL link will be updated for version 6.0 public release found via the module page: http://qualityindicators.ahrq.gov/Modules/psi_resources.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** PSI04_Technical_Specifications_v6.0_160527.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
As standard protocol, the AHRQ QI program annually updates all measures with Fiscal Year coding changes, refinements based on stakeholder input, refinements to improve specificity and sensitivity based on additional analyses, and necessary software changes. In addition, approximately every two years, AHRQ updates the risk-adjustment parameter estimates based on the most recent year of data (i.e., the most current reference population possible). The refined measures are tested and confirmed to be valid and reliable prior to release of the updated software.

Since the last update, the following changes have been made to the indicator:
This revised version will be implemented in forthcoming version 6.0 specifications in 2016. This version (v6.0) includes the following changes from the previously-endorsed version (v4.4):

- An ICD-10-CM/PCS version has been created.
- STRATUM_SHOCK (previously Stratum D):
  - Abortion-related shock diagnosis codes added to Denominator
    63450 - SPON ABORT W SHOCK-UNSP
    63451 - SPON ABORT W SHOCK-INC
    63452 - SPON ABORT W SHOCK-COMP
    63550 - LEGAL ABORT W SHOCK-UNSO
    63551 - LEGAL ABORT W SHOCK-INC
    63552 - LEGAL ABORT W SHOCK-COMP
    63650 - ILLEG AB W SHOCK-UNSO
    63651 - ILLEG ABORT W SHOCK-INC
    63652 - ILLEG ABORT W SHOCK-COMP
    63750 - ABORT NOS W SHOCK-UNSO
    63751 - ABORT NOS W SHOCK-INC
    63752 - ABORT NOS W SHOCK-COMP
    6385 - ATTEM ABORTION W SHOCK
  - Codes removed from denominator (eliminating overlap with STRATUM_SEPSIS):
    78552 – SEPTIC SHOCK
    99802 – POSTOP SHOCK, SEPTIC
  - Code added to Denominator principal diagnosis exclusion: 53021 (ULCER OF ESOPHAGUS WITH BLEEDING)
- STRATUM_SEPSIS (previously Stratum C):
  - Codes removed from Denominator (eliminating overlap with STRATUM_SHOCK):
    78559 (SHOCK W/O TRAUMA NEC)
    99800 (POSTOPERATIVE SHOCK, NOS)
  - Codes added to the Denominator Exclusion:

70700 PRESSURE ULCER, SITE NOS
70701 PRESSURE ULCER, ELBOW
70702 PRESSURE ULCER, UPR BACK
70703 PRESSURE ULCER, LOW BACK
70704 PRESSURE ULCER, HIP
70705 PRESSURE ULCER, BUTTOCK
70706 PRESSURE ULCER, ANKLE
70707 PRESSURE ULCER, HEEL
70709 PRESSURE ULCER, SITE NEC

- STRATUM_PNEUMONIA (previously Stratum B):
  - Codes added to Denominator: 481 – PNEUMOCOCCAL PNEUMONIA [STREPTOCOCCUS PNEUMONIAE PNEUMONIA
  - Codes added to Denominator exclusions: ICD-9-CM Lung cancer procedure codes for thoracoscopic surgery (3230, 3241, 3250)
- STRATUM_DVT (previously Stratum A):
  - Codes removed from Denominator: 45342 (AC DVT/EMB DISTL LOW EXT),
- STRATUM_GI_HEM (previously Stratum E)

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Number of deaths (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator.

S.5. Time Period for Data *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
For users with a complete all-payer sample of hospital discharge, the recommended time period is two years for measurement of hospital rates. This recommendation is based on testing of reliability of the measure; this reliability testing is specific to all-payer hospital populations. Reliability estimates often vary when the measure is applied to other hospital populations, such as Medicare-only populations. Reliability is sensitive to numerator and denominator size as well as the distribution of hospital rates. For populations other than all-payer hospital populations fewer or more than 2 years of data may be recommended, depending on changes in reliability estimates. Note that the signal variance parameters embedded in the AHRQ QI software assume at least a one-year time period. Users may use longer time periods if desired.

S.6. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Please see attached excel file in S.2b. for version 6.0 specifications.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Surgical discharges, for patients ages 18 through 89 years or MDC 14 (pregnancy, childbirth, and puerperium), with all of the following:

- any-listed ICD-9-CM or ICD-10-PCS procedure codes for an operating room procedure; and
- the principal procedure occurring within 2 days of admission or an admission type of elective (ATYPE=3); and
- meet the inclusion and exclusion criteria for STRATUM_SHOCK (shock or cardiac arrest), STRATUM_SEPSIS (sepsis), STRATUM_PNEUMONIA (pneumonia), STRATUM_DVT (deep vein thrombosis or pulmonary embolism), or STRATUM_GI_HEM (gastrointestinal hemorrhage or acute ulcer)

STRATUM_SHOCK (shock or cardiac arrest)

- any secondary ICD-9-CM or ICD-10-CM diagnosis codes or any-listed ICD-9-CM or ICD-10-PCS procedure codes for shock or cardiac arrest

STRATUM_SEPSIS (sepsis)

- any secondary ICD-9-CM or ICD-10-CM diagnosis codes for sepsis.

STRATUM_PNEUMONIA (pneumonia)

- any secondary ICD-9-CM or ICD-10-CM diagnosis codes for pneumonia or pneumonitis.

STRATUM_DVT (deep vein thrombosis or pulmonary embolism)
- any secondary ICD-9-CM or ICD-10-CM diagnosis codes for deep vein thrombosis or pulmonary embolism.

STRATUM_GI_HEM (gastrointestinal hemorrhage or acute ulcer)
- any secondary ICD-9-CM or ICD-10-CM diagnosis codes for gastrointestinal hemorrhage or acute ulcer.

Surgical discharges are defined by specific MS-DRG codes and ICD-9-CM/ICD-10-PCS codes indicating "major operating room procedures."

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Please see attached excel file in S.2b. for v6.0 specifications.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exclude cases:
- transferred to an acute care facility (DISP = 2)
- with missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing)

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Please see attached excel file in S.2b. for v6.0 specifications.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Please see attached excel file in S.2b. for v6.0 specifications.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age (in 5-year age groups, except for the youngest age range), Modified Diagnosis Related Groups (ie. MS-DRGs without any distinction for "comorbidity and complications" (CC/MCC), Elixhauser Comorbidity Index (https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp), Major Diagnosis Categories (MDC) based on the principal diagnosis, and transfer in from another acute care hospital. A parsimonious model was identified using a backward stepwise selection procedure with bootstrapping. 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). The risk-adjusted rate for the overall PSI 04 is calculated as the observed to expected ratio multiplied by the reference population rate, where the observed and expected values are summed across five strata (categories) of PSI 04 risk. This approach differs from other AHRQ Patient Safety Indicators without strata, in that each discharge-record's expected value is computed using one of five distinct stratum-specific risk adjustment models that correspond to an assigned PSI 04 stratum. The five PSI 04 strata group records together based on secondary diagnoses that represent complications of care, and place the patient at risk of death (which is the numerator of PSI 04).

Additional information on methodology can be found in the Empirical Methods document on the AHRQ Quality Indicator website (www.qualityindicators.ahrq.gov). The Empirical Methods are also attached in the supplemental materials.

The specific covariates for this measure are provided for each Stratum as part of the Technical Specifications attached to section S.2b.
Source: http://www.qualityindicators.ahrq.gov/Modules/psi_resources.aspx

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
The observed rate is the number of discharge records where the patient experienced the PSI adverse event divided by the number of discharge records at risk for the event. The expected rate is a comparative rate that incorporates information about a reference population that is not part of the user's input dataset – what rate would be observed if the expected level of care observed in the reference population and estimated with risk adjustment regression models, were applied to the mix of patients with demographic and comorbidity distributions observed in the user's dataset. The expected rate is calculated only for risk-adjusted indicators.

The following descriptions are for the expected rate and risk-adjusted rate. These rates are calculated using models for each individual stratum.

The expected rate is estimated using the stratum specific model for each record using a generalized estimating equations (GEE) approach to account for correlation at the hospital or provider level. Records are assigned to the stratum for which they qualify with the highest observed mortality rate.

The risk-adjusted rate is a comparative rate that also incorporates information about a reference population that is not part of the input dataset – what rate would be observed if the level of care observed in the user's dataset were applied to a mix of patients with demographics and comorbidities distributed like the reference population? The risk-adjusted rate for the overall PSI 04 is calculated as the observed to expected ratio multiplied by the reference population rate, where the observed and expected values are summed across five strata (categories) of PSI 04 risk. This approach differs from other AHRQ Patient Safety Indicators without strata, in that each discharge-record's expected value is computed using one of five distinct stratum-specific risk adjustment models that correspond to an assigned PSI 04 stratum. The five PSI 04 strata group records together based on secondary diagnoses that represent complications of care, and place the patient at risk of death (which is the numerator of PSI 04).

The smoothed rate is the weighted average of the risk-adjusted rate from the user's input dataset and the rate observed in the reference population; the smoothed rate is calculated with a shrinkage estimator to result in a rate near that from the user's dataset if the provider's rate is estimated in a stable fashion with minimal noise, or to result in a rate near that of the reference population if the variance of the estimated rate from the input dataset is large compared with the hospital-to-hospital variance estimated from the reference population. Thus, the smoothed rate is a weighted average of the risk-adjusted rate and the reference population rate, where the weight is the signal-to-noise ratio. In practice, the smoothed rate brings rates toward the mean, and tends to do this more so for outliers (such as rural hospitals).

For additional information, please see the supplemental materials for the AHRQ QI Empirical Methods.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Exclude cases with missing gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing), or principal diagnosis (DX1=missing). Missingness on these variables, in aggregate, almost never exceeds 1% of eligible records.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
While the measure is tested and specified using data from the Healthcare Cost and Utilization Project (HCUP) (see section 1.1 and 1.2 of the measure testing form), the measure specifications for numerators, denominators and observed rates and software are specified to be used with any ICD-9-CM- or ICD-10-CM/PCS coded administrative billing/claims/discharge dataset. Software to calculate risk-adjusted and smoothed rates is available for ICD-9-CM only. One year of ICD-10-CM/PCS coded data is necessary before risk adjustment will be available for ICD-10-CM/PCS versions of the software.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PSI04_Measure_Testing_Form_160615.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0351
Measure Title: Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04)
Date of Submission: 5/31/2016

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in**

performance;
OR
there is evidence of overall less-than-optimal performance.
2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.
2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements 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 of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. 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 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

All analyses were completed using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), 2011-2013. HCUP is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ).[1] HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. The HCUP SID contain the universe of the inpatient discharge abstracts in participating States, translated into a uniform format to facilitate multi-State comparisons and analyses. All states provide data for community hospitals and together, the SID encompasses about 97 percent of all U.S. community hospital discharges. For the analyses presented here, we use 34 states representing about 89 percent of the U.S. community hospital discharges, for a total of about 30 million hospital discharges from community hospitals. As defined by the American Hospital Association, community hospitals are all non-Federal, short-term, general or other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are public and academic medical centers, specialty hospitals such as obstetrics–gynecology, ear–nose–throat, orthopedic and pediatric institutions. Short-stay rehabilitation, long-term acute care hospitals are excluded from the data used for the reported analyses.

Each of the 34 states included in the dataset report information about whether a diagnosis was present on admission (POA) and information on the timing of procedures during the hospitalization. POA data[2] is important to distinguish complications that occur in-hospital from diagnoses that existed prior to hospitalization. Edit checks on POA were developed using a separate analysis of HCUP databases that examined POA coding in the 2013 SID at hospitals that were required to report POA to CMS. The edits identify general patterns of suspect reporting of POA. The edits do not evaluate whether a valid POA value (e.g., Y or N) is appropriate for the specific diagnosis. There are three hospital-level edit checks:

1. Indication that a hospital has POA reported as Y on all diagnoses on all discharges
2. Indication that a hospital has POA reported as missing on all non-Medicare discharges
3. Indication that a hospital reported POA as missing on all nonexempt diagnoses for 15 percent or more of discharges. The cut-point of 15 percent was determined by 2 times the standard deviation plus the mean of the percentage for hospitals required to report POA to CMS.

Hospitals that failed any of the edit checks were excluded from the dataset.
The SID data elements include International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) coded principal and secondary diagnoses and procedures, additional detailed clinical and service information based on revenue codes, admission source and discharge status, patient demographics, expected payment source (Medicare, Medicaid, private insurance as well as the uninsured), total charges and length of stay (www.hcup-us.ahrq.gov).

1.3. What are the dates of the data used in testing?

HCUP data included calendar years 2011-2013. Further explanation of the years used for each analysis are in section 1.7.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency

[1] HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)
[2] Present-on -Admission was added as a data element to the uniform bill form (UB-04) effective October 1, 2007, and hospitals incurred a payment penalty for not including POA on Medicare records beginning October 1, 2008. Each of the several diagnoses in a discharge record can be flagged as "present at the time the order for inpatient admission occurs" or not (see http://www.cdc.gov/nchs/icd/icd9cm_addenda_guidelines.htm).

☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Table 1a. Reference Population Observed Rate for Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04), 2011-2013

Overall Reference Population Rate				
Year[2]	Number of Hospitals	Outcome of Interest (Numerator)[1]	Population at Risk (Denominator)[1]	Observed Rate Per 1000 Surgical Discharges[1]
2013	2,783	21,242	182,512	116.3869
2012	2,860	21,897	185,872	117.8069
2011	2,748	21,403	181,317	118.0419

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

[1]The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population of all hospitals included in the reference population data (denominator).

[2]Reference population is limited to states with present on admission data (POA). Since many states did not report POA data prior to 2011 we have not included testing prior to 2011.

Table 1b. Distribution of Hospital Performance for Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04) in 2-year Pooled Data (2011-2012, 2012-2013)[1]

Distribution of Hospital-level Observed Rates in Reference Population								
Year[3]	Number of Hospitals	Rates per 1000 Surgical Discharges (p=percentile)[2]						
		Mean	SD[2]	p5	p25	Median	p75	p95
2011-2012	3,212	103.83	85.55	0.00	58.06	102.61	140.35	217.39
2012-2013	3,398	100.76	88.28	0.00	51.28	99.92	137.25	212.12

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

[1]Consistent with the recommended minimum reporting time period, results are presented for data combining 2 years of data: 2011 and 2012, 2012 and 2013. All data from 2012 are included in both time periods reported. Limitations in present on admission data (POA) data availability (see below) do not allow for use of earlier years.

[2]The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals in the dataset with at least one case in the denominator, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile. Standard deviation refers to the spread in observed values in relation to the mean.

[3]Reference population is limited to states with present on admission data (POA). Since many states did not report POA data prior to 2011 we have not included testing prior to 2011.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

See 1.5 (Table 1a)

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

For tests requiring hospital rates we combine 2 years of hospital data prior to calculating rates and testing the measure (termed "2-year pooled data"). The tests that used pooled 2012 and 2013 data include: reliability testing (Table 2), validity testing (described in text, section 2.b) and performance discrimination (Table 7). The hospital rate distributions (Table 1b) are reported for two 2-year pooled data periods, 2011 – 2012 and 2012 – 2013. All other tests that do not use hospital rates are calculated using 2013 data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Age and sex were the only patient sociodemographic characteristics that were available and analyzed in the data used for measure development and testing. Many of the HCUP SID include race/ethnicity, and all of the HCUP SID include the primary expected source of payment and zip code of residence, which could be used to capture socioeconomic characteristics at an ecological (community) level. While these variables were used to assess disparities at the national level, these variables were not used in the current risk adjustment model, based on our conceptual description (i.e., logical rationale or theory informed by literature and content experts) of the causal pathway between these factors, patient clinical factors, quality of care, and outcome, described in Section 2b4.3 below.

2a2. RELIABILITY TESTING

Note: *If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

The AHRQ QIs use signal-to-noise ratios to assess reliability**.** The signal-to-noise ratio is a measure of reliability that is calculated at the hospital level and then summarized across the entire population of US hospitals. It compares the degree to which risk adjusted rates differ across hospitals (the signal) to the degree of precision of the rates within hospitals (the noise). This metric is a stringent measure of reliability that takes into account the observed distribution of risk adjusted rates within a reference population. An indicator with a low signal-to-noise ratio may not be able to distinguish differences in performance among hospitals, or may identify differences inconsistently within the same time period. An indicator with a high signal-to-noise ratio will be more likely to consistently distinguish performance differences among hospitals (e.g. one hospital performs better than others).

The signal-to-noise ratio is estimated for each hospital. The overall signal-to-noise estimate is an average of hospital-level signal to noise ratios weighted by a value of one divided by the signal plus the hospital's noise for PSI 04. Hospitals with smaller denominators (the number of patients at risk) will have lower weight, and less influence on the overall signal-to-noise ratio, because of higher noise. Weighting reduces the influence of hospitals that have less reliable rates due to very small denominators (the number of patients at risk) on the overall signal-to-noise ratio estimate.

Because the signal-to-noise ratio quantifies the ability to consistently discriminate one hospital's performance from the other hospitals in the population, it is sensitive to the distribution of hospital sizes as well as the distribution of risk-adjusted rates in the reference population. If the hospitals in a population all have performance in a narrow range (low signal), it is more difficult to reliably distinguish among hospitals' performance than when hospital performance is spread out over a much wider range (high signal). For example, if all hospitals have nearly perfect performance, it will be impossible to distinguish among them. As a consequence, if the distribution of hospital rates changes over time, or if the measured population is restricted (e.g. Medicare patients), or if a different subset of hospitals is included, the signal-to-noise ratio will also change.

There is no universally accepted threshold of "adequate" signal to noise ratio. Different methods of calculating reliability and signal-to-noise (e.g., split sample or test-retest reliability of the data, different methods of calculating the hospital signal-to-noise ratio) result in different distributions of reliability scores. In addition, "adequate" depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8, using hospital signal-to-noise ratios as an indicator of reliability. To account for the uncertainty (noise) in a hospital's performance due to low volume, a longer period of data can be used or smoothed rates can be calculated.

For reference, the following text in black was previously submitted to NQF. Most of the information is outdated. 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 inhospital 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]

AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million discharges

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] Hude Quan, MD, PhD; Saskia Drösler, 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, Shepheard 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

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.g., *percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 2 shows the most recent reliability testing for PSI 04.

Table 2. Signal-to-Noise Ratio by Hospital Size Decile for Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04) in 2-year Pooled Data (2012-2013)

Hospital Size Decile	Number of Hospitals	Avg. Number of Discharges per Hospital in Decile	Avg. Signal-to-Noise Ratio for Hospitals in Decile
1 (smallest)	319	5.0	0.0579
2	320	11.9	0.1063
3	320	22.9	0.1732
4	320	37.4	0.2279
5	320	56.2	0.3094
6	320	81.6	0.3954
7	320	113.8	0.4738
8	320	157.2	0.5582
9	320	226.5	0.6464
10 (largest)	320	437.5	0.7765
Overall	3,199	115.0	0.6040

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.e., *what do the results mean and what are the norms for the test conducted?*)

For all-payer populations and across all hospitals in the AHRQ QI POA reference population, the overall signal to noise ratio for this indicator is moderate to good with an overall signal-to-noise ratio of 0.60. Hospitals with more than 82 qualifying discharges on average have risk adjusted rates with moderate to high reliability (average signal-to-noise ratio of 0.40 to 0.78). Signal-to-noise ratios were smaller for hospitals with fewer than approximately 82 qualifying discharges per year (average signal-to-noise ratio less than 0.40). Smoothed rates, which are recommended for all hospitals (and are implemented in the AHRQ software), address reliability concerns particularly for small hospitals.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

In this section, in addition to the empirical analyses completed, we summarize the most relevant literature, and provide a full evidence summary in the attached Evidence Form.

Empirical Validity Analyses

As part of the original testing and NQF endorsement process in 2007, we collaborated with Silber et al. to compare PSI 04 with other specifications of the failure-to-rescue concept, including related measure NQF 0352 ("Failure to Rescue In-Hospital Mortality (risk adjusted)").[3] This analysis was based on Medicare inpatient fee-for-service claims for general surgical admissions from July 1, 1999 through June 30, 2000, linked to the 2000 American Hospital Association Annual Survey. This data set included 1,467 hospitals and 403,679 Medicare beneficiaries between 65 and 90 years of age. To assess construct validity, we estimated logit models using detailed patient characteristics and 5 hospital characteristics shown to be associated with better quality of care in previous studies: (1) teaching status (member of the Council of Teaching Hospitals); (2) high technology status (open heart surgery or organ transplantation); (3) size greater than 200 beds; (4) bed-to-nurse ratio (RN plus LPN FTE positions); and (5) nursing skill mix ratio [RN/(RN + LPN)]. We report both "marginal" and "partial" results for each regression; marginal analyses used one hospital characteristic at a time along with all patient characteristics, whereas partial analyses adjusted for all hospital and patient variables simultaneously. Finally, the omega statistic represents the ratio of the squared sum of the log odds for patient characteristics at the discharge-level variables divided by the corresponding quantity for hospital-level variables. All else equal, outcome measures that have lower omega ratios may be more desirable quality indicators, because the lower the omega, the greater the hospital's impact on the outcome relative to the patient's impact.

In more recent analyses, we confirmed the association between teaching status and risk-adjusted PSI 04 rates using the 2012 and 2013 HCUP SID reference data set described above. These analyses used a broader definition of teaching status, as implemented in the HCUP program: "a hospital is considered to be a teaching hospital if it has an American Medical Association (AMA)-approved residency program, is a member of the Council of Teaching Hospitals, or has a ratio of full-time equivalent interns and residents to beds of .25 or higher."

Systematic Assessment of Face Validity

We utilized a structured panel review to evaluate face validity (from a clinical perspective) of the Patient Safety Indicators. The panels were convened in 2002. It is anticipated that the results of face validity review would be similar if panels were convened in more recent years, given that the clinical characteristics of these events, treatment and prevention approaches, and sequelae have not changed substantially since 2002. The clinical panel review process was based on the RAND appropriateness method, a modified Delphi process also known as a nominal group technique.

Twenty-one professional clinical organizations were invited to submit nominations. These organizations were selected based on the applicability of the specialty or subspecialty to potential Patient Safety Indicators. Clinical areas represented by the panels included internal medicine, cardiology, radiology, geriatrics, surgical and critical care nursing, anesthesiology, pharmacy, inpatient medicine and surgery (including thoracic, neurology, orthopedic, colorectal, urology, spine, and transplant surgical subspecialties). For assignments to each panel, a list of applicable specialties was identified for the indicators to be evaluated by that panel. Panelists were selected so that each panel had diverse membership in terms of practice characteristics and setting. For PSI 04, 7 members of a multispecialty panel completed the evaluation in full. Additional details of panel composition are available online at http://archive.ahrq.gov/clinic/tp/hospdatp.htm.

Panelists completed a 10-item questionnaire, tailored to each specific indicator. Following the initial rating of the indicators, panelists participated in a moderated 90-minute conference call, where opinions about the indicators were discussed. The panelists then completed the same 10-item questionnaire again, and submitted their final ratings. Ratings were summarized in accordance with the RAND Appropriateness Method.[4]

---

[3] Silber JH, Romano PS, Rosen AK, Wang Y, Ross RN, Even-Shoshan O, Volpp K. Failure-to-rescue: Comparing definitions to measure quality of care. Med Care 2007; 45:918-925.

[4] McDonald KM, Romano PS, Geppert J, Davies SM, Duncan BW, Shojania KG. Measures of Patient Safety Based on Hospital Administrative Data: The Patient Safety Indicators. Technical Review Number 5. Rockville, MD: Agency for Healthcare Research and Quality, 2002

For reference, the following text in black was previously submitted to NQF. Most of the information is outdated. 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]

The Agency for Healthcare Research and Quality Patient Safety Indicators (PSIs) were used to identify 14 PSIs among 161,004 surgeries. [5]

A likelihood ratio test of the hypothesis that the coefficients on all of these variables were equal to 0 (lambda) = 35.3, p< .01). [1]

We used propensity score matching and multivariate regression analyses to predict expenditures and outcomes attributable to the 14 PSIs. [5]

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]

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 Demen. 2005 Jan-Feb;20(1):27-36. PMID: 15751451.

[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.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Performance Measure Score

As summarized in the Evidence Form, numerous studies have linked failure to rescue measures, including PSI 04, to structure and process measures. Multiple studies have found lower FTR rates in hospitals with higher nurse-to-bed ratios[5,6,7,8,9,10,11,12], better nurse skill mix ratios[13,14,15,16,17,18,19], and better US-trained nurse ratios.[20]

[5] Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA 2002; 288:1987-1993.

[6] Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. N Engl J Med 2002; 346:1715-1722.

[7] Friese CR, Aiken LH. Failure to rescue in the surgical oncology population: implications for nursing and quality improvement. Oncol Nurs Forum 2008; 35:779-785. PMC2562164.

[8] Ghaferi AA, Osborne NH, Birkmeyer JD, Dimick JB. Hospital characteristics associated with failure to rescue from complications after pancreatectomy. J Am Coll Surg 2010; 211:325-330.

[9] Needleman J, Buerhaus PI, Vanderboom C, Harris M. Using present-on-admission coding to improve exclusion rules for quality metrics: the case of failure-to-rescue. Med Care 2013; 51:722-730.

Higher hospital volume was associated with lower FTR rates in at least 6 studies.[21] In addition, studies have found that hospitals with the highest patient satisfaction scores and hospitals with better compliance with NQF Safe Practices had lower risk adjusted odds of FTR.[22,23]

Empirical Validity Analyses

In the marginal analysis, we report the odds ratio and p value for each structural measure of quality when all patient characteristics are included in the model with only one hospital characteristic at a time. In the partial analysis, we report the same odds ratios and p values, using all patient discharge-level and hospital variables simultaneously. The latter approach may be more difficult to interpret due to collinearities among hospital characteristics. Teaching hospitals demonstrated odds ratios of 0.807 and 0.852 in marginal and partial analyses, respectively ($p<0.0001$ for both).

High technology hospitals demonstrated odds ratios of 0.924 ($p<0.005$) and 1.049 (NS) in marginal and partial analyses, respectively.

Large hospitals (>200 beds) demonstrated odds ratios of 0.872 ($p<0.0001$) and 0.917 ($p<0.01$) in marginal and partial analyses, respectively.

Less well staffed hospitals (with one additional bed per licensed nurse FTE) demonstrated odds ratios of 1.108 ($p<0.005$) and 1.044 (NS) in marginal and partial analyses, respectively.

Hospitals with better nursing skill mix (100% RN) demonstrated odds ratios of 0.832 and 0.870 in marginal and partial analyses, respectively ($p<0.0001$ for both).

The omega ratio summarizing the contribution of patient characteristics at the discharge-level versus hospital-level variables for PSI 04 was 57, compared with omega ratios of 189 for the overall risk-adjusted surgical mortality rate and 128 for NQF 0352.

We used all-payer data from 34 states in 2012-2013, described above, to confirm the association between hospital teaching status and lower PSI 04 rates. In these analyses, a much broader definition of teaching status was used, capturing not just COTH members but all hospitals with approved residency programs or more than

---

[10] Clarke SP, Aiken LH. Failure to rescue. Am J Nurs 2003; 103:42-47

[11] Schmid A, Hoffman L, Happ MB, Wolf GA, DeVita M. Failure to rescue: a literature review. J Nurs Adm 2007; 37:188-198.

[12] Park SH, Blegen MA, Spetz J, Chapman SA, De Groot H. Patient turnover and the relationship between nurse staffing and patient outcomes. *Res Nurs Health.* 2012;35(3):277-288.

[13] Blegen, M. A., et al. (2013). "Baccalaureate education in nursing and patient outcomes." J Nurs Adm 43(2): 89-94.

[14] Kendall-Gallagher D., Aiken L.H., Sloane D.M., and Cimiotti J.P.: Nurse specialty certification, inpatient mortality, and failure to rescue. J Nurs Scholarsh 2011; 43: pp. 188-194

[15] Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational levels of hospital nurses and surgical patient mortality. JAMA 2003; 290:1617-1623. PMC3077115.

[16] Silber JH, Romano PS, Rosen AK, Wang Y, Even-Shoshan O, Volpp KG. Failure-to-rescue: comparing definitions to measure quality of care. *Med Care.* 2007;45(10):918-925.

[17] Kendall-Gallagher D, Aiken LH, Sloane DM, Cimiotti JP. Nurse specialty certification, inpatient mortality, and failure to rescue. J Nurs Scholarsh 2011; 43:188-194. PMC3201820.

[18] Needleman J, Buerhaus PI, Vanderboom C, Harris M. Using present-on-admission coding to improve exclusion rules for quality metrics: the case of failure-to-rescue. Med Care 2013; 51:722-730.

[19] Seago JA, Williamson A, Atwood C. Longitudinal analyses of nurse staffing and patient outcomes: More about failure to rescue. J Nurs Adm 2006; 36:13-21.

[20] Neff, D. F., et al. (2013). "Utilization of non-US educated nurses in US hospitals: implications for hospital mortality." Int J Qual Health Care 25(4): 366-372.

[21] Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. N Engl J Med 2002; 346:1715-1722.

[22] Sacks GD, Lawson EH, Dawes AJ, et al. Relationship Between Hospital Performance on a Patient Satisfaction Survey and Surgical Quality. *JAMA Surg.* 2015;150(9):858-864.

[23] Brooke BS, Dominici F, Pronovost PJ, Makary MA, Schneider E, Pawlik TM. Variations in surgical outcomes associated with hospital compliance with safety practices. *Surgery.* 2012;151(5):651-659.

0.25 residents per bed. In unadjusted analyses, teaching hospitals demonstrated a risk ratio of 1.131 (12.29% versus 10.86%). In adjusted analyses, using the current V6 risk-adjustment model, this risk ratio reversed to 0.976 (11.46% versus 11.74%). In adjusted analyses, using the proposed V7 risk-adjustment model, this risk ratio further improved to 0.975 (11.45% versus 11.75%).

Systematic Assessment of Face Validity

The multi-specialty Panel and Surgical Panel both rated the indicator as acceptable on overall usefulness as an indicator of quality of care.

Table 4. Clinician Panel Evaluations of the Face Validity for Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04) [4]

Multi-specialty Panel (MSP) Evaluation		
Overall Rating[1]	**Agreement[2]**	**Acceptability[3]**
7	Indeterminate	Acceptable

[1]Median panel overall rating of the indicator on a scale from 1 to 9, with the higher rating indicating better measurement
[2]Level of agreement, where "agreement" corresponds to little dispersion of opinion, "indeterminate" means that the opinion ranged but did not reach the point of clear "disagreement", the final category where there were panelists with diametrically different opinions
[3]"Acceptable" indicates that the indicator was rated as useful by almost all panelists. "Acceptable (-)" indicates that the indicator was rated as useful by most panelists, although a few rated it as less useful (but not as poor). "Unclear" indicates that panelists rated the usefulness of the indicator as moderate. For further details of methods, see http://archive.ahrq.gov/clinic/tp/hospdatp.htm
[4]PSI 04 was evaluated under a previous name (i.e. Failure to Rescue).

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The consistent associations with structural measures of hospital quality, including higher nurse staffing, better nursing skill mix, higher hospital volume (beds), and teaching status, suggest that PSI 04 is a valid measure. These findings are supported by other studies (summarized in the Evidence Form) that showed lower PSI 04 or failure-to-rescue rates at hospitals with better patient satisfaction and higher adherence to NQF Safe Practices. Teaching hospitals had higher unadjusted PSI 04 rates, but lower adjusted PSI 04 rates, relative to nonteaching hospitals. However, this effect was less pronounced with a more inclusive definition of teaching hospitals (and all-payer data instead of Medicare data).

PSI 04 has acceptable face/content validity based on clinical panel evaluation.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Empirical Evaluation of Exclusions: Using the 2013 data from 34 states, we examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Table 5 shows the results of the most recent exclusions analysis.

Table 5. Number and Percent of Discharges Excluded, by Denominator Exclusion Criteria, for Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04) [1]

PSI 04	Denominator			Numerator		
Exclusion Name	**Exclusion Count**	**After Exclusions**	**% Change**	**Exclusion Count**	**After Exclusions**	**% Change**
No Exclusions applied		**329,716**			**37,698**	
Exclude Transfers to an acute care facility	9,889	319,827	3.0%	0	37,698	0.0%
Stratum: PE/DVT						
Exclude principal dx PE/DVT	884	328,832	0.3%	7	37,691	0.0%
Exclude principal dx Abortion-Related and Postpartum Obstetric Pulmonary Embolism	6	329,710	0.0%	0	37,698	0.0%
Stratum: Pneumonia						
Exclude principal dx Pneumonia	184	329,532	0.1%	12	37,686	0.0%
Exclude principal dx respiratory Complications	126	329,590	0.0%	0	37,698	0.0%
Exclude Diagnosis Viral Pneumonia	343	329,373	0.1%	17	37,681	0.0%
Exclude Dx of Immunocompromised	8,696	321,020	2.6%	632	37,066	1.7%
Exclude Procedure of Immunocompromised	172	329,544	0.1%	1	37,697	0.0%
Exclude MDC 4	9,006	320,710	2.7%	428	37,270	1.1%
Exclude Lung Cancer Procedure	47	329,669	0.0%	4	37,694	0.0%
Stratum Sepsis						
Exclude principal dx Septicemia	45,202	284,514	13.7%	3,381	34,317	9.0%
Exclude principal dx of Infection	14,982	314,734	4.5%	843	36,855	2.2%
Exclude Diagnosis of Immunocompromised	4,830	324,886	1.5%	1,050	36,648	2.8%
Exclude Procedure of Immunocompromised	44	329,672	0.0%	5	37,693	0.0%
Exclude Length of Stay	1,476	328,240	0.4%	703	36,995	1.9%

Less than 4						
Stratum: Shock/Cardiac Arrest						
Exclude principal dx Shock	99	329,617	0.0%	55	37,643	0.1%
Exclude principal dx Trauma	0	329,716	0.0%	0	37,698	0.0%
Exclude principal dx Hemorrhage	453	329,263	0.1%	81	37,617	0.2%
Exclude principal dx GI Hemorrhage	3,620	326,096	1.1%	2,239	35,459	5.9%
Exclude principal dx Abortion-Related Shock	947	328,769	0.3%	175	37,523	0.5%
Exclude MDC 4 or 5	27,819	301,897	8.4%	6,611	31,087	17.5%
Stratum GI hemorrhage / Acute ulcer						
Exclude principal dx GI Hemorrhage-Acute Ulcer	714	329,002	0.2%	12	37,686	0.0%
Exclude principal dx Blood Loss / Anemia	141	329,575	0.0%	0	37,698	0.0%
Exclude principal dx Trauma	3,117	326,599	0.9%	78	37,620	0.2%
Exclude principal dx Alcoholism	354	329,362	0.1%	11	37,687	0.0%
Exclude MDC 6 or 7	14,134	315,582	4.3%	128	37,570	0.3%
All Exclusions applied		**182,429**			**21,226**	

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

[1]This indicator does not have numerator exclusion criteria.

[2]Potential numerator cases are those that would have qualified for the numerator if not for a particular denominator exclusion criterion.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The stratum-specific exclusions are meant to exclude cases for which the "complication" was actually the principal reason for admission or the primary indication for surgery. Some exclusions (e.g. immunocompromised state) are intended to exclude patients for whom death may be the expected outcome (i.e., less preventable). For example, patients presenting with acute hemorrhagic shock due to ongoing blood loss are unlikely to survive.

All patients transferred to other hospitals must be excluded from the analysis because the relevant outcome of these patients (i.e., dead or alive at the time of discharge from the acute inpatient setting) cannot be ascertained without social security numbers or other data elements to support linkage.

Although many of these exclusions are rare, they ensure face validity. Note that AHRQ does NOT exclude patients with complications that were present on admission (i.e., upon transfer from another hospital, emergency department, or ambulatory surgery center). This decision was based on Needleman et al.'s analysis, showing

that PSI 04 mortality was higher among patients with hospital-acquired complications than among patients with present-on-admission complications (22% versus 13%), and that nurse staffing (licensed nurse hours per bed) and skill mix were highly associated with PSI 04 rates, regardless whether complications present on admission were included or excluded.[24]

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 362 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Clinical Factors

For each PSI 04 stratum, we considered a standard set of covariates grouped into four categories: demographics, severity of illness, comorbidities and transfer-in status. Covariates that were considered as potential risk adjusters included gender and age (in mutually exclusive 5-year age categories), Major Diagnostic Categories (MDCs), Modified Diagnostic Related Groups (MDRGs) (defined as the base MS-DRG without comorbidity or complication distinctions), AHRQ Comorbidity Software categories, and whether the patient was transferred in from another facility. Only those covariates present in at least 30 records for that PSI 04 strata are retained. A parsimonious model was identified using backward stepwise selection with bootstrapping.

The omitted covariate within mutually exclusive categories is the reference group for those categories. Reference categories are usually 1) the most common and/or 2) the least risk. The choice of omitted reference category does not affect predicted probabilities or model performance.

For the MDRGs, the risk reported is the residual risk after adjustment for the MDC to which the MDRG belongs. Likewise, the risk reported for MDCs represents the average risk of all MSDRGs in that MDC not included in the model.

The risk-adjusted rate for the overall PSI 04 is calculated as the observed to expected ratio multiplied by the reference population rate, where the observed and expected values are summed across five strata (categories) of PSI 04 risk. This approach differs from other AHRQ Patient Safety Indicators without strata, in that each discharge-record's expected value is computed using one of five distinct stratum-specific risk adjustment models that correspond to an assigned PSI 04 stratum. The five PSI 04 strata group records together based on secondary diagnoses that represent complications of care, and place the patient at risk of death (which is the numerator of PSI 04).

Additional details are available in the *AHRQ Quality Indicator Empirical Methods* document, included in the supplemental file and available on the AHRQ QI website.

[24] Needleman J, Buerhaus PI, Vanderboom C, Harris M. Using present-on-admission coding to improve exclusion rules for quality metrics: the case of failure-to-rescue. *Med Care.* 2013;51(8):722-730.

Sociodemographic Factors

There is no evidence or causal model to suggest that socioeconomic factors are associated with death following serious surgical complications independent of quality of care, or are mediated by pre-hospital care (which may not fall within the proper realm of hospital accountability). Accordingly, consistent with the guidance provided by NQF in the SDS Trial Period FAQs, AHRQ believes that it would be inappropriate to include other SDS variables in the risk-adjustment approach for PSI 04, which is an in-hospital outcome measure.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

This section includes a summary of the selected risk factors for each stratum, used together to construct the risk model for the overall PSI 04 measure. Details of the current risk adjustment coefficients for each PSI 04 stratum can be found in the attached technical specifications.

STRATUM_SHOCK: The risk model includes 75 risk categories, including 24 age-gender categories in 5-year age categories between ages 30 and 89, and 2 age-gender categories below age 30 (i.e. 18-29), transfer in from another acute care facility and 14 comorbidities. The remainder of selected risk factors account for the reason for admission and the type of surgery that was performed during the hospitalization, including MDC and MS-DRGs collapsed to remove Complication or Comorbidity/ Major Complication or Comorbidity (CC/MCC) distinctions.

STRATUM_SEPSIS: The risk model includes 81 risk categories, including 24 age-gender categories in 5-year age categories between ages 30 and 89, and 2 age-gender categories below age 30 (i.e. 18-29), transfer in from another acute care facility and 18 comorbidities. The remainder of selected risk factors account for the reason for admission and the type of surgery that was performed during the hospitalization, including MDC and MS-DRGs collapsed to remove Complication or Comorbidity/ Major Complication or Comorbidity (CC/MCC) distinctions.

STRATUM_PNEUMONIA: The risk model includes 89 risk categories, including 24 age-gender categories in 5-year age categories between ages 30 and 89, and 2 age-gender categories below age 30 (i.e. 18-29), transfer in from another acute care facility and 22 comorbidities. The remainder of selected risk factors account for the reason for admission and the type of surgery that was performed during the hospitalization, including MDC and MS-DRGs collapsed to remove Complication or Comorbidity/ Major Complication or Comorbidity (CC/MCC) distinctions.

STRATUM_DVT: The risk model includes 56 risk categories, including 24 age-gender categories in 5-year age categories between ages 30 and 89, and 2 age-gender categories below age 30 (i.e. 18-29), transfer in from another acute care facility and 10 comorbidities. The remainder of selected risk factors account for the reason for admission and the type of surgery that was performed during the hospitalization, including MDC and MS-DRGs collapsed to remove Complication or Comorbidity/ Major Complication or Comorbidity (CC/MCC) distinctions.

STRATUM_GI_HEM: The risk model includes 61 risk categories, including 24 age-gender categories in 5-year age categories between ages 30 and 89, and 2 age-gender categories below age 30 (i.e. 18-29), transfer in from another acute care facility and 15 comorbidities. The remainder of selected risk factors account for the reason for admission and the type of surgery that was performed during the hospitalization, including MDC and MS-DRGs collapsed to remove Complication or Comorbidity/ Major Complication or Comorbidity (CC/MCC) distinctions.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Not applicable (see above)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

This analysis evaluates the performance of the risk adjustment model(s) with respect to the event of interest (i.e., in-hospital death). The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the c-statistic (also known as the area under a receiver operating characteristic curve). The c-statistic is computed by assigning each observation a predicted probability of the outcome from the risk-adjustment model, based on the value of the observed covariates and the parameter estimates from the risk-adjustment model. Two copies of the dataset are sorted, first from highest to lowest predicted probability and second from lowest to highest predicted probability. Random sampling is used to create a set of paired observations. Pairs that consist of one event and one non-event (discordant pairs) are kept and concordant pairs are discarded. The c-statistic represents the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the observation without the event. C-statistics above 0.70 and below 0.80 have moderate discrimination. Above 0.80, the discrimination is considered high. We did not employ common "goodness of fit" tests because these tests tend to be uninformative with large samples.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to ***2b4.9***

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

The c-statistic for the overall PSI 04 model is 0.829 in the 2013 HCUP data (described above).

Note that there are actually five distinct risk models; one for each type of complication or stratum. These five models currently have c statistics that range from 0.726 to 0.860. Enhancements now being tested for version 7 (i.e., adjusting for the severity and timing of the triggering complication, as well as all of the factors listed in 2b4.4a above) will increase these c statistics to 0.779 to 0.877. The overall c statistic reported in above represents the overall performance of all five models.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Table 6. Risk adjustment Model Discrimination and Calibration for Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04), 2013

Predicted Rate Decile	Number of Discharges per Decile	Predicted Rate (per 1,000 surgical discharges)	Observed Rate (per 1,000 surgical discharges)	Observed to Predicted Ratio
1 (lowest)	18,251	5.4184	3.0683	0.57
2	18,251	13.9581	11.4514	0.82
3	18,251	23.4806	17.4785	0.74
4	18,251	34.4656	34.3543	1.00
5	18,251	49.3254	53.2026	1.08
6	18,252	70.4652	71.8825	1.02

7	18,251	102.9786	108.9255	1.06
8	18,251	157.9172	162.8952	1.03
9	18,252	248.7309	253.7804	1.02
10 (highest)	18,251	457.1238	446.8248	0.98

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
See calibration by decile in Table 6 in 2b4.7

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The risk-adjustment model has moderately high discrimination, based on a c statistic of 0.829 (i.e., in 83% of randomly selected pairs of discordant observations, the patient who experienced PSI 04 had a higher probability of experiencing the event than the patient who did not). A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles. This indicator is well calibrated, as the observed to predicted ratio values across the deciles range between 0.74 to 1.08 for all deciles except the lowest decile. For patients with very low predicted rates, the relative difference between observed and predicted values is greater, but this is not particularly concerning due to the very small number of events that occur in this risk stratum.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

This analysis assesses the probability that a hospital is higher or lower than a benchmark or threshold, given hospital size. It reflects whether the indicator can discriminate the best performing hospitals from the lower performing hospitals.

For this analysis, "benchmark" refers to the smoothed indicator rate based on the 20th percentile of the reference population (i.e., 20% of hospitals have a lower mortality rate or better performance). "Threshold" refers to the indicator rate based on the 80th percentile (i.e., 80% have lower mortality or better performance). Assuming an underlying Gamma distribution for the smoothed rates of the measure, the benchmark and threshold values are identified using population reference rates and signal variances computed from the entire AHRQ QI POA Reference Population. Hospital-level 90% confidence limits for smoothed rates are also computed from the Gamma distribution.

The analysis is reported by size decile, based on the denominator cases, demonstrating performance across hospitals of various sizes. Each hospital is assumed to have an underlying distribution of smoothed rates that follows a Gamma distribution. The parameters of a Gamma distribution are shape and scale. For each hospital the shape is calculated as ((smoothed rate)^2/ smoothed rate variance), and the scale is calculated as (smoothed rate variance / smoothed rate). The smoothed rate variance (aka posterior variance) is calculated as the signal variance – (reliability weight * signal variance). The reliability weight is calculated as (signal variance / (signal variance + noise variance)). Hospitals are

ranked by size and grouped into 10 equal categories of size (deciles). The Benchmark and Threshold are compared to the Gamma distribution of the smoothed rates for each hospital to determine if the hospital rate is better or worse than the Benchmark and Threshold rates with 95% probability. This provides a 95% confidence interval for the Benchmark and Threshold rate.

Table 7 reports the proportion of hospitals above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The hospitals not classified as either better or worse have rates that fall within the 95% confidence interval.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 7. Performance Categories Using Smoothed Hospital Rates by Hospital Size Decile for Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI 04) in 2-year Pooled Data (2012-2013)

			Benchmark			Threshold		
Hospital Size Decile	Number of Hospitals	Average Number of Denominator Discharges Per Hospital	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
(smallest) 1	319	5.0	0.0000	0.0063	0.9937	0.0000	0.0000	1.0000
2	320	11.9	0.0000	0.0219	0.9781	0.0000	0.0000	1.0000
3	320	22.9	0.0000	0.0938	0.9063	0.0125	0.0000	0.9875
4	320	37.4	0.0000	0.1281	0.8719	0.0156	0.0000	0.9844
5	320	56.2	0.0000	0.2188	0.7813	0.0625	0.0031	0.9344
6	320	81.6	0.0000	0.3281	0.6719	0.1344	0.0000	0.8656
7	320	113.8	0.0000	0.3375	0.6625	0.2344	0.0125	0.7531
8	320	157.2	0.0031	0.4094	0.5875	0.2344	0.0031	0.7625
9	320	226.5	0.0031	0.4500	0.5469	0.3781	0.0188	0.6031
10 (largest)	320	437.5	0.0094	0.6188	0.3719	0.4344	0.0438	0.5219
Overall	3,199	115.0	0.0016	0.2613	0.7371	0.1507	0.0081	0.8412

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2012 - 2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov (AHRQ QI Software Version 6.0)

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

As hospital size increases, the discrimination also increases such that for hospitals in the largest 2 deciles the algorithm classifies 40% - 48% of hospitals against the threshold and 45%-63% of hospitals against the benchmark, based on conventional statistical criteria. Over all hospitals, using smoothed rates, this indicator has limited discrimination for identifying low or high performing hospitals; 16% of hospitals can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold) and 27% as better or worse than the benchmark (the percentage classified as either above or below the benchmark), based on conventional statistical criteria. In this example, use of smoothed rates "shrinks" the performance distribution across hospitals, which typically decreases performance discrimination. Although this means that hospitals with few inpatient post-surgical complications (the PSI 04 denominator) cannot be identified as low

or high performers unless their PSI 04 rates vary widely from the benchmark/threshold, this shrinkage approach also makes the measure less likely to classify hospitals as low performing when they are not.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

Not applicable

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not applicable

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The AHRQ QIs use frequently reported administrative data variables. PSI 04 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all hospital discharge records. The frequency of missing data for each variable is available by state and year from the AHRQ HCUP website (http://www.hcup-us.ahrq.gov/cdstats/cdstats_search.jsp).

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

For these variables, frequencies of missing data are typically less than 1% of the state database. It is unlikely that bias would occur from such a low frequency of missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Exclusion of cases with missing data for these variables is appropriate.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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 and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Because the indicator is based on readily available administrative billing and claims data and U.S. Census data, feasibility is not an issue.
The AHRQ QI software has been publicly available at no cost since 2001; Users have over ten years of experience using the AHRQ QI software in SAS and Windows.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees. Software is freely available from the AHRQ Quality Indicators website (http://www.qualityindicators.ahrq.gov/).

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website http://pub.azdhs.gov/hospital-discharge-stats/2012/AboutQualityRatings.html CareChex (Division of Quantros) http://www.carechex.com/QualityIndicators.aspx CMS Medicare Hospital Compare Program https://www.medicare.gov/hospitalcompare/Data/Measures-Displayed.html# Commonwealth Fund, Why Not the Best http://whynotthebest.org/methodology Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings Connecticut Hospital Association http://www.cthosp.org/advocacy/quality-and-patient-safety/hospital-quality-reporting-website/ Consumer Reports http://www.consumerreports.org/health/resources/pdf/how-we-rate-hospitals/How%20We%20Rate%20Hospitals.pdf HealthGrades https://d2dcgio3q2u5fb.cloudfront.net/54/98/f79cdfd84640a03792ea092f20a8/2014-patient-safety-methodology.pdf Hospital Safety Score http://www.hospitalsafetyscore.org/media/file/HospitalSafetyScore_ScoringMethodology_Spring2015_Final.pdf Iowa Healthcare Collaborative https://iowareport.ihconline.org/Public/Reports.aspx?FID=778&F1ID=0&F2ID=0&F3ID=0&CID=2&PID=4 Kentucky Cabinet for Health and Family services https://prd.chfs.ky.gov/MONAHRQ/2012/MONAHRQ/AboutQualityRatings.html Kentucky Hospital Association Quality Data http://info.kyha.com/qualitydata/psisite/SelectPSIReport.asp?IndID=PSI4&TimePeriod=5&GroupOpt=none&SortOrder=hospital&SortDir=ASC Louisiana Hospital Inform http://lahospitalinform.org/index.html Maryland Health Care Commission, MONAHRQ Website http://www.hscrc.state.md.us/documents/md-maphs/wg-meet/di/2014-03-04/MHCC%20Inpatient%20Measures%20Inventory%20QBR%20highlights.pdf Maine Health Data Organization (MHDO), MONAHRQ Website https://mhdo.maine.gov/monahrq/#/resources/AboutQualityRatings Minnesota Community Measurement http://mncm.org/wp-content/uploads/2014/02/2013-HCQR-Final-2.4.2014.pdf

	Nevada Compare Care, MONAHRQ website http://nevadacomparecare.net/MQ2014/index.html#/professional/resources/AboutQualityRatings Niagara Health Quality Coalition, New York State Hospital Report Card http://www.myhealthfinder.com/newyork15/main_byproc.php Norton Healthcare http://www.nortonhealthcare.com/QualityReport Oklahoma State Department of Health, MONAHRQ https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html South Dakota Association of Healthcare Organizations http://www.sdhospitalquality.org/search.php http://healthdata.dshs.texas.gov/Hospital/PatientSafetyQualityIndicators Texas Department of State Health Services Texas Health Resources https://www.texashealth.org/Documents/System/Quality_Patient_Safety/Reports/03-02-2016_Surgery.pdf U.S. News and World Report http://www.usnews.com/pubfiles/BH2015-16MethodologyReport.pdf Utah Department of Health, MONAHRQ website https://health.utah.gov/myhealthcare/monahrq/ Virginia Health Information http://www.vhi.org/MONAHRQ/default.asp?yr=2013 Washington State, MONAHRQ website http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/AboutQualityRatings WHA Information Center (Wisconsin Hospital Association) http://www.whainfocenter.com/uploads/PDFs/Publications/QualityIndicators/2012_WI_IQIReport.pdf CMS Hospital Quality Initiative: Outcome Measures http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/OutcomeMeasures.html Quality Improvement with Benchmarking (external benchmarking to multiple organizations) CMS Hospital Compare http://www.medicare.gov/hospitalcompare/Data/Measures-Displayed.html University HealthSystem Consortium/Vizient https://www.vizientinc.com/clinical-analytics-and-benchmarking.htm Quality Improvement (Internal to the specific organization) BayCare https://baycare.org/quality-report-card/surgical-complications-0715 Blue Cross Blue Shield of North Carolina http://www.bcbsnc.com/content/providers/hqp/index.htm Greenville Health System, Quality and Safety Report http://www.ghs.org/upload/docs/Reports/2013-April-Quality-Report.pdf Northwestern Memorial Hospital, Patient Safety Indicator Monitoring Plan https://www.nm.org/location/northwestern-memorial-hospital/quality-nmh/view-our-quality-ratings-nmh/surgery-nmh/general-surgery-nmh Upstate University Hospital http://qoc.upstate.edu/QualityOfCare.cfm?quality_measure_group_id=7

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Public Reporting:

Arizona Department of Health Services, AZ Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Arizona
http://pub.azdhs.gov/hospital-discharge-stats/2012/AboutQualityRatings.html

CareChex (Division of Quantros)
Provides comprehensive reports of hospitals to consumers, providers and purchasers
http://www.carechex.com/QualityIndicators.aspx

CMS Medicare Hospital Compare Program
Publically available database containing information about the quality of care at over 4,000 Medicare-certified hospitals across the U.S.
https://www.medicare.gov/hospitalcompare/Data/Measures-Displayed.html#

CMS Hospital Quality Initiative: Outcome Measures
Produces a chartbook of hospital outcome measures
http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/OutcomeMeasures.html

Commonwealth Fund, Why Not the Best
Provides performance and quality ratings for most US hospitals
http://whynotthebest.org/methodology

Connecticut Department of Health Services, CT Hospital Compare, MONAHRQ website
Hospital quality ratings from all hospitals in Connecticut
http://ctmonahrq.ct.gov/2012/index.html#/resources/AboutQualityRatings

Connecticut Hospital Association
Provide quality of care for hospitals in Connecticut
http://www.cthosp.org/advocacy/quality-and-patient-safety/hospital-quality-reporting-website/

Consumer Reports
Hospital measure performance compared to external hospitals
http://www.consumerreports.org/health/resources/pdf/how-we-rate-hospitals/How%20We%20Rate%20Hospitals.pdf

HealthGrades
Healthgrades measures 40 million patient records from 4,500 hospitals nationwide for the most recent three-year period. Consumer-targeted hospital and provider ratings
https://d2dcgio3q2u5fb.cloudfront.net/54/98/f79cdfd84640a03792ea092f20a8/2014-patient-safety-methodology.pdf

Hospital Safety Score
PSI 04 is one component of a single composite score that represents a hospital's overall performance in patient safety
http://www.hospitalsafetyscore.org/media/file/HospitalSafetyScore_ScoringMethodology_Spring2015_Final.pdf

Iowa Healthcare Collaborative
Hospital quality ratings from hospitals in Iowa
https://iowareport.ihconline.org/Public/Reports.aspx?FID=778&F1ID=0&F2ID=0&F3ID=0&CID=2&PID=4

Kentucky Cabinet for Health and Family services
Hospital quality ratings from hospitals in Kentucky
https://prd.chfs.ky.gov/MONAHRQ/2012/MONAHRQ/AboutQualityRatings.html

Kentucky Hospital Association Quality Data
Hospital quality ratings from most hospitals in Kentucky
http://info.kyha.com/qualitydata/psisite/SelectPSIReport.asp?IndID=PSI4&TimePeriod=5&GroupOpt=none&SortOrder=hospital&SortDir=ASC

Louisiana Hospital Inform

Hospital quality ratings from hospitals in Louisiana
http://lahospitalinform.org/index.html

Maryland Health Care Commission, MONAHRQ Website
Collects and provides quality ratings on hospitals across Maryland
http://www.hscrc.state.md.us/documents/md-maphs/wg-meet/di/2014-03-04/MHCC%20Inpatient%20Measures%20Inventory%20QBR%20highlights.pdf

Maine Health Data Organization (MHDO), MONAHRQ Website
Hospital quality ratings from all hospitals in Maine
https://mhdo.maine.gov/monahrq/#/resources/AboutQualityRatings

Minnesota Community Measurement
Minnesota Community Measurement is a nonprofit healthcare data reporting organization. Provides quality ratings on hospitals across Minnesota.
http://mncm.org/wp-content/uploads/2014/02/2013-HCQR-Final-2.4.2014.pdf

Nevada Compare Care, MONAHRQ website
Hospital quality ratings from most hospitals in Nevada
http://nevadacomparecare.net/MQ2014/index.html#/professional/resources/AboutQualityRatings

Niagara Health Quality Coalition, New York State Hospital Report Card
Consumer focused public report of quality indicator performance for NY hospitals.
http://www.myhealthfinder.com/newyork15/main_byproc.php

Norton Healthcare
Report patient satisfaction scores in Norton Healthcare hospitals and their performance on nationally recognized quality indicators and practices http://www.nortonhealthcare.com/QualityReport

Oklahoma State Department of Health, MONAHRQ
Compares quality ratings on hospitals across Oklahoma
https://www.phin.state.ok.us/ahrq/MONAHRQ%202010/Methodology.html

South Dakota Association of Healthcare Organizations
Use PSI 04 in a composite of serious complications in report of South Dakota hospital quality.
http://www.sdhospitalquality.org/search.php

Texas Department of State Health Services
Texas Health Care Information Collection
http://healthdata.dshs.texas.gov/Hospital/PatientSafetyQualityIndicators

Texas Health Resources
Provides quality and safety reports for all Texas Health Resources
https://www.texashealth.org/Documents/System/Quality_Patient_Safety/Reports/03-02-2016_Surgery.pdf

U.S. News and World Report
National publication that lists ratings of U.S. medical centers based on performance
http://www.usnews.com/pubfiles/BH2015-16MethodologyReport.pdf

Utah Department of Health, MONAHRQ website
Report hospital quality for all hospitals in Utah
https://health.utah.gov/myhealthcare/monahrq/

Virginia Health Information
Compares quality ratings on hospitals across Virginia
http://www.vhi.org/MONAHRQ/default.asp?yr=2013

Washington State, MONAHRQ website
Information system of inpatient care utilization, quality, and potentially avoidable stays in Washington State's community hospitals
http://www.wamonahrq.net/MONAHRQ_5p0_WA_2012/index.html#/resources/AboutQualityRatings

WHA Information Center (Wisconsin Hospital Association)
Wisconsin Inpatient Hospital Quality Indicators Report
http://www.whainfocenter.com/uploads/PDFs/Publications/QualityIndicators/2012_WI_IQIReport.pdf

Quality Improvement (external benchmarking to multiple organizations):

CMS Hospital Compare
Publically available performance measures for hospitals
http://www.medicare.gov/hospitalcompare/Data/Measures-Displayed.html

University HealthSystem Consortium/Vizient
Internal quality improvement efforts, documentation, and evaluation of AHRQ PSIs for quality improvement by its members
https://www.vizientinc.com/clinical-analytics-and-benchmarking.htm

Quality Improvement (internal to the specific organization):

BayCare
Provide information on quality of hospital care within the BayCare health system
https://baycare.org/quality-report-card/surgical-complications-0715

Blue Cross Blue Shield of North Carolina
Stimulate improvements in quality and safety within hospitals
http://www.bcbsnc.com/content/providers/hqp/index.htm

Greenville Health System, Quality and Safety Report
All data was collected from four hospitals in the Greenville Health system and compared with internal rates
http://www.ghs.org/upload/docs/Reports/2013-April-Quality-Report.pdf

Northwestern Memorial Hospital, Patient Safety Indicator Monitoring Plan
Quality improvement initiative at 894-bed academic hospital
https://www.nm.org/location/northwestern-memorial-hospital/quality-nmh/view-our-quality-ratings-nmh/surgery-nmh/general-surgery-nmh

Upstate University Hospital
Report of hospital rates against national benchmark (published online)
http://qoc.upstate.edu/QualityOfCare.cfm?quality_measure_group_id=7

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
n/a

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
n/a

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in

use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

See Table 1 in response to question 1b.2 (also included in supplemental materials)

We observe that PSI 04 rates have been relatively stable from 2011-2013 in the AHRQ QI POA Reference Population data (116-118 deaths per 1000 patients with perioperative or postoperative complications). An earlier study of administrative data showed a decrease by 6.05% per year (p<0.0001) (Downey et al., 2012).

Downey, J. R., et al. (2012). "Is patient safety improving? National trends in patient safety indicators: 1998-2007." Health Serv Res 47(1 Pt 2): 414-430.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
n/a

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No evidence has been identified suggesting unintended consequences for this measure.
Coding professionals follow detail guidelines, are subject to training and credentialing requirements, peer review and audit.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0352 : Failure to Rescue In-Hospital Mortality (risk adjusted)
0353 : Failure to Rescue 30-Day Mortality (risk adjusted)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
n/a

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**

Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
NQF 0353 uses 30-day mortality (dated from the date of the surgical admission), regardless of location, for the numerator. This is a different outcome from in-hospital mortality, and is only available in a very limited number of data sets, so NQF 0353 is a related (not competing) measure. NQF 0352 is a measure of in-hospital mortality, similar to PSI 04 (NQF 0351), but it has a different target population, so NQF 0352 is a related (not competing) measure. Specifically, the denominator for NQF 0352 and NQF 0353 is limited to surgical MS-DRGs in MDC 6 (Digestive System), MDC 7 (Hepatobiliary), MDC 9 (Skin, subcutaneous tissue, breast), MDC 10 (Endocrine, nutritional, metabolic), MDC 8 (Musculoskeletal and connective tissue), and MDC 5 (Circulatory system). By contrast, the denominator for PSI 04 (NQF 0351) also includes patients undergoing transplantation, neurosurgical, ophthalmologic, otolaryngologic (ENT), pulmonary/respiratory, urologic, gynecologic, hematologic, infection-related, trauma-related, and burn-related major procedures (if they otherwise qualify for the denominator). Therefore, the clinical/specialty breadth of the current measure is substantially greater than that of NQF 0352. Although all three of these measures are focused on "surgical patients between ages 18 and 90 admitted to an acute care hospital," the available risk-adjustment for NQF 0352 and NQF 0353 is based on Medicare fee-for-service claims data, which greatly limits the usefulness of these two measures for users with all-payer data sets (i.e., hospitals and hospital systems/associations, state and regional health data agencies, regional quality collaboratives and other "report card" sponsors, and researchers using HCUP or similar data). By contrast, the publicly available risk-adjustment for PSI 04 (NQF 0351) is based on all-payer data from 34 US states. The target population for PSI 04 (NQF 0351) is substantially broader than the target population for NQF 0352 and NQF 0353, as described above. Another key difference in denominator specifications is that PSI 04 (NQF 0351) only includes patients who experienced one or more of five broad categories of perioperative or postoperative complications, as defined by the strata. By contrast, the denominators of NQF 0352 and NQF 0353 include patients with a much wider set of 38 perioperative or postoperative complications. More importantly, in-hospital death after surgery automatically qualifies a patient for the denominator of NQF 0352, regardless whether the patient had any reported complication. As a result, the numerator of NQF 0352 includes ALL in-hospital deaths after eligible operations, whereas the numerator of PSI 04 (NQF 0351) only includes in-hospital deaths that follow one or more of the stratum-defining complications. Previous studies suggest that PSI 04 (NQF 0351) captures about 42-49% of all in-hospital deaths after qualifying operations, whereas NQF 0352 captures 100% of these deaths. The clinical rationale for this difference is that focusing on a narrower subset of deaths provides an easier target for quality improvement efforts and makes the indicator more sensitive to nursing-related quality of care (i.e., nurses are presumably less likely to be able to "rescue" patients from sudden unexpected deaths or "planned" deaths, in which physicians' orders and/or advance directives do not allow cardiopulmonary resuscitation or similar efforts). Specifically, a 2007 analysis cited in the Testing Form showed that the omega ratio summarizing the contribution of patient characteristics at the discharge-level versus hospital-level variables for explaining PSI04 (NQF 0351) was 57, compared with omega ratios of 189 for the overall risk-adjusted surgical mortality rate and 128 for NQF 0352. In other words, NQF 0352 is more heavily influenced by patient characteristics, whereas PSI 04 (NQF 0351) better isolates the hospital quality effect (albeit at the price of lower reliability, given that it only captures 42-49% of all in-hospital deaths after qualifying operations).

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
n/a

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PSI04_Supplemental_file_160531.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Agency for Healthcare Research and Quality
Co.2 **Point of Contact:** Pamela, Owens, Pam.Owens@ahrq.hhs.gov, 301-427-1412-
Co.3 **Measure Developer if different from Measure Steward:** Agency for Healthcare Research and Quality
Co.4 **Point of Contact:** Mamatha, Pancholi, Mamatha.Pancholi@ahrq.hhs.gov, 301-427-1470-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
In 2002, a workgroup convened and provided feedback on key indicator development decisions and methodology, including the usefulness of the Death Rate among Surgical Inpatients with Serious Treatable Complications (PSI04), formerly known as Failure to Rescue (PSI04). The active members of the panel were:

Michael Barrett, MD, Internist and Cardiologist
Blue Bell, PA
Medical College of Pennsylvania Hospital
Nominated by the American College of Physicians

William Golden, MD, Professor of medicine, Internist
Little Rock, AR
University of Arkansas for Medical Sciences
Nominated by the American College of Physicians

Constantine Manthous, MD, Critical care physician
Hamden CT
Yale University
Nominated by the American Thoracic Society

Brenda Snyder, RN, MS, CNS, CCRN, Critical care nurse
Evans, CO
University of Northern Colorado
Nominated by the American Association of Critical-Care Nurses

Mark W. Thomas, RPh, MS, Pharmacist, Pediatrics
Minneapolis, MN
Children´s Hospital and Clinics-Minneapolis, St. Paul
Nominated by the American Society of Health-system Pharmacists

Mark Williams, MD, Hospitalist
Atlanta, GA
Emory University of Medicine
Nominated by the National Association of Inpatient Physicians

Charles Yowler, MD, Surgeon, Critical Care - Burn Surgery
Cleveland, OH
Case Western Reserve University
Nominated by the American College of Surgeons

In 2013, ten panels of experts were convened to support the process of converting the AHRQ QIs from ICD-9-CM to ICD-10-CM/PCS in an accurate and transparent manner, to improve the validity and usefulness of the QIs. Four of these panels –focused on Cancer, Infection, Medicine, and Surgery - advised AHRQ on the ICD-10-CM/PCS specifications for PSI 04. The active members of these panels were:

Ann Borzecki, MD, MPH
Bedford, MA

Dept. of Health Policy and Management, Boston University School of Public Health, and Section of General Internal Medicine, Boston University School of Medicine, and Center for Health Quality, Outcomes and Economic Research Bedford VAMC

B. Ashleigh Guadagnolo, MD, MPH
Houston, TX
The University of Texas MD Anderson Cancer Center

Danil Victor Makarov, MD
New York, NY
Dept of Urology, New York University School of Medicine

Gail Grant, MD, MPH, MBA
Los Angeles, CA
Resource & Outcomes Management, Cedars-Sinai Health System

Joel V. Brill, MD, AGAF
Bethesda, MD
AGA Digestive Health Outcomes Registry
Fair Health, Inc., New York

John Maa, MD
San Francisco, CA
Dept of Surgery, UCSF

Kay Schwebki, MD, MA, MPH
Eden Prairie, MN
OptumInsight

Richard Dutton, MD, MBA
Park Ridge, IL
Anesthesia Quality Institute

Robert S. Gold, MD (expired 2016)
Atlanta, GA
CEO, DCBA, Inc

Coding Professionals:
Bobbi Moore, MBA, RHIT
Grand Rapids, MI
AHIMA Approved ICD-10-CM/PCS Trainer
Quality & Safety Department, Spectrum Health

Gloryanne Bryant, BS, RHIA, RHIT, CCS, CDIP, CCDS
Oakland, CA
AHIMA Approved ICD-10-CM/PCS Trainer
NCAL Revenue Cycle – HIM, Kaiser Foundation Health Plan, Inc.

Jennifer Hornung Garvin, PhD, MBA, RHIA, CPHQ, CCS, CTR
Salt Lake City, UT
FAHIMA, AHIMA Approved ICD-10-CM/PCS Trainer
IDEAS Research Center, VA Salt Lake Health Care System and University of Utah Dept of Biomedical Informatics

Lou Ann Schraffenberger, MBA, RHIA, CCS, CCS-P
Oak Brook, IL
FAHIMA, AHIMA Approved ICD-10-CM/PCS Trainer
Advocate Health Care

Mary Johnson, RHIT, CCS-P
Germantown, OH
Dept of Veteran Affairs

Monica VanSuch, MBA, RHIA
Rochester, MN
Division of Health Care Policy and Research, Mayo Clinic

Nancy Andersen, RHIA, CCS, CRCR,
Oakland, CA
AHIMA Approved ICD-10-CM/PCS Trainer
National Compliance, Ethics, and Integrity Office, Kaiser Foundation Health Plan, Inc.

Rayna Scott, MS, RHIA, CHDA
Oakbrook Terrace, IL
Division of Healthcare Quality Evaluation
The Joint Commission

Sandra Bailey, RHIA
Warrenton, VA
AHIMA Approved ICD-10-CM/PCS Trainer
Cooper Thomas

Sandra Seabold, MBA, RHIA
Cleveland, OH
Cleveland Clinic

Nurses:
Irene Lopez, BSN, RN,CSTR
Austin, TX
Trauma Services Administration, University Medical Center Brackenridge

Karen Snyder, BSN, RN
Cleveland, OH
Cleveland Clinic

Kathleen Hartman, RN, MSN
Cleveland, OH
Cleveland Clinic

Kathryn Fiandt, PhD, RN, FNP-BC, FAANP, FAAN
Galveston, TX
School of Nursing, University of Texas Medical Branch

Marybeth Farquhar, PhD, MSN, RN
Washington, DC
URAC

Patricia Hildebrand, RN, MSN, CCS-P, CPHQ, FACHE
Sugarland, TX
Hildebrand Healthcare Consulting, LLC

Other Professionals:
Anthony Warmuth, MPA, FACHE, CPHQ
Cleveland, OH
Office of Quality, Cleveland Clinic

Catherine Fulton, BS, MS, CPHQ
Montpelier, VT
Vermont Program for Quality in Health Care, Inc

James Notaro, PhD
New York, NY
Clinical Support Services, Inc.

Tina Hernandez-Boussard, PhD, MPH
Palo Alto, CA
Division of General Surgery, Stanford University School of Medicine

Wendy Patterson, MPH
Albany, NY
Office of Quality & Patient Safety, New York State Dept of Health

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2002

Ad.3 **Month and Year of most recent revision:** 06, 2016

Ad.4 **What is your frequency for review/update of this measure?** Annually

Ad.5 **When is the next scheduled review/update for this measure?** 06, 2016

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 1543
De.2. **Measure Title:** Postoperative Stroke or Death in Asymptomatic Patients undergoing Carotid Artery Stenting (CAS)
Co.1.1. **Measure Steward:** Society for Vascular Surgery
De.3. **Brief Description of Measure:** Percentage of patients 18 years of age or older without carotid territory neurologic or retinal symptoms within 120 days immediately proceeding carotid angioplasty and stent (CAS) placement who experience stroke or death during their hospitalization for this procedure. This measure is proposed for both hospitals and individual interventionalists.
1b.1. **Developer Rationale:** Better patient selection to avoid treating high risk patients who will likely experience stroke or death after CAS for asymptomatic patients which eliminates any benefit of the procedure.

S.4. **Numerator Statement:** Patients over age 18 without preoperative carotid territory neurologic or retinal symptoms within one year of their procedure who experience stroke or death during their hospitalization following elective carotid artery angioplasty and stent placement.
S.7. **Denominator Statement:** Patients over age 18 without preoperative carotid territory neurologic or retinal symptoms within one year immediately preceding carotid artery stenting.
S.10. **Denominator Exclusions:** Per PQRS Specifications for 2016: DENOMINATOR EXCLUSIONS:
Symptomatic carotid stenosis: Ipsilateral carotid territory TIA or stroke less than 120 days prior to procedure: 9006F
OR Other carotid stenosis: Ipsilateral TIA or stroke 120 days or greater prior to procedure or any prior contralateral carotid territory or vertebrobasilar TIA or stroke: 9007F

De.1. **Measure Type:** Outcome
S.23. **Data Source:** Electronic Clinical Data : Registry
S.26. **Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Facility

Original Endorsement Date: Jan 31, 2012 **Most Recent Endorsement Date:** Jan 31, 2012

IF this measure is paired/grouped, NQF#/title:
De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** Submitted SVS measure: Postoperative Stroke or Death in Asymptomatic Patients undergoing Carotid Endarterectomy

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported

by the stated rationale.

Summary of evidence:

- The developer notes better patient selection for CAS is needed for asymptomatic patients to avoid treating patients at higher risk for stroke or death following CAS.
- Guidelines from the American Heart Association recommend carotid endartectomy (CEA) for patients if the risk of death or stroke is less than 3%; the developers state this same threshold for asymptomatic patients undergoing CAS although could also be used although there is no published guideline for CAS.
- Cochrane Database analysis of stroke or death within 30 days of CAS for asymptomatic carotid stenosis showed no difference between CEA and CAS in all patients as well as for a subset of patients deemed not suitable for surgery.
- CAPTURE-2 and EXACT stent trials demonstrated outcomes for CAS in asymptomatic patients that were comparable to those established by the AHA for patients treated with CEA.
- Additional citations for evidence are linked here.

Updates

The developer submitted updated evidence for this measure

- "Safety of stenting and endarterectomy by symptomatic status in the Carotid Revascularization Endarterectomy Versus Stenting Trial (CREST)." Silver FL(1), Mackey A, Clark WM, Brooks W, Timaran CH, Chiu D, Goldstein LB, Meschia JF, Ferguson RD, Moore WS, Howard G, Brott TG; CREST Investigators. Stroke. 2011 Mar;42(3):675-80. doi: 10.1161/STROKEAHA.110.610212.
- "Randomized Trial of Stent versus Surgery for Asymptomatic Carotid Stenosis". Rosenfield K(1), Matsumura JS(1), Chaturvedi S(1), Riles T(1), Ansel GM(1), Metzger DC(1), Wechsler L(1), Jaff MR(1), Gray W(1); ACT I Investigators. N Engl J Med. 2016 Mar 17;374(11):1011-20. doi: 10.1056/NEJMoa1515706. Epub 2016 Feb 17.
- "Experience matters more than specialty for carotid stenting outcomes" Sgroi, Michael D. et al. Journal of Vascular Surgery 2015, Volume 61, Issue 4, 933 - 938.
- Experience and outcomes with carotid artery stenting: an analysis of the CHOICE study (Carotid Stenting for High Surgical-Risk Patients; Evaluating Outcomes Through the Collection of Clinical Evidence).JACC Cardiovasc Interv. 2014 Nov;7(11):1307-17. doi: 10.1016/j.jcin.2014.05.027.

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance from the Evidence Algorithm
Health outcome (Box 1)→relationship between outcome and one action is supported by rationale (Box 2) → Pass
Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- Data by the developer in the previous submission show that CAS procedures were completed for asymptomatic patients in 65% of patients in VSGNE undergoing CAS.
- An analysis of 2010-2015 VQI self-reported data (175 centers, 544 providers, 3,342 procedures) showed stroke or death within 30 days after CAS to be 2.1%.
    - The developer also notes a decrease in the percentage of cases with a reported death within 30 days of CAS has fallen from 2% to 1.6%, with the exception of in 2012 where the rate was 3.3% .
    - By center over 2010-2015, the interquartile range was 0% to 1.7% per center with the number of centers increasing each year.

Disparities

- In the 2016 analysis of 3,342 patients reported from 2010-2015 in the SVS VQI, the developer reports that the patients experiencing stroke or death within 30 days of a CAS procedure were older, had Medicare, and were slightly more likely to be female. The developer notes they did not see specific differences by race.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Is there expected variation in performance if reported at the physician level versus at the facility level?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

- An appropriate outcome strongly linked to the index intervention.

  I would like to hear from the experts about the number of patients eligible for this measure -- it is very hard to make asymptomatic patients better. Is CAS really indicated over statin therapy/observation?

  Are there data available on the risk for stroke/death during observation based on patient demographics, size or location of the plaque? It seems these should be part of the indications for the procedure, otherwise every healthy patient over 18 would be eligible!

  As with many advanced medical procedures, the best results are likely to arise in populations that don't need the intervention -- these groups have fewer co-morbidities and better outcomes with or without the procedure. How do the developers intend to address the risk of overuse?
- The evidence for the outcome applies directly. The proposed outcome of stroke/death are standard and important quality indicators of carotid intervention. The relationship between the measured outcome (in-hospital stroke/death) are supported by the stated rationale since appropriate patient selection is crucial to getting the stated benefit of the procedure.

1b. Performance Gap

- This measure documents the most critical piece of information desired by patients contemplating this procedure; the measure is therefore appropriate for public accountability reporting even if variability is low. 2) There is a moderate gap in performance demonstrated, with variability especially between low volume and high volume centers. Variability among surgeons in the registry is small, but median score is zero, so statistical power at the level of individual surgeons is likely very low.

  Variability data presented are for 30-day stroke risk, but the measure is specified to hospital discharge, which might be only 1 day. Can the developers shed light on the likely difference between 30 day and 1 day outcomes?

- The volume-outcome relationship and provider experience with CAS is well-known. It will be important to consider these features if reporting of the measure is endorsed. CAS generally carries a 2-fold greater risk of stroke/death even for asymptomatic patients when compared to CEA in a non-trial environment with variably experienced providers. The tremendous variability in patient selection and outcome for CAS even for asymptomatic disease represents an important performance gap. It is not clear that the VQI sample of CAS procedures reflects national patterns in utilization and outcomes.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- **Data source(s):** Electronic Clinical Data: Registry. Registries used for this measure are The Society for Vascular Surgery Vascular Quality Initiative (SVS VQI) and the Vascular Study Group of New England (VSGNE). VQI and VSGNE include hospitalization details and symptom status.

Specifications:

- This measure is specified at the clinician: group/practice, clinician: individual/facility level for the hospital/acute care facility setting.
- The numerator includes patients over age 18 without preoperative carotid territory neurologic or retinal symptoms within one year of their procedure who experience stroke or death during their hospitalization following elective carotid artery angioplasty and stent placement.
- The denominator includes patients over age 18 without preoperative carotid territory neurologic or retinal symptoms within one year immediately preceding carotid artery stenting.
- Patients are excluded if they are symptomatic or have other carotid stenosis less than 120 days prior to procedure.
- The measure is calculated as the number of asymptomatic patients undergoing CAS who have in hospital stroke or death divided by the number of asymptomatic patients undergoing CAS.
- The developer notes that this measure is to be reported each time a CAS is performed during the reporting period.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate data elements and definitions included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Testing has not been updated since the previous endorsement. A summary of testing is provided below.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- A random sample of 100 patient records was reviewed, representing 5 relevant procedures from 5 different hospitals from data collected during the “past 2 years”. (Note that the ‘past 2 years’ refers to the time period prior to when the measure was first endorsed).
- Hospital reporting is proposed for every 12 months based on sufficient volume.
- Annual reporting of the last 50 consecutive procedures for surgeons (which may span more than one year) with suppression of <10 procedures is recommended.
- In-hospital mortality was examined using claims based analysis of 7,205 patients discharged and recorded in the VSGNE registry between 2003 to 2007.
- Chart abstraction was completed with results compared to registry data.
- Developers analyzed the level of agreement between the chart and registry data using the Kappa statistic.

- For mortality validation, claims data from each of 12 hospitals participating in the VSGNE registry were matched to patient data within the registry to compare discharge status (alive vs dead).
- Any discrepancies were further evaluated based on medical record audit.

Results of reliability testing

- Data element validity was used to support the reliability of the measure.
- Kappa statistics indicated strong agreement for identification of the correct procedure (CAS) performed (1.0), hospital mortality (.91), hospital stroke (1.0), and asymptomatic 120 days before treatment (.90).

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
- *Is the data element level testing provided enough to also confirm reliability and validity for physician/clinician level performance?*

Guidance from the Reliability Algorithm

Precise specifications (Box 1) → Empirical reliability testing (Box 2) →Patient level data validity (Box 3) → (Box 10 of validity algorithm) →Appropriate method to assess data elements (Box 11) → Moderate certainty that the data used in the measure are valid →Highest possible rating is moderate.

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Testing has not been updated since the previous endorsement. A summary of testing is provided below.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Method(s) of validity testing

- A random sample of 100 patient records was reviewed, representing 5 relevant procedures from 5 different hospitals.
- In-hospital mortality was examined using claims based analysis of 7,205 patients discharged and recorded in the VSGNE registry between 2003 to 2007.
- Chart abstraction was completed with results compared to registry data.
- For mortality validation, claims data from each of 12 hospitals participating in the VSGNE registry were matched to patient data within the registry to compare discharge status (alive vs dead).

Results of validity testing

- Kappa statistics indicated strong agreement for identification of the correct procedure (CAS) performed (1.0), hospital mortality (.91), hospital stroke (1.0), and asymptomatic 120 days before treatment (.90).
- The developer reports that the percentage of asymptomatic patients treated in VSGNE (60%) and the postoperative stroke or death rate of 2.2% correspond to published data on asymptomatic patients.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients are excluded from the measure if they have NASCET criteria symptoms within the one year proceeding CAS; are symptomatic carotid stenosis less than 120 days prior to procedure, or other carotid stenosis 120 days or greater prior to procedure.
- Exclusions analysis was completed on 805 asymptomatic patients undergoing elective CEA from the SVS registry.
  - Death rate of 2.0% and a stroke rate of 2.11% among 287 providers in 58 centers. The interquartile range for the combined endpoint was 0.3% to 8.6%.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The measure is not risk adjusted.
- The developer gave the rationale that "risk adjustment is implicit within this quality measure" since the decision to perform this procedure "requires the interventionist to calculate the patient's risk-benefit ratio".
- The developer provided a list of carotid artery stenting definitions that includes a number of factors for pre-operative consideration.
- In addressing disparities, the developer states that such data will become available if this measure is adopted for reporting and used by more centers with more varied population demographics than found in the New England region. In analysis of disparities, the developer reports that the patients experiencing stroke or death within 30 days of a CAS procedure were older, had Medicare, and were slightly more likely to be female. The developer notes they did not see specific differences by race.

Questions for the Committee:

- Do you agree with the developer's rationale regarding risk adjustment?
- What is the Committee's expectation regarding consideration of SDS factors in maintenance measures?

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Data by the developer in the previous submission show that CAS procedures were completed for asymptomatic patients in 65% of patients in VSGNE undergoing CAS.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- The developer reports no other data sources are available.

2b7. Missing Data

- The developer notes less than 1% missing data from both VSGNE and VQI.

Guidance from the Validity Algorithm
Precise specifications (Box 1) → All threats to validity assessed (Box 2) →Insufficient

Potential threats to validity around risk adjustment and SDS factors result in preliminary rating.

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications: Reliability-Specifications

- I'd be interested in a little more information on this statement:
  Patients are excluded from the measure if they have NASCET criteria symptoms within the one year proceeding CAS; are symptomatic carotid stenosis less than 120 days prior to procedure, or other carotid stenosis 120 days or greater prior to procedure.

  Is it asymptomatic for a year before CAS, or 120 days?

- No issues with the data elements other than proposing the question as to whether or not the number of CAS procedures in the VQI truly reflects national trends in utilization and outcomes. Would favor non-risk adjusted data given relatively benign natural history of asymptomatic carotid disease with the narrow therapeutic window for carotid revascularization procedures. The reliability algorithm is clearly presented and the data element validity testing has moderate to high acceptability.

2a2. Reliability - Testing:

- Again, kappa is greater for stroke definition than for mortality. What's up with that? Both scores are high, however, lending support to the reliability of data capture, at least in the measured population.

- Sample size was modest- unclear if the VQI CAS patients reflect national trends in utilization and outcomes. Moderate reliability was the preliminary rating assigned.

2b.1 Validity – Specifications

- Highly valid measure from the public perspective and also from the clinician's point of view. This measure has very good alignment between clinical and patient-centered goals.\

- No specific threats are identified.

2b2. Validity - Testing

- Again, what is the participation in the registry of facilities and providers doing CAS?

  Assuming decent and increasing penetration, then this measure has high face validity -- easy to understand and highly relevant.

- Validity testing was adequate (100 random sample patients, 7205 VSGNE patients checked for mortality; direct chart abstraction w/ kappa statistic testing was excellent between procedure identification and mortality/stroke outcomes in the claims data and VSGNE data.

2b3-7. Threats to Validity

- The greatest concern would be centers/proceduralists who don't report this measure or participate in the appropriate registry. If this is a large number, or systematically biased (e.g. neuroradiologists) then public reporting of this measure might paint the wrong picture for the public.

- Threats to validity would be if/how risk adjustment is used. Recommend that no risk adjustment be performed for previously stated reasons and this is supported by the quality measure developers statements. Stroke/death are important quality indicators for carotid revascularization, especially for prophylactic operations. They will directly identify possible important meaningful differences that get at quality of care delivery for CAS.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data elements are generated and used by healthcare personnel during the provision of care and coded by someone other than the person obtaining the original information.
- All data elements are in defined fields with electronic clinical data
- The developer reports that in using VSGNE (which has been tracking stroke or death since 2005) and VQI, they have not had any difficulty obtaining these data.
- The developers note percent missing for this variable at less than 1%.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are there fees to belong to the registry?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

- Requires individual assessment of medical records to ensure accurate diagnosis and capture of post-procedure stroke.
  How many CAS are performed by non-surgeons? Do they participate in the registries cited? If not, how would they report this measure? How would their results likely compare?

- Stroke/death are routinely reported and reliably abstracted from EHR.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

- Centers for Medicare & Medicaid Services, Physician Quality Reporting System (PQRS) (measure #345)

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- PQRS is a reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs). PQRS measures are used for public reporting on the Physician Compare website and for the quality component of the Value-Based Payment Modifier (VBPM).

Improvement results

- Although this measure is reported in PQRS, the developer did not provide these data.
- An analysis of 2010-2015 VQI self-reported data (175 centers, 544 providers, 3,342 procedures) found an outcome of 2.1% stroke or death within 30 days after a CAS procedure.
- The developer also notes a decrease in the percentage of cases with a reported death within 30 days of CAS has fallen from 2.1% to 1.6%, with the exception of in 2012 where the rate was 3.3%.

	Free from outcome	With Outcome
Overall, N=3342	3273 (97.9%)	69 (2.1%)
2010, n=100	98 (98%)	2 (2%)
2011, n=226	221 (97.8%)	5 (2.2%)
2012, n=549	531 (96.7%)	18 (3.3%)
2013, n=739	725 (98.1%)	14 (1.9%)
2014, n=749	735 (98.1%)	14 (1.9%)
2015, n=979	963 (98.4%)	19 (1.6%)

Unexpected findings (positive or negative) during implementation

- The developer notes that data definitions regarding asymptomatic status based on NASCET criteria have eliminated confusion about symptoms.
- The developer also notes they have not had any challenges with this measure.

Potential harms

- No potential harms reported.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4. Usability and Use

- Assuming that it can capture most providers of this procedure, it seems like a good measure, with high face validity for what the publis would like to know.

- The PQRS already provides a construct for public reporting and accountability for the proposed quality measure. The author of the measure did not provide any data from PQRS on these measures.

Criterion 5: Related and Competing Measures

Related/competing measures

- N/A

Harmonization

- N/A

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1543 **NQF Project:** Surgery Endorsement Maintenance 2010

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three sub criteria must be met to pass this criterion. See guidance on evidence.

Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. **(evaluation criteria)**

1c.1 **Structure-Process-Outcome Relationship** *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process-health outcome; intermediate clinical outcome-health outcome)***:**

discussed above

1c.2-3 **Type of Evidence** *(Check all that apply)***:**

1c.4 **Directness of Evidence to the Specified Measure** *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)***:**

The combined endpoint of stroke/death is the accepted primary endpoint for both CAS and carotid endarterectomy. Variation in outcome has been established in randomized trials, cohort studies and meta analyses. This outcome measure has face validity among all providers of this service. Studies cited above have shown substantial variation in outcomes by provider when CEA is performed in asymptomatic patients. While such data does not yet exist for CAS, similar findings are expected due to the same patient population being treated.

1c.5 **Quantity of Studies in the Body of Evidence** *(Total number of studies, not articles)***:**

1c.6 **Quality of Body of Evidence** *(Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):*

1c.7 **Consistency of Results across Studies** *(Summarize the consistency of the magnitude and direction of the effect):*

1c.8 **Net Benefit** *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded?

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.11 System Used for Grading the Body of Evidence: Expert opinion.

1c.12 If other, identify and describe the grading scale with definitions:

1c.13 Grade Assigned to the Body of Evidence: Stroke/death after CAS is the reporting standard recommended by the Society for Vascular Surgery.

1c.14 Summary of Controversy/Contradictory Evidence: The endpoint of stroke, death or myocardial infarction is a frequent endpoint in CAS studies. However, this is seldom used in CEA studies, and recent studies have shown that the impact of MI is much less than the impact of stroke after CAS. Thus, we favor stroke/death as the primary endpoint for this measure.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

1.) Carotid Artery Angioplasty and Stent Placement: Quality Improvement Guidelines to Ensure Stroke Risk Reduction, J Vasc Interv Radiol 2003;14;S317-9. 2.) Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. Endarterectomy for asymptomatic carotid artery stenosis, JAMA 1995;273:1421-8. 3.) Management of Atherosclerotic Carotid Artery Disease: Clinical Practice Guidelines of the Society for Vascular Surgery, J Vasc Surg 2008;48:480-6. 4.) Clinical Competence Statement on Carotid Stenting: Training and Credentialing for Carotid Stenting-Multispecialty Consensus Recommendations, J Vasc Surg 2005;41:160-8. 5.) Percutaneous Transluminal Angioplasty and Stenting for Carotid Artery Stenosis, Cochrane Database Syst Rev 2007;(4):CD000515. 6.) Endarterectomy vs Stenting for Carotid Artery Stenosis: A Systematic Review and Meta-analysis, J Vasc Surg 2008;48:487-93. 7.) Carotid Stenting and Angioplasty, Circulation 1998;97:121-3. 8. Risk-adjusted 30-day outcomes of carotid stenting and endarterectomy: Results from the SVS Vascular Registry, J Vasc Surg 2008.

Added for 2016 Maintenance:

CAS Citations for evidence:

1. "Safety of stenting and endarterectomy by symptomatic status in the Carotid Revascularization Endarterectomy Versus Stenting Trial (CREST)." Silver FL(1), Mackey A, Clark WM, Brooks W, Timaran CH, Chiu D, Goldstein LB, Meschia JF, Ferguson RD, Moore WS, Howard G, Brott TG; CREST Investigators. Stroke. 2011 Mar;42(3):675-80. doi: 10.1161/STROKEAHA.110.610212.

2. "Randomized Trial of Stent versus Surgery for Asymptomatic Carotid Stenosis". Rosenfield K(1), Matsumura JS(1), Chaturvedi S(1), Riles T(1), Ansel GM(1), Metzger DC(1), Wechsler L(1), Jaff MR(1), Gray W(1); ACT I Investigators. N Engl J Med. 2016 Mar 17;374(11):1011-20. doi: 10.1056/NEJMoa1515706. Epub 2016 Feb 17.

3. "Experience matters more than specialty for carotid stenting outcomes" Sgroi, Michael D. et al. Journal of Vascular Surgery 2015 , Volume 61 , Issue 4 , 933 - 938.

4. Experience and outcomes with carotid artery stenting: an analysis of the CHOICE study (Carotid Stenting for High Surgical-Risk Patients; Evaluating Outcomes Through the Collection of Clinical Evidence).JACC Cardiovasc Interv. 2014 Nov;7(11):1307-17. doi: 10.1016/j.jcin.2014.05.027.

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:

Presently there is no published guideline that places a threshold for acceptable stroke and death rates following CAS for the treatment of asymptomatic carotid stenosis. There is, however, an acceptable and published threshold of 3% for patients treated with the established surgical alternative, CEA. The AHA has determined that CEA in particular should only be

performed for asymptomatic carotid stenosis if the risk of the procedure was les than 3% stroke and/or death (2). It has been suggested that this is fairly generalizable to any form of intervention (1)

1c.17 **Clinical Practice Guideline Citation:** Risk-adjusted 30-day outcomes of carotid stenting and endarterectomy: Results from the SVS Vascular Registry, J Vasc Surg 2008.

1c.18 **National Guideline Clearinghouse or other URL:** NA

1c.19 **Grading of Strength of Guideline Recommendation.** Has the recommendation been graded?

1c.20 **If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:**

1c.21 **System Used for Grading the Strength of Guideline Recommendation:** NA

1c.22 **If other, identify and describe the grading scale with definitions:**

1c.23 **Grade Assigned to the Recommendation:** NA

1c.24 **Rationale for Using this Guideline Over Others:**

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: 1c.26 Quality: 1c.27 Consistency:

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
1543_Evidence_MSF5.0_Data_-1-_2016.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Better patient selection to avoid treating high risk patients who will likely experience stroke or death after CAS for asymptomatic patients which eliminates any benefit of the procedure.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Stroke or death following CAS has been the primary clinical endpoint for a number of clinical CAS trials. Stroke or death within 30 days following intervention is captured in the SVS Registry, Vascular Quality Initiative (VQI). This endpoint is easy to capture from claims data and registries. This outcome is particularly important for asymptomatic patients undergoing CAS, since this is a prophylactic procedure being proposed to prevent future stroke. Guidelines from the American Heart Association recommend CEA for such patients only if the risk of surgical death or stroke combined is less than 3%. While there is no similar level published as a guideline, the same clinical threshold of 3% can be used for asymptomatic patients undergoing CAS. Cochrane Database analysis of stroke or death within 30 days of CAS for asymptomatic carotid stenosis showed no difference between CEA and CAS in all patients as well for a subset of patients deemed "not suitable for surgery" (CEA). Similarly, two large industry-sponsored carotid stent trials, CAPTURE-2 and EXACT, both demonstrated outcomes for CAS in asymptomatic patients that were "comparable to those established

by the AHA for patients treated with CEA".

Stroke is defined as an acute neurological deficit due to an occlusive or hemorrhagic brain lesion that persists more than 24 hours. It can be substantiated by a new stroke seen on brain imaging, but this is not a requirement, i.e., clinical symptoms alone are sufficient. Both minor and major strokes will be counted, as long as the symptoms persist more than 24 hours. Stroke in either carotid distribution, or vertebrobasilar stroke is included, i.e., any postprocedural new neurologic deficit attributed to an occlusive or hemorrhagic brain lesion lasting more than 24 hours.

While stroke or death following CAS is an appropriate quality measure for either symptomatic or asymptomatic patients, we believe that the former group would require risk adjustment to allow fair comparisons, while we do not believe this is necessary for asymptomatic patients. For asymptomatic patients, it is incumbent upon the interventionalist to select only those patients of low periprocedural risk to benefit from CAS.

We propose that the denominator for this measure should be patients who have never been symptomatic in either the cerebral hemisphere ipsilateral to the carotid lesion, the contralateral hemisphere or the vertebrobasilar circulation(dizziness or lightheadedness alone are not considered symptoms). This group has the lowest risk of stroke with carotid intervention and also the lowest risk of stroke with medical therapy alone.

Adopting this outcome measure would likely have immediate impact on improving quality. Regional data have shown that feedback of the key outcome of stroke and death, in addition to some process measures after carotid endarterectomy reduced this outcome from 5.6% to 5.0% and in asymptomatic patients from 4.1% to 3.8%. The same is likely to hold true for CAS. Reporting time frame for hospitals should be on a yearly basis. The time frame for interventionalists should be cumulative over their career.

In an analysis of the VQI self-reported data for the time period of 2010 - 2015, across 175 centers with 544 providers reporting on 3,342 procedures, we found an outcome of 2.1% of a stroke or death within 30 days after a CAS procedure for all reported cases. And, with the exception of 2012 where it jumped to 3.3%, the percentage of cases with a reported death within 30 days of the CAS procedures has fallen,2% - 1.6%, even as the number of incidents has increased as the number of patients included in the denominator has increased.

In a review by center over the five year period, we found an interquartile range of 0% to 1.7% per center with the number of centers increasing each year.

While there has been some improvement over the last five years, there continues to be a performance gap regarding the number of deaths in this 2016 study.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
To date, there is no strong evidence that CAS for asymptomatic carotid stenosis provides a significant benefit to patients over best medical therapy. Nevertheless, CAS is being performed for the treatment of asymptomatic stenosis in multiple centers in the US. The results of controlled randomized trials are pending and should soon provide the Level 1 evidence required.

Although CAS is not approved for reimbursement by CMS for asymptomatic patients, this procedure is performed for asymptomatic patients in 65% of patients in VSGNE undergoing CAS. We suspect overuse in many of these patients.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Such data will become available if this measure is adopted for reporting and used by more centers with more varied population demographics than found in the New England region.

In our 2016 analysis of the 3,342 patients reported over the time frame of 2010 - 2015 to the SVS VQI, we found that the patients still experiencing a stroke or death within 30 days of a CAS procedure were older, had Medicare as their insurance, and were slightly more likely to be female. We did not see any specific differences related to race.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from**

the literature that addresses disparities in care on the specific focus of measurement. Include citations.
not available

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, Frequently performed procedure, A leading cause of morbidity/mortality, High resource use, Severity of illness, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Percutaneous carotid intervention is a rapidly emerging field. Published trial results have established carotid stenting (CAS) in high risk surgical patients to be an effective alternative to carotid endarterectomy (CEA). It is well established that CEA benefits patients with asymptomatic >60% stenosis only if performed with a high degree of technical proficiency on appropriately selected patients. The same is proposed to hold true for CAS. This is particularly important when considering an asymptomatic population where the relative risk reduction with intervention is narrow when compared to medical management. Numerous publications have noted variation in the combined endpoint of stroke and death following carotid angioplasty and stent placement with embolic protection (5). Adoption of this outcome measure in the United States would likely disclose disperate results between hospitals and between providers, and lead to quality improvement when this information was provided to individual providers and participating centers. The SVS Vascular Registry has shown that outcome results are good for CAS, but variations exist between interventionalists and centers (8). Postoperative stroke or death is the accepted outcome parameter for this procedure, and its measurement and reporting would demonstrate variation and opportunity for improvemement. CAS is an elective procedure in nearly all cases. Patients can be referred or transferred to a center with the personnel and experience to perform this procedure with a high level of competence and any procedure that has "stroke" as a potential risk should be performed only by individuals with appropriate training and experience. (1)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1.) Carotid Artery Angioplasty and Stent Placement: Quality Improvement Guidelines to Ensure Stroke Risk Reduction, J Vasc Interv Radiol 2003;14;S317-9. 2.) Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. Endarterectomy for asymptomatic carotid artery stenosis, JAMA 1995;273:1421-8. 3.) Management of Atherosclerotic Carotid Artery Disease: Clinical Practice Guidelines of the Society for Vascular Surgery, J Vasc Surg 2008;48:480-6. 4.) Clinical Competence Statement on Carotid Stenting: Training and Credentialing for Carotid Stenting-Multispecialty Consensus Recommendations, J Vasc Surg 2005;41:160-8. 5.) Percutaneous Transluminal Angioplasty and Stenting for Carotid Artery Stenosis, Cochrane Database Syst Rev 2007;(4):CD000515. 6.) Endarterectomy vs Stenting for Carotid Artery Stenosis: A Systematic Review and Meta-analysis, J Vasc Surg 2008;48:487-93. 7.) Carotid Stenting and Angioplasty, Circulation 1998;97:121-3. 8. Risk-adjusted 30-day outcomes of carotid stenting and endarterectomy: Results from the SVS Vascular Registry, J Vasc Surg 2008.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the

Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*
Surgery : Vascular Surgery

De.6. Cross Cutting Areas *(check all the areas that apply):*
Safety : Complications

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.vascularqualityinitiative.org/wp-content/uploads/2016_PQRS_Information-v2-1.pdf

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
Attachment:

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** CAS defs v.01.09.doc

S.3. For endorsement maintenance, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
There are no changes since the last endorsement date.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients over age 18 without preoperative carotid territory neurologic or retinal symptoms within one year of their procedure who experience stroke or death during their hospitalization following elective carotid artery angioplasty and stent placement.

S.5. Time Period for Data *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Since hospitals have sufficient annual volume to generate accurate reporting levels, these are proposed for reporting every 12 months for hospital. Since surgeons have lower individual volume, we recommend annual reporting of the last 50 consecutive procedures, which may span more than one year, with suppression if < 10 procedures (ie, reported as too low volume to report).

S.6. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
ANY registry that includes hospitalization details and symptom status within 120 days is required to identify patients for numerator inclusion. The Society for Vascular Surgery Vascular Quality Initiative (SVS VQI) and the Vascular Study Group of New England (VSGNE) are examples of registries that record such information, but the measure is not limited to these registries. Other registries that collect this same information could report these measures. Patients who were asymptomatic within one year of the CAS (CPT code 37215) who died or had a stroke recorded in the registry during that admission.ANY registry that includes hospitalization details and symptom status within 120 days is required to identify patients for numerator inclusion. The Society for Vascular Surgery Vascular Quality Initiative (SVS VQI) and the Vascular Study Group of New England (VSGNE) are examples of registries that record such information, but the measure is not limited to these registries. Patients who were asymptomatic within one year of the CAS (CPT code 37215) who died or had a stroke recorded in the registry during that admission.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Patients over age 18 without preoperative carotid territory neurologic or retinal symptoms within one year immediately preceding carotid artery stenting.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
ANY registry that includes hospitalization details and symptom status within one year is required to identify patients for numerator inclusion. The Society for Vascular Surgery Vascular Quality Initiative (SVS VQI) and the Vascular Study Group of New England (VSGNE) are examples of registries that record such information, but the measure is not limited to these registries. Patients who were asymptomatic within one year of the CAS (CPT code 37215) are included.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Per PQRS Specifications for 2016:

DENOMINATOR EXCLUSIONS:
Symptomatic carotid stenosis: Ipsilateral carotid territory TIA or stroke less than 120 days prior to procedure: 9006F
OR
Other carotid stenosis: Ipsilateral TIA or stroke 120 days or greater prior to procedure or any prior contralateral carotid territory or vertebrobasilar TIA or stroke: 9007F

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Patients with NASCET criteria neurologic symptoms (transient ischemic attack, amaurosis, or stroke) within the one year immediately proceeding CAS.

DENOMINATOR EXCLUSIONS per PQRS 2016 specifications:
Symptomatic carotid stenosis: Ipsilateral carotid territory TIA or stroke less than 120 days prior to procedure: 9006F
OR
Other carotid stenosis: Ipsilateral TIA or stroke 120 days or greater prior to procedure or any prior contralateral carotid territory or vertebrobasilar TIA or stroke: 9007F

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not required

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
See "Scientific Acceptablility" section for rationale

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
Number of asymptomatic patients undergoing CAS who have in hospital stroke or death / Number of asymptomatic patients undergoing CAS

INSTRUCTIONS:
This measure is to be reported each time a CAS is performed during the reporting period. It is anticipated that clinicians who provide services of CAS, as described in the measure, based on the services provided and the measure-specific denominator coding will report this measure. This measure may be reported by clinicians who perform the quality actions described in the measure based on the services provided and the measure-specific denominator coding.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Society for Vascular Surgery Vascular Quality Initiative Registry
Vascular Study Group of New England Registry

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual, Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting*

rules, or calculation of individual performance measures if not individually endorsed.)

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
1543_MeasureTesting_MSF5.0_Data_v1.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 1543 **NQF Project:** Surgery Endorsement Maintenance 2010

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. **(evaluation criteria)**
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
A random sample of 100 patient records representing 5 procedures relevant to the measure from 5 different hospitals based on data collected during the past 2 years. In addition, in-hospital mortality was examined by claims based analysis of 7,205 patients discharged and recorded in the VSGNE registry between 2003 to 2007. These measures were originally tested in 2011 and this was the most recent data. All of the testing was approved by the NQF Steering Committee at the time that the measures were first approved in 2012. These measures are approved for PQRS reporting and working well. Regarding the sample and the data, this is an accepted testing practice to pull a sample for chart review to then compare to the data that was submitted to a registry.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
A nurse abstractor completed a form based on medical record review for the variables relevant to this measure. The results of this chart review were then compared with the original registry data. The Kappa statistic was used to judge reliability of the data. For mortality validation, claims data from each of 12 hospitals were matched to patient identified data within the VSGNE registry to compare discharge status (alive vs. dead). Any discrepencies were then further evaluated based on a medical record audit.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
The key variables for this measure and testing results were:

1. Correct procedure (carotid artery stenting) performed. Kappa =1.0
2. Hospital mortality: Kappa = .91 (SE .01)
3. Hospital stroke: Kappa = 1.0
4. Asymptomatic 120 days pre-Rx: Kappa = .90 (SE .07)

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* **are consistent with the evidence cited in support of the measure focus (***criterion 1c)* **and identify any differences from the evidence:**

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
see reliability

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
Multiple sources from the medical record were used as the gold standard, and rates compared with literature. Please see the evidence listed in the NQF form under importance.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
The percentage of asymptomatic patients being treated in VSGNE of 60% corresponds to published data on this cohort. The postop stroke or death rate of 2.2% also correponds to published results for asymptomatic patients.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
SVS Vascular Registry 805 asymptomatic patients undergoing elective CEA

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
measure calculation

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
Death rate 2.0%, stroke rate 2.11% among 287 provider in 58 centers
Interquartile range was 0.3-8.6% for the combined endpoint

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
See "Scientific Acceptablility" section for rationale. Risk adjustment is implicit within this quality measure as judged by the sponsor, the Society for Vascular Surgery, for the following reason. CAS in an asymptomatic patients is a prophylactic procedure designed to prevent future stroke. The decision to perform such a procedure requires the interventionist to calculate the patient´s risk-benefit ratio, in order to avoid post-CAS stroke or death that eliminate the benefit of the procedure. Risk adjustment based on patient factors should not be applied, since high risk patients should not undergo this prophylactic procedure, and using risk adjustment would reward interventionists who selected high risk patients for treatment.

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
N/A

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
N/A

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:
N/A

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a*

sample, characteristics of the entities included):
see section 1.b.3 and above 2,d,5

2b5.2 **Analytic Method** *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
Standard statistial analysis to determine 95% confidence interval for hospitals and providers to determine practical difference from mean

2b5.3 **Results** *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 **Data/Sample** *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
no other data sources available

2b6.2 **Analytic Method** *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:

2b6.3 **Testing Results** *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 **If measure is stratified for disparities, provide stratified results** *(Scores by stratified categories/cohorts)*: N/A

2c.2 **If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:**
No disparities have been reported. Please see the new data under the importance sections of the NQF regular form that was required as part of the measure maintenance check list.

2.1-2.3 **Supplemental Testing Methodology Information:**

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition,
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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 and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

In the VSGNE experience which has been tracking stroke or death as a major endpoint since 2005, we have not experienced any difficulty with obtaining data related to this endpoint. Our percent missing for this variable has been less than 1%. This has also been the case with the VQI.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Payment Program PQRS Approved Measure www.cms.hhs.gov

	Quality Improvement (Internal to the specific organization) Vascular Quality Initiative www.vascular.org

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

This measure is an approved measure for PQRS reporting. It is PQRS measure number 345. PQRS is the physician quality and payment program operated by the Centers for Medicare and Medicaid Services. It is a national program. We are not aware from CMS how may entities are reporting this measure nor how often it has been reported.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Overall (N=3342) Free from outcome (N=3273, 97.9%) With outcome (N=69, 2.1%)

RATE BY YEAR

2010	100 (3%)	98 (98%)	2 (2%)
2011	226 (6.8%)	221 (97.8%)	5 (2.2%)
2012	549 (16.4%)	531 (96.7%)	18 (3.3%)
2013	739 (22.1%)	725 (98.1%)	14 (1.9%)
2014	749 (22.4%)	735 (98.1%)	14 (1.9%)
2015	979 (29.3%)	963 (98.4%)	16 (1.6%)

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the**

negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.
Data definitions regarding asymptomatic status based on NASCET criteria have eliminated confusion about symtoms. Death is an accurate endpoint. Stroke has been accurately collected as judged by chart audits and comparison to claims data that has been done within VSGNE.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Society for Vascular Surgery
Co.2 **Point of Contact:** Sarah, Murphy, smurphy@vascularsociety.org, 312-334-2305-
Co.3 **Measure Developer if different from Measure Steward:** Society for Vascular Surgery
Co.4 **Point of Contact:** Jill, Rathbun, Jill_Rathbun@galileogrp.com, 703-217-7224-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
N/A

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2010
Ad.3 **Month and Year of most recent revision:** 05, 2016
Ad.4 **What is your frequency for review/update of this measure?** Annual
Ad.5 **When is the next scheduled review/update for this measure?** 12, 2017

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

CAROTID ARTERY STENT DEFINITIONS (Include only carotid bifurcation or internal carotid artery stents) v.01.09

If more than one response applies, select the most severe (highest number) response for each data field.

Pre-op

Smoking: Prior = quit ≥ 1 year ago. Current = still smoking within last 12 months. Include cigarettes, pipe, or cigar.
HTN (Hypertension): Defined as ≥ 140/90, either systolic or diastolic, at admission or within last 6 months, or clearly documented in medical record.
Beta-blockers: Peri-operative = started w/in one month before surgery or during surgery. Chronic = >than one month before surgery.
Symptoms (Coronary artery disease**):** Stable angina = stable pattern or symptoms with or without antianginal medication. Unstable angina = new onset, increasing frequency, lasting > 20 min and/or rest angina.
CABG/PTCA: Coronary artery bypass, angioplasty, or stent.
CHF (Congestive Heart Failure): Documented CHF: Mild = SOB on exertion; Severe = SOB at rest, pulmonary edema, or pitting ankle edema. (Use 2 = mild if severity not documented.)
COPD: Not treated = COPD documented in record but not treated with medication. Meds include theophylline, aminophylline, inhalers or steroids
Dialysis: Transplant = patient has functioning kidney transplant; Dialysis = currently on hemo- or peritoneal dialysis.
Creatinine: Last available measurement taken before procedure. If multiple measurements, use highest within 30 days of surgery.
Stress Test: Includes stress EKG, stress echo, nuclear stress scans, within 2 years of surgery.
Pre-admin living: Use last living status before any current, acute hospitalization or rehab unit.
Previous Arterial:
Bypass - Any non-cardiac arterial bypass for occlusive disease
CEA - Carotid endarterectomy
Aneurysm Repair – Any known true arterial aneurysm repair (excluding cerebral or pseudo-aneurysm)
PTA/Stent – Of any non-cardiac artery
Major Amputation – Any amputation above the foot or hand
Pre-Op Medications: Taken within 36 hours of surgery. Statins include any HMG-CoA reductase inhibitor, such as Lipitor, Mevacor, Pravachol, Zocor, Lescol, etc. If Plavix is disontinued prior to surgery it should be coded = 0.
Pre-op Hemoglobin: Most recent pre-op hemoglobin within past 30 days.
Symptoms: Ocular: unilateral visual loss or major blurring, etc. Cortical: unilateral motor and/or memory loss, or dysphagia/aphasia, etc. Vertebrobasiliar: bilateral motor, sensory, or visual loss, diplopia, ataxaia, etc. Major cortical or vertebrobasilar stroke = disability causing non-independent living status. Minor stroke is non-disabling. Major ocular stroke = blindness, otherwise minor. Stroke<1 month means stroke within previous month before surgery, etc. TIA=transient ischemic attack completely resolved within 24 hours.
Non-specific: Not clearly a carotid or vertebrobasilar TIA, e.g., light-headedness, dizziness
Ipsilat stroke on CT/MRI: Carotid territory only.
Medical high risk: At least one factor required: > 80 years old, severe O2 dependent pulmonary disease, CHF w/in one month, or abnormal stress test.
Anatomic high risk: Previous endarterectomy, previous neck surgery or radiation, tracheal or pharyngeal stoma, lesion above C3, contralat laryngeal nerve palsy, or contralateral carotid occlusion.
Refused for surgery: Surgeon has evaluated patient and refuses to operate due to excessive risk.

ICA stenosis: Use most severe category by modality thought to be most accurate if multiple modalities used.

Procedure

Urgency: Urgent = surgery within 24 hrs of admit or patient can't be discharged; emergent = surgery within 6 hrs of admission.
Lesion length: Length of stenosis intended to be covered with stent.
Prophylactic Anti-bradyarrhythmic: Atropine or Glycopyrolate given prior to angioplasty
Pre-dilate before protection device: Angioplasty required in order to cross lesion with a protection device.
Proximal CCA stent: Stent placement in the origin of the CCA.
Bradyarrhythmia requiring tx: Any dose given post post-dilation.
Technical failure: Can't complete procedure – CAS procedure defined as starting with attempting to place long sheath into CCA.
Protection device failure: Can't cross lesion, filter clogged, difficulty removing filter, ICA spasm requiring treatment, neurological change during procedure.

Post-op

Cranial nerve injury: Any occurrence, transient or persisting: VII-facial droop or more severe; IX-swallowing difficulty unless other diagnosis confirmed; X- hoarseness unless laryngoscopy normal; XII-any tongue deviation or dis-coordination
Ipsilat/Contralat neurologic event: Cerebral or ocular. TIA = cortical or ocular symptoms <24hrs duration. Major cortical or vertebrobasilar stroke = disability causing non-independent living status. Otherwise, minor. Major ocular stroke = blindness, otherwise minor. Minor stroke is non-disabling.
Time of Onset Ipsila/Contralat: Time when first noticed, but if noted on awakening from anesthesia code as 1=intra-op. Use 2=≤6 hrs post-op if normal at completion of procedure, and then neurologic event developed.
2b3a Inhibitor: Integrilin, Aggrastat.
Reperfusion Symptoms: Seizures associated with headache, or hemorrhage on CT/MRI.
IV meds required: Indicates continuous infusion or more than one dose required more than one hour after surgery.
Myocardial Infarction: Troponin: by local standards for MI. EKG: new Q waves, new ST and T wave changes. Clinical: documentation of MI by clinical criteria or ECHO or other imaging modality.
Dysrhythmia: New rhythm disturbance requiring treatment with medications or cardio-version.
CHF: Pulmonary edema with requirement for monitoring or treatment in ICU.
Access site cx: Complications at puncture site. PA=pseudo-aneurysm.

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3020
De.2. **Measure Title:** PBM-04: Initial Transfusion Threshold
Co.1.1. **Measure Steward:** The Joint Commission
De.3. **Brief Description of Measure:** This measure assesses the proportion of various pre-transfusion hemoglobin levels in patients age 18 and over receiving the first unit of a whole blood or packed cell transfusion. Over time, in a patient blood management program, there should be a higher proportion of patients receiving blood at the lower hemoglobin threshold and a lower proportion receiving blood at the higher hemoglobin thresholds. It also identifies patients who receive transfusions that should be reviewed by hospital transfusion/blood usage committees so that appropriate educational programs can be developed as part of a patient blood management program.
1b.1. **Developer Rationale:** All published sources indicate that a strict transfusion strategy is preferable to a liberal strategy, since transfusion can be harmful and contributes to higher mortality, infection, and other complications.1,2,3,4 Most guidelines recommend a threshold of 7.0 or 8.0 grams of hemoglobin or less as an indication for transfusion, and if the hemoglobin level is 10.0 or greater there is agreement that the transfusion is rarely indicated. There is disagreement among guidelines, however, when patients have underlying cardiac disease, postoperative status, or other clinical conditions5, others, and there is some concern about a lack of robust evidence to support some guidelines.6,7 Caution is always advised that when determining the appropriateness of transfusion, underlying clinical conditions and symptoms should be taken into consideration.
The purpose of this measure is to allow facilities to profile blood usage according to initial transfusion hemoglobin thresholds. Strata are defined to direct facility review to the appropriateness of selected transfusions, taking into account clinical symptoms combined with hemoglobin measurements. By this review, facilities will be able to determine the best approaches to enhance blood conservation and management and over time, there should be a gradual decline in the proportion of initial units given from the higher hemoglobin values to those lower values supported in a restrictive transfusion strategy in the literature and guidelines as part of a blood management program.

1. Carson JL, Grossman BJ, Kleinman S, Tinmouth at, et al. Red Blood Cell Transfusion: A Clinical Practice Guideline From the AABB. Ann Intern Med. 2012;157(1):49-58.
2. Goodnough LT, Shander A. Patient Blood Management. Anesthesiology v116; No 6, June 2012
3. Paone G, Likosky DS, Brewer R, Theurer PF, et al. Transfusion of 1 and 2 Units of Red Blood Cells is Associated With Increased Morbidity and Mortality. Ann Thorac Surg 2014; 97:87-94.
4. Shander A, Goodnough LT. Can Blood Transfusion Be Not Only Ineffective, But Also Injurious? Ann Thorac Surg 2014; 97: 11-4.
5. Shander A, Gross I, Hill S, Javidroozi M, Sledge S. A new perspective on best transfusion practices. Blood Transfus 2013; 11: 193-202.
6. Carson JL, Carless PA, Hebert PC. Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion (review). The Cochrane Collaboration, April 2012.
7. Wilkinson KL, Brunskill SJ, Doree C, Hopewell S, et al.The Clinical Effects of Red Blood Cell Transfusions: An Overview of the Randomized Controlled Trials evidence Base. Transfusion Medicine Reviews, Vol 25, No.2 (April), 2011, pp 145-155 e2.

S.4. **Numerator Statement:** Patients whose hemoglobin level measured prior to the transfusion and closest to the transfusion was:
- less than 7.0 grams
- >=7.0 and <8.0 grams
- >=8.0 and <9.0 grams
- >=9.0 and <10.0 grams
- 10.0 grams or greater

S.7. **Denominator Statement:** Patients age 18 and over receiving the first unit of a whole blood or packed cell transfusion
S.10. **Denominator Exclusions:** • Patients who have a surgical procedure performed to address a traumatic injury
• Patients who have a solid organ transplant
• Patients undergoing extracorporeal membrane oxygenation (ECMO) treatment at the time of initial transfusion.
• Patients whose first unit of whole blood or packed red blood cells was given while an Emergency Department patient.
• Patients with sickle cell disease or hereditary hemoglobinopathy

De.1. **Measure Type:** Process
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Laboratory
S.26. **Level of Analysis:** Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** This measure is not paired or grouped.

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

- The developer provides the following path to support assessment of the proportion of various pre-transfusion hemoglobin levels in patients age 18 and over receiving the first unit of a whole blood or packed cell transfusion and outcomes:
  1. Process: Assessment of pre-transfusion hemoglobin level, and administration of transfusion if hemoglobin below 7 or 8, or if clinical presentation requires transfusion at higher hemoglobin levels (restrictive strategy).
  2. Outcomes: A. Avoidance of transfusion and attendant complications. B. Reduced length of stay, reduced morbidity and mortality. C. Blood resource conservation.
- The rationale for the measure is supported by clinical guideline recommendations:
  1. AABB:
     - *Recommendation 1: The AABB recommends adhering to a restrictive transfusion strategy (7 to 8 g/dL) in hospitalized, stable patients. (*Grade: strong recommendation; high-quality evidence.)
     - *Recommendation 2: The AABB suggests adhering to a restrictive strategy in hospitalized patients with preexisting cardiovascular disease and considering transfusion for patients with symptoms or a hemoglobin level of 8 g/dL or less. (*Grade: weak recommendation; moderate-quality evidence.)
     - *Recommendation 3: The AABB cannot recommend for or against a liberal or restrictive transfusion threshold for hospitalized, hemodynamically stable patients with the acute coronary syndrome.* (Grade: uncertain recommendation; very low-quality evidence.)
  2. Society of Thoracic Surgeons/ The Society of Cardiovascular Anesthesiologists: *With hemoglobin levels below*

6 g/dL, red blood cell transfusion is reasonable since this can be lifesaving. Transfusion is reasonable in most postoperative patients whose hemoglobin is less than 7 g/dL but no high level evidence supports this recommendation. (Level of evidence C. CLASS IIa - Additional studies with focused objectives needed. IT IS REASONABLE to perform procedure/administer treatment.)

3. Society of Critical Care Medicine: *"A "restrictive" strategy of RBC transfusion (transfusion when Hb <7 g/dL) is as effective as a "liberal" strategy (transfusion when Hb < 10 g/dL) in critically ill patients with hemodynamically stable anemia, except possibly in patients with acute myocardial ischemia". (*Grade-Level 1. The recommendation is convincingly justifiable based on the available scientific information alone.)

- Findings from 2 systematic reviews included a Cochrane Review that addressed evidence for the effect of transfusion thresholds on the use of allogeneic and/or autologous red cell transfusion, and the evidence for any effect on clinical outcomes. The developer also presented a Salpeter Meta-Analysis and Systematic Review that looked at randomized controlled trials evaluating a restrictive transfusion trigger of <7 g/dL.

Guidance from the Evidence Algorithm

Process measure based on SR/grading (Box 3) → QQC provided (Box 4) → Moderate quality evidence (Box 5b) → MODERATE

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

Although there is no performance data on the measure as specified, the developer provided data on blood transfusion appropriateness and rate of hospitalization with blood transfusion that indicates opportunity for improvement.

Disparities

The developed indicated that no disparity data are available.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are data available to show the percent of transfusions when hemoglobin levels are at the various lower levels specified?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

- I think this is too hard a concept to measure. Has the strong unintended consequence of delaying transfusion in patients who need it. Also does not account for the vector (how fast is bleeding?) and the need to make decisions in the fog of war.
  Evidence is more controversial than stated.
  Experts moving away from absolute threshold and towards "is patient bleeding?" and "how are they doing?"

- This is a Trial Approval Program process and outcome eMeasure for restrictive transfusion strategy.

  There are 2 Cochrane systematic reviews specific to this measure; quality, quantity and consistency of the evidence has been provided; the evidence is graded; and is supported by clinical guideline recommendations from the AABB, STS/Society of CV Anesthesiologists and Society of Critical Care Medicine.

1b. Performance Gap

- No performance or disparity data was provided.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): EHR

Specifications: HQMF specifications are provided – see technical review

- Numerator Statement: Patients whose hemoglobin level measured prior to the transfusion and closest to the transfusion was:
  - less than 7.0 grams
  - >=7.0 and <8.0 grams
  - >=8.0 and <9.0 grams
  - >=9.0 and <10.0 grams
  - 10.0 grams or greater
- Denominator Statement: Patients age 18 and over receiving the first unit of a whole blood or packed cell transfusion
- Denominator Exclusions:
  - Patients who have a surgical procedure performed to address a traumatic injury
  - Patients who have a solid organ transplant
  - Patients undergoing extracorporeal membrane oxygenation (ECMO) treatment at the time of initial transfusion.
  - Patients whose first unit of whole blood or packed red blood cells was given while an Emergency Department patient.
  - Patients with sickle cell disease or hereditary hemoglobinopathy

eMeasure Technical Advisor(s) review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously;

	Submitted with Bonnie results
Feasibility Testing	The feasibility analysis submitted by the measure developer meets the requirements to be considered for eMeasure Trial Approval.

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

Initial reliability testing was conducted in the Bonnie test deck; the overall patient simulation included 48 patients. The developer stated that Bonnie testing confirms that the measure logic performs as expected and that the terminologies used are applied consistently. As a measure under consideration for the Trial Approval program, the developers must indicate if they have a plan in place for full testing (reliability and validity) and this information will be submitted and evaluated by NQF prior to any consideration of full measure endorsement. The Testing attachment indicates a plan for reliability and validity testing.

Questions for the Committee:

- *The Committee will not be asked to vote on Reliability for this eMeasure since it is being considered for Trial Use; however, questions regarding the testing plan and other concerns about reliability are welcome for discussion.*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Based on the information provided, and intent of the measure, do you feel the specifications are consistent with evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

The only testing completed to date includes Bonnie testing and some review for feasibility. Additionally, the developer stated that findings from public comment support the face validity of this measure. The public comment was open for 30 days and the Joint Commission received 150 responses. Respondents were asked to rate the measure on a number of parameters, using a Likert scale ranging from 1 to 5, where 1=Disagree and 5=Agree. The table below presents the average rating for these parameters.

PARAMETER	RATING
Numerator clearly describes the activity being measured	4.41
Denominator clearly describes the activity being measured	4.40
Numerator inclusions clear and appropriate	4.46
Denominator inclusions clear and appropriate	4.42
Numerator exclusions clear and appropriate	4.44

Denominator exclusions clear and appropriate	4.36
Accurately assesses the process of care to which it is addressed	4.31

This measure is being considered for trial use, thus full validity testing results are not expected and the Committee will not vote on this criterion.

2b3-2b7. Threats to Validity

2b3. Exclusions:
When data are available, The Joint Commission will analyze exclusion frequency and variability across providers. These data elements to be analyzed include:

- Patients with a traumatic injury <=48 hours prior to or during the encounter.
- Patients with a solid organ transplant <=48 hours prior to or during the encounter.
- Patients who have an ECMO procedure during the inpatient encounter.
- Patients with sickle cell disease and related blood disorders

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

Unknown at this time.

2b6. Comparability of data sources/methods:
N/A

2b7. Missing Data

The developer stated that eMeasures are calculated using only the structured data collected in certified EHR technology. Data not present in the structured field from which the measure draws will not be included in the measure calculation.

The Committee will only vote on one portion of Scientific Acceptability: 2b1 – to determine if the measure specifications are consistent with evidence. This is a must pass criteria.

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

- The specifications are consistent with the evidence, though full testing of the eMeasure, as specified has not been conducted to assess reliability or validity.

2b2. Validity – Testing

- In addition to Bonnie testing, TJC held a 30 day public comment period with 150 responses that support the face validity of the measure, per the developer.

  Full reliability and validity testing must be completed prior to NQF consideration of full measure endorsement.

2b3-7. Threats to Validity

- Data not present in an EHR structured field for the measure will not be included in the measure calculation.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The feasibility analysis submitted by the measure developer meets the requirements to be considered for eMeasure Trial Approval. Based on the findings of the eMeasure Technical Review, the submitted eMeasure specification is capable of being processed and interpreted by clinical information systems and is ready for implementation in real world settings.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

- Following feasibility analysis the measure meets the requirements to be considered for eMeasure Trial Approval.

  Data source is the EHR. Not all facilities have an EHR. The denominator exclusions may not be complete. Abstraction of the chart may require significant modification of the data fields in the EHR or addition of staff to abstract the paper record.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details The Joint Commission maintains a certification program in Blood Management, which is a voluntary program for hospitals to achieve excellence in patient blood management. The measures in this set can be made available within a year for hospitals to use in fulfilling the requirements for certification.

Improvement results N/A

Unexpected findings (positive or negative) during implementation N/A

Potential harms None identified

Feedback :

None identified

***Questions for the Committee**:*

- *Does the Committee consider the certification program in Blood Management to be an accountability program?*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4. Usability and Use

- This measure is not publically reported. Planned to be used in an accountability program. TJC maintains a voluntary Blood Management certification program.

Criterion 5: Related and Competing Measures

Related or competing measures

N/A

Harmonization

N/A

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: PBM-04: Initial Transfusion Threshold

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 5/20/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to

demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. 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 multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Assessment of the proportion of various pre-transfusion hemoglobin levels in patients age 18 and over receiving the first unit of a whole blood or packed cell transfusion.☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

3. Process: Assessment of pre-transfusion hemoglobin level, and administration of transfusion if hemoglobin below 7 or 8, or if clinical presentation requires transfusion at higher hemoglobin levels (restrictive strategy).
4. Outcomes: A. Avoidance of transfusion and attendant complications. B. Reduced length of stay, reduced morbidity and mortality. C. Blood resource conservation.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

A. AABB

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Carson JL, Grossman BJ, Kleinman S, Tinmouth at, et al. Red Blood Cell Transfusion: A Clinical Practice Guideline From the AABB. *Ann Intern Med.* 2012;157(1):49-58.

http://annals.org/article.aspx?articleid=1206681

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Recommendation 1: The AABB recommends adhering to a restrictive transfusion strategy (7 to 8 g/dL) in hospitalized, stable patients.

Recommendation 2: The AABB suggests adhering to a restrictive strategy in hospitalized patients with preexisting cardiovascular disease and considering transfusion for patients with symptoms or a hemoglobin level of 8 g/dL or less.

Recommendation 3: The AABB cannot recommend for or against a liberal or restrictive transfusion threshold for hospitalized, hemodynamically stable patients with the acute coronary syndrome.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Recommendation 1: Grade: strong recommendation; high-quality evidence.

Recommendation 2: Grade: weak recommendation; moderate-quality evidence.

Recommendation 3: Grade: uncertain recommendation; very low-quality evidence.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

The strength of recommendations (for or against intervention) is graded as "strong" (indicating judgment that most well-informed people will make the same choice; "We recommend . . . "), "weak" (indicating judgment that a majority of well-informed people will make the same choice, but a substantial minority will not; "We suggest . . . "), or "uncertain" (indicating that the panel made no specific recommendation for or against interventions; "We cannot recommend . . . ").

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Same.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

B. Society of Thoracic Surgeons/ The Society of Cardiovascular Anesthesiologists

1a.4.1. Guideline citation (*including date*) and URL for guideline (*if available online*):

Ferraris V, Brown JR, Despotis GJ, Hammon JW, et al. 2011 Update to the Society of Thoracic Surgeons and the Society of Cardiovascular Anesthesiologists Blood Conservation Clinical Practice Guidelines.

Ann Thorac Surg 2011;91:944-82.

http://www.annalsthoracicsurgery.org/article/S0003-4975(10)02888-2/pdf

1a.4.2. Identify guideline recommendation number and/or page number and quote verbatim, the specific guideline recommendation.

(No Number Table 2): "With hemoglobin levels below 6 g/dL, red blood cell transfusion is reasonable since this can be lifesaving. Transfusion is reasonable in most postoperative patients whose hemoglobin is less than 7 g/dL but no high level evidence supports this recommendation. (Level of evidence C)."

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Class IIA **CLASS IIa,** *Benefit >> Risk*
Additional studies with focused objectives needed. **IT IS REASONABLE** to perform procedure/administer treatment.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

CLASS I, *Benefit >>> Risk*
Procedure/Treatment **SHOULD** be performed/administered

CLASS IIb, *Benefit ≥ Risk*
Additional studies with broad objectives needed; additional registry data would be helpful.
Procedure/Treatment **MAY BE CONSIDERED**

CLASS III, *Risk ≥ Benefit*
Procedure/Treatment should **NOT** be performed/administered **SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL**

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*):

Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines

http://professional.heart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → *complete section 1a.7*

☒ No → *report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7*

C. American Red Cross Guideline

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Vassallo R, et al. A Compendium of Transfusion Practice Guidelines, Second Edition, 2013. American Red Cross, page 8.
http://www.redcrossblood.org/sites/arc/files/59802_compendium_brochure_v_6_10_9_13.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Page 15: A restrictive RBC transfusion strategy (Hgb 7–8 g/dL trigger) is recommended in stable hospitalized patients.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

No grade assignment

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

n/a

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

n/a

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

D. Society of Critical Care Medicine:

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Napolitano L, et al. Clinical Practice Guideline: Red blood cell transfusion in adult trauma and critical care. Crit Care Med 2009 Vol 37, No 12.

http://journals.lww.com/ccmjournal/Abstract/2009/12000/Clinical_practice_guideline__Red_blood_cell.19.aspx

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Page 3127, recommendation 3: "A "restrictive" strategy of RBC transfusion (transfusion when Hb <7 g/dL) is as effective as a "liberal" strategy (transfusion when Hb < 10 g/dL) in critically ill patients with hemodynamically stable anemia, except possibly in patients with acute myocardial ischemia".

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Level 1. The recommendation is convincingly justifiable based on the available scientific information alone. This recommendation is usually based on Class I data, however strong Class II evidence may form the basis for a Class 1 recommendation.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Level 2. The recommendation is reasonably justifiable by available scientific evidence and strongly supported by expert opinion. This recommendation is usually supported by Class II data or a preponderance of Class III evidence.

Level 3. The recommendation is supported by available data but adequate scientific evidence is lacking.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Same, p. 3126.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

A. Cochrane Review

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Carson JL, Carless PA, Hebert PC. Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion (review). *The Cochrane Collaboration*, April 2012.

http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD002042.pub3/full

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Same as for evidence review (see full text article); no grading applied.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

C. Salpeter Meta-Analysis and Systematic Review:

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Salpeter SR, Buckley JS, Chatterjee s> Impact of More Restrictive Blood Transfusion Strategies on Clinical Outcomes: A Meta-analysis and Systematic Review. *The American Journal of Medicine*, Vol 127, No 2, February 2014.

http://www.ncbi.nlm.nih.gov/pubmed/24331453

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Same as for evidence review (see full text article); no grading applied.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

A. **Cochrane Review:**

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Evidence for the effect of transfusion thresholds on the use of allogeneic and/or autologous red cell transfusion, and the evidence for any effect on clinical outcomes.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

No grade assigned.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

n/a

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: "Unrestricted"

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

19 controlled trials were ultimately included, covering a span of 55 years and 6,264 patients.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Selection bias – Low for 9 trials, insufficient information for 9 trials, high risk for 1 trial

Allocation bias – low for 4 trials, unclear for 13 trials, high risk for 2 trials

Blinding of physicians – low risk for all trials

Incomplete outcome data – low risk for 14 trials, unclear for 5 trials

Selective reporting – none

Eighteen of nineteen trials presented data suitable for inclusion in the meta-analysis.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Results were that "In patients who do not have acute coronary artery disease, blood transfusion can probably be withheld in the presence of haemoglobin levels as low as 7.0 g/dL to 8.0 g/dL as long as there is no notable bleeding." In addition, "…restrictive transfusion strategies were associated with a reduction of more than one-third in the number of patients receiving blood, a red cell transfusion requirement that was approximately one unit lower, and a haemoglobin concentration (average postoperative) that was around 1.5 g/dL lower than in the blood transfusion group."

Ratings of statistical significance and confidence intervals varied by studied outcome; heterogeneity between studies was statistically significant (P<0.00001, Chi 96.82).

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

"None of the outcomes evaluated, including mortality, cardiac morbidity, infections, and length of hospital stay, appear to be adversely affected by the lower use of red cell transfusions."

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

n/a

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

B. **Clinical Practice Guideline, AABB:**

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

The GRADE system (39) uses the following 4 ratings for quality of evidence:

"High" indicates considerable confidence in the estimate of effect. The true effect probably lies close to the estimated effect, and future research is unlikely to change the estimate of the health intervention's effect.

"Moderate" indicates confidence that the estimate is close to the truth. Further research is likely to have an important effect on confidence in the estimate and may change the estimate of the

health intervention's effect.

"Low" indicates that confidence in the effect is limited. The true effect may differ substantially from the estimate, and further research is likely to have an important effect on confidence in the estimate of the effect and is likely to change the estimate.

"Very low" indicates little confidence in the effect estimate. Any estimate of effect is very uncertain.

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

C. **Salpeter Meta-Analysis and Systematic Review:**

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Randomized controlled trials evaluating a restrictive transfusion trigger of <7 g/dL.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

No grade assigned.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

n/a

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).

Date range: 1966 to April 2013.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Primary analysis – 3 RCTs (trigger <7 g/dL) Secondary analysis – 19 RCTS (trigger 7.5 – 10 g.dL).

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Unstated

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

In the primary analysis, pooled results from 3 trials with 2364 participants showed that a restrictive hemoglobin transfusion trigger of <7 g/dL resulted in reduced in-hospital mortality (risk ratio [RR], 0.74; confidence interval [CI], 0.60-0.92), total mortality (RR, 0.80; CI, 0.65-0.98), rebleeding (RR, 0.64; CI, 0.45-0.90), acute coronary syndrome (RR, 0.44; CI, 0.22-0.89), pulmonary edema (RR, 0.48; CI, 0.33-0.72), and bacterial infections (RR, 0.86; CI, 0.73-1.00), compared with a more liberal strategy. Pooled data from randomized trials with less restrictive transfusion strategies showed no significant effect on outcomes.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

In-hospital mortality, total mortality, ACS, pulmonary edema, and infection occurrence were all assessed, with reduction in occurrence of all noted in a more restrictive strategy.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

n/a

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

In January 2015 a literature search of EMBASE, Pub Med, MEDLINE and other relevant sources including professional association websites, The Cochrane Library, the National Guideline Clearinghouse, and other sources was conducted, using search terms such as anemia, preoperative testing, and other relevant search terms, requesting English language publications from 2009 – 2014. Identified publications were searched for additional relevant reference documents.

1a.8.2. Provide the citation and summary for each piece of evidence.

New York State Department of Health:

"Hemoglobin concentration:

Hb >10 g/dL– Transfusion is rarely indicated

Hb 6-10 g/dL – indications for transfusion should be based on the patient's risk of inadequate oxygenation from ongoing bleeding and/or high-risk factors, such as age, cardiovascular compromise, or respiratory disease.

Hb <6 g/dL – transfusion is almost always indicated"

New York State Council on Human Blood and Transfusion Services. Guidelines for Transfusion Options and Alternatives, 2010. Downloaded from www.wadswoth.org/labcert/blood_tissue July 2015.

American Academy of Family Physicians: "The threshold for transfusion of red blood cells should be a hemoglobin level of 7g/dL (70 g/L) in most adults and children." Evidence Rating A, RCTs in adults and children with critical illness. Transfusion of Blood and Blood Products: Indications and Complications. Am Fam Physician 2011;83(6): 719-24.

Sharma S, Sharma P, Tyler L. Transfusion of Blood and Blood Products: Indications and Complications. *American Family Physician*. March 15, 2011: volume 83 No. 6, p.720.
"The threshold for transfusion of red blood cells should be a hemoglobin level of 7 g per dL (70 g per L) in adults and most children." (Evidence rating A).

Shander et al. Appropriateness of Allogeneic Red Blood Cell Transfusion: The International Consensus Conference on Transfusion Outcomes. *Transfusion Medicine Reviews*, Vol 25, No 3 (July), 2011: pp 232-246.e53.
An international multidisciplinary panel of 15 experts reviewed 494 published articles and used the RAND/UCLA Appropriateness Method to determine the appropriateness of allogeneic red blood cell (RBC) transfusion based on its expected impact on outcomes of stable nonbleeding patients in 450 typical inpatient medical, surgical, or trauma scenarios. Panelists rated allogeneic RBC transfusion as appropriate in 53 of the scenarios (11.8%), inappropriate in 267 (59.3%), and uncertain in 130 (28.9%). Red blood cell transfusion was most often rated appropriate (81%) in scenarios featuring patients with hemoglobin (Hb) level 7.9 g/dL or less, associated comorbidities, and age older than 65 years. Red blood cell transfusion was rated inappropriate in all scenarios featuring patients with Hb level 10 g/dL or more and in 71.3% of scenarios featuring patients with Hb level 8 to 9.9 g/dL. Conversely, no scenario with patient's Hb level of 8 g/dL or more was rated as appropriate. Nearly one third of all scenarios were rated uncertain, indicating the need for more research. The observation that allogeneic RBC transfusions were rated as either inappropriate or uncertain in most scenarios in this study supports a more judicious transfusion strategy.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
PBM_04_evidence_attachment.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
All published sources indicate that a strict transfusion strategy is preferable to a liberal strategy, since transfusion can be harmful and contributes to higher mortality, infection, and other complications.1,2,3,4 Most guidelines recommend a threshold of 7.0 or 8.0 grams of hemoglobin or less as an indication for transfusion, and if the hemoglobin level is 10.0 or greater there is agreement that the transfusion is rarely indicated. There is disagreement among guidelines, however, when patients have underlying cardiac disease, postoperative status, or other clinical conditions5, others, and there is some concern about a lack of robust evidence to support some guidelines.6,7 Caution is always advised that when determining the appropriateness of transfusion, underlying clinical conditions and symptoms should be taken into consideration.
The purpose of this measure is to allow facilities to profile blood usage according to initial transfusion hemoglobin thresholds. Strata are defined to direct facility review to the appropriateness of selected transfusions, taking into account clinical symptoms combined with hemoglobin measurements. By this review, facilities will be able to determine the best approaches to enhance blood conservation and management and over time, there should be a gradual decline in the proportion of initial units given from the higher hemoglobin values to those lower values supported in a restrictive transfusion strategy in the literature and guidelines as part of a blood management program.

1. Carson JL, Grossman BJ, Kleinman S, Tinmouth at, et al. Red Blood Cell Transfusion: A Clinical Practice Guideline From the AABB. Ann Intern Med. 2012;157(1):49-58.
2. Goodnough LT, Shander A. Patient Blood Management. Anesthesiology v116; No 6, June 2012
3. Paone G, Likosky DS, Brewer R, Theurer PF, et al. Transfusion of 1 and 2 Units of Red Blood Cells is Associated With Increased Morbidity and Mortality. Ann Thorac Surg 2014; 97:87-94.
4. Shander A, Goodnough LT. Can Blood Transfusion Be Not Only Ineffective, But Also Injurious? Ann Thorac Surg 2014; 97: 11-4.
5. Shander A, Gross I, Hill S, Javidroozi M, Sledge S. A new perspective on best transfusion practices. Blood Transfus 2013; 11: 193-202.
6. Carson JL, Carless PA, Hebert PC. Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion (review). The Cochrane Collaboration, April 2012.
7. Wilkinson KL, Brunskill SJ, Doree C, Hopewell S, et al.The Clinical Effects of Red Blood Cell Transfusions: An Overview of the Randomized Controlled Trials evidence Base. Transfusion Medicine Reviews, Vol 25, No.2 (April), 2011, pp 145-155 e2.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This is a new measure for which approval for trial use is requested.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
A. Agency for Healthcare Research and Quality (AHRQ): Blood transfusion was the most common listed procedure performed during hospitalizations in 2010 (11 percent of stays with a procedure); the rate of hospitalization with blood transfusion has more than doubled since 1997. The percentage change in rate of all stays with a blood transfusion from 1997 – 2010 is 126%. Most Frequent Procedures Performed in U.S. Hospitals, 2010. Healthcare Cost and Utilization project, statistical brief, February 2013, AHRQ.

B. An international multidisciplinary panel of 15 experts reviewed 494 published articles and used the RAND/UCLA Appropriateness Method to determine the appropriateness of allogeneic red blood cell (RBC) transfusion based on its expected impact on outcomes of stable nonbleeding patients in 450 typical inpatient medical, surgical, or trauma scenarios. Panelists rated allogeneic RBC transfusion as appropriate in 53 of the scenarios (11.8%), inappropriate in 267 (59.3%), and uncertain in 130 (28.9%). Red blood cell transfusion was most often rated appropriate (81%) in scenarios featuring patients with hemoglobin (Hb) level 7.9 g/dL or less, associated comorbidities, and age older than 65 years. Red blood cell transfusion was rated inappropriate in all scenarios featuring patients with Hb level 10 g/dL or more and in 71.3% of scenarios featuring patients with Hb level 8 to 9.9 g/dL. Conversely, no scenario with patient´s Hb level of 8 g/dL or more was rated as appropriate. Nearly one third of all scenarios were rated uncertain, indicating the need for more research. The observation that allogeneic RBC transfusions were rated as either inappropriate or uncertain in most scenarios in this study supports a more judicious transfusion strategy. Appropriateness of Allogeneic Red Blood Cell transfusion: The International Consensus Conference on Transfusion Outcomes. Transfusion Medicine Reviews, Vol 25, No. 3 (July)

2011.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
No disparities were identified.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
No disparity data was found in the literature.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, Frequently performed procedure, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
1. Agency for Healthcare Research and Quality (AHRQ): Blood transfusion was the most common of all listed procedures performed during hospitalizations in 2010 (11 percent of stays with a procedure); the rate of hospitalization with blood transfusion has more than doubled since 1997. The percentage change in rate of all stays with a blood transfusion from 1997 – 2010 is 126%.
2. AABB: If a restrictive transfusion strategy were widely implemented and replaced a liberal strategy, exposure of patients to red blood cell (RBC) transfusions would decrease by an average of approximately 40% (relative risk, 0.61 [confidence interval (CI), 0.52 to 0.72]). This would have a large effect on blood use and the risks for infectious and noninfectious complications of transfusion. Unnecessary transfusions increase costs and expose patients to potential infectious or noninfectious risks, such as hepatitis B and C virus, human immunodeficiency virus (HIV), transfusion-associated circulatory overload, transfusion-related acute lung injury, fatal hemolysis, life-threatening reaction, and fever.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
1. Most Frequent Procedures Performed in U.S. Hospitals, 2010. Healthcare Cost and Utilization project, statistical brief, February 2013, AHRQ.
2. Red blood cell transfusion: a clinical practice guideline from the AABB. Ann Intern Med. 2012 Jul 3;157(1):49-58. [63 references]

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not a PRO-PM.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across

organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*
Surgery

De.6. Cross Cutting Areas *(check all the areas that apply):*
Overuse, Safety

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.jointcommission.org/measure_development_initiatives.aspx

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** PBM-04_InitialTransfusionThreshold.zip

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** InitialTransfusionThreshold_v4_3_Wed_Jun_08_10.20.18_CDT_2016.xls

S.3. For endorsement maintenance, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
n/a

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients whose hemoglobin level measured prior to the transfusion and closest to the transfusion was:
- less than 7.0 grams
- >=7.0 and <8.0 grams
- >=8.0 and <9.0 grams
- >=9.0 and <10.0 grams
- 10.0 grams or greater

S.5. Time Period for Data *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Hemoglobin level prior to and closest to the transfusion is represented by a code from the following Value Set and associated QDM datatype:
- "Laboratory Test, Performed: Hemoglobin blood serum plasma" using "Hemoglobin blood serum plasma LOINC Value Set (2.16.840.1.113762.1.4.1104.4)

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Patients age 18 and over receiving the first unit of a whole blood or packed cell transfusion

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

Inpatient encounters are represented by a code from the following value set and associated QDM datatype:
•"Encounter, Performed: Encounter Inpatient" using "Encounter Inpatient SNOMEDCT Value Set (2.16.840.1.113883.3.666.5.307)"

Patients who receive the first unit of a packed cell or whole blood transfusion are represented by a code from the following Value Set and associated QDM datatype:
"Procedure, Performed: Blood Transfusion Administration" using "Blood Transfusion SNOMEDCT Value Set (2.16.840.1.113762.1.4.1029.24)

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

- Patients who have a surgical procedure performed to address a traumatic injury
- Patients who have a solid organ transplant
- Patients undergoing extracorporeal membrane oxygenation (ECMO) treatment at the time of initial transfusion.
- Patients whose first unit of whole blood or packed red blood cells was given while an Emergency Department patient.
- Patients with sickle cell disease or hereditary hemoglobinopathy

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

Patients who have a surgical procedure performed to address a traumatic injury are represented by a code from the following Value Set and associated QDM datatype:
"Attribute: Diagnosis: Traumatic Injury" using "Traumatic Injury Grouping Value Set (2.16.840.1.113762.1.4.1029.10)
Patients who have a solid organ transplant are represented by a code from the following Value Set and associated QDM datatype:
"Procedure, Performed: Solid Organ Transplant" using "Solid Organ Transplant Grouping Value Set (2.16.840.1.113762.1.4.1029.11)"
Patients who undergo ECMO at the time of initial transfusion are represented by a code from the following Value Set and associated QDM datatype:
"Procedure, Performed: ECMO" using "ECMO Grouping Value Set (2.16.840.1.113762.1.4.1029.22)
Patients whose first unit is given while an Emergency Department patient are implicity excluded as blood administered in an ED location is not captured in this measure.

Patients with sickle cell disease or hereditary hemoglobinopathy are represented by a code from the following Value Set and associated QDM datatype:
Attribute: "Diagnosis: Sickle Cell Disease and Related Blood Disorders" using "Sickle Cell Disease and Related Blood Disorders Grouping Value Set (2.16.840.1.113762.1.4.1029.35)"

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*

Stratification 1 =
AND: Most Recent: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma" <= 45 day(s) starts before start of "Occurrence A of Procedure, Performed: Blood Transfusion Administration"
AND: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma (result < 7.0 g)"

Stratification 2 =
AND: Most Recent: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma" <= 45 day(s) starts before start of "Occurrence A of Procedure, Performed: Blood Transfusion Administration"
AND: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma" satisfies all:
(result >= 7.0 g)
(result < 8.0 g)

Stratification 3 =
AND: Most Recent: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma" <= 45 day(s) starts before start of "Occurrence A of Procedure, Performed: Blood Transfusion Administration"
AND: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma" satisfies all:

(result >= 8.0 g)
(result < 9.0 g)

Stratification 4 =
AND: Most Recent: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma" <= 45 day(s) starts before start of "Occurrence A of Procedure, Performed: Blood Transfusion Administration"
AND: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma" satisfies all:
(result >= 9.0 g)
(result < 10.0 g)

Stratification 5 =
AND: Most Recent: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma" <= 45 day(s) starts before start of "Occurrence A of Procedure, Performed: Blood Transfusion Administration"
AND: "Occurrence A of Laboratory Test, Performed: Hemoglobin blood serum plasma (result >= 10.0 g)"

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
n/a

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
n/a

S.16. Type of score:
Count
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Score within a defined interval

S.18. Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
See attached HQMF file.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Records are not sampled.

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on*

minimum response rate.)
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Measure is not based on a survey; not a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
eMeasures are calculated using only the structured data collected in certified EHR technology (CEHRT). Data not present in the structured field from which the measure draws will not be included in the measure calculation.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Laboratory

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Hospitals report EHR data using Certified Electronic Health Record Technology (CEHRT), and by submitting Quality Reporting Document Architecture Category 1 (QRDA-1).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not a composite measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
PBM_04_testing_form_for_trial_use.docx,PBM04_CMS608v0_Bonnie_Export.xlsx

National Quality Forum
Measure Testing Form for Trial Approval Program

Measure Title: PBM-04: Initial Transfusion Threshold

Date of Submission: 5/31/2016

Type of Measure:

☐ Composite –	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

A measure submision that is to be considered for the Trial Approval Program must complete this form in its entirety. Either a test data set provided by the measure developer, or the use of the Bonnie tool is acceptable to provide prelminary testing results,

For all measures being submitted for potential acceptance into the Trial Approval Program, each section must be filled out as completely as possible.

Respond to all questions as instructed with answers immediately following the question. All information on testing of either a sample data set or results from Bonnie testing that can demonstrate, to the extent possible, the the measure meets the reliability and validity must be in this form..

If you are unable to check a box, please highlight or shade the box for your response.

Maximum of 10 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***

Contact NQF staff regarding questions at trialmeasures@qualityforum.org

DATA and SAMPLING INFORMATION

1. DATA/SAMPLE USED FOR PRELMINARY TESTING OF THIS MEASURE

It is important that the measure developer use a data set to conduct preliminary testing in order to evauate the measure logic and the inclusions/exclusions for the population used in the measure.

What type of data was used for testing? (*The measure developer must provide a test data set that will provide some initial information to be used for the evaluation, or the Bonnie testing tool can use can be used to create a sample data set using synthesized patients.)* Please indicate whether the test data set used was provided through the measure developer, or through the Bonnie tool.

> The Bonnie testing tool was used to simulate a testing environment where measure specifications and HQMF output are tested against synthetic test data. Measure developers rely on the results in Bonnie to confirm whether the measure logic is performing as expected.
>
> Reference the eCQI Resource Center website (https://ecqi.healthit.gov/ecqm-tools/tool-library/bonnie) or the Bonnie testing tool website (https://bonnie.healthit.gov/) for more information about Bonnie functionality and its role in measure development. Please also reference the Bonnie testing worksheet attachment for detailed Bonnie test cases and testing results for this measure.

If Bonnie was NOT used, please identify the specifications for the test dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured)*

Not Applicable

What levels of analysis were tested (either through the test data set or Bonnie)? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*) *in order to determine its suitability for inclusion into the Trial Approval Program.,*

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency

☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other:

1.4. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis)*

48 unique synthetic patient records were created in the BONNIE testing system for this measure. Cases were used to test the validity of each data element and timing relationship in the measure. Bonnie testing was also performed for each stratum specified in the measure. Patient characteristics such as age, diagnosis, and length of stay were pre-determined to provide a variety of scenarios that adequately tested for patients passing each data element and failing each data element. Data included in cases and tested for this measure included all data elements required to calculate the measure and the measure denominator exclusions.

For further information on the characteristics of the patients included in the analysis, please refer to the attached BONNIE testing spreadsheet.

1.5. Please refer to the guidance for Bonnie testing found at this link. Bonnie testing results may be compiled into spreadsheet or table, which must be completed in its entirety, to the extent possible, in order to provide a basis for evaluation to determine the acceptability of the measure for inclusion in the Trial Approval program. *Any questions regarding the completion of this form can be directed to NQF Staff at trialmeasures@qualityforum.org.*

Please refer to the attached BONNIE testing spreadsheet.

<u>RELIABILITY AND VALIDITY ASSESSMENTS</u>

<u>Note</u>: The information provided in this next section is intended to aid the Standing Committee and other stakeholders in understanding to what degree the measure is both reliable and valid. While it is not possible to provide comprehensive results due to the lack of actual testing data, the developer needs to provide as much information as possible based on their interpretation of the results from the sample test data.

2.1 Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score. **What is your interpretation of the results in terms of demonstrating reliability**? (i.*e., what do the sample results mean and what are the norms for the test conducted?*) Please summarize the plan for future testing of reliability if the measure is accepted into the Trial Approval Program. **Include descriptions of:**

Inter-abstractor reliability, and data element reliability of all critical data elements

Computation of the performance measure score (e.g., signal-to-noise analysis)?

All data elements within the measure are specified using nationally accepted standard terminologies, including LOINC, SNOMEDCT, ICD10CM, and ICD10PCS. Bonnie testing confirms that the measure logic performs as expected and that the terminologies used are applied consistently. This suggests that organizations using these terminologies within the EHR should be able to produce repeatable and reliable results. For further discussion of measure feasibility, please review the attached feasibility scorecard and feasibility report.

When data are available, The Joint Commission will perform extensive tests of measure reliability at the data element and measure level. Testing will include re-abstraction to the eCQM specification to evaluate missing data and assure inter-rater reliability, as well as analysis of agreement rates for data elements used to compute measure rates for PBM-04.

2.2 Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score. **What is your interpretation of the results in terms of demonstrating validity**? (i.*e., what do the results mean and what are the norms for the test conducted?*). Please summarize the plan for future testing of validity if the measure is accepted into the Trial Approval Program. Include the method(s) of validity testing and what it will test (describe the steps—do not just name a method; what will be tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis will be used used)

Face validity was established through public comment.

Public comment was open for 30 days from March 20 to April 19, 2015. The Joint Commission received 150 responses to the call for comment. Respondents were asked to rate the measure on a number of parameters, using a Likert scale ranging from 1 to 5, where 1=Disagree and 5-Agree. The table below presents the average rating for these parameters.

PARAMETER	RATING
Numerator clearly describes the activity being measured	4.41
Denominator clearly describes the activity being measured	4.40
Numerator inclusions clear and appropriate	4.46
Denominator inclusions clear and appropriate	4.42
Numerator exclusions clear and appropriate	4.44
Denominator exclusions clear and appropriate	4.36
Accurately assesses the process of care to which it is addressed	4.31

Findings from public comment support the face validity of this measure.

The Bonnie testing tool and environment were used to establish content and construct validity through testing of the measure logic and value sets. Each data element and logic statement was tested to confirm actual results met expectations. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure. An example of negative testing would be to include test cases with pediatric ages to ensure that pediatric patients are not included in the measure.

Initial Population and Denominator test cases positively test to ensure that only patients >= 18 years of age who have a surgical procedure performed <=48 hours prior to the inpatient encounter or during the inpatient encounter are included. Negative test cases ensure that patients who do not meet these criteria to do not pass into the denominator. For example, cases test patients who have a surgical procedure at 49 hours and 48 hours prior to the start of the encounter. Patients who have a surgical procedure 48 hours

prior to the start of the encounter were included in the denominator, while patients with a surgical procedure at 49 hours prior to the encounter were not.

Numerator test cases positively test to ensure patients who have a hemoglobin result recorded <= 45 days and >=14 days prior to the start of surgery are included in the numerator. Negative test cases ensure that a patient who did not meet these criteria are not included. For example, test cases in which hemoglobin results were recorded >45 days prior to surgery or after surgery confirmed that such patients would not be included in the numerator.

Denominator exclusion test cases for this measure ensure that patients are properly removed from the denominator if they have specific documented procedures or encounter diagnoses. Negative test cases for the denominator exclusion ensure that patients without these diagnoses or procedures fall in to the denominator population. Testing confirmed patients meeting the exclusion criteria are removed from the measure appropriately, while those that do not meet the criteria are retained in the denominator population.

Once pilot data are available, The Joint Commission will evaluate construct validity though an examination of the degree of association between measure results for PBM-04 and other measures in this set, using the Pearson Correlation Coefficient. The Joint Commission would hypothesize that a relationship exists between this measure and other measures in the Patient Blood Management set.

In addition, data element validity would be assessed for accuracy and clarity in reliability testing, using the data element values obtained in the reliability study as the gold standard.

2.3 Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion. **What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results?** (*i.e., the value outweighs the burden of increased data collection and analysis).* Please summarize the plan for future testing of exclusions if the measure is accepted into the Trial Approval Program. Describe the method of testing exclusions and what it will test (describe the steps—do not just name a method; what will be tested, e.g., whether exclusions affect overall performance scores; what statistical analysis will be used)

When data are available, The Joint Commission will analyze exclusion frequency and variability across providers. These data elements to be analyzed include:

- Patients with a traumatic injury <=48 hours prior to or during the encounter.
- Patients with a solid organ transplant <=48 hours prior to or during the encounter.
- Patients who have an ECMO procedure during the inpatient encounter.
- Patients with sickle cell disease and related blood disorders

2.4 Risk Stratification (applicable ONLY to outcome or resource use measures). If an outcome or resource use measure will not be risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities. If risk adjustment/stratification is needed then please describe the conceptual/clinical and statistical methods and criteria that will be used to select patient factors (clinical factors or sociodemographic factors) that will be used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Not Applicable, not an outcome measure.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment **Attachment:** PBM04_NQF_Measure_Feasibility_Assessment_Report.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.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 and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

No modifications have been made

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets. Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.html)

There are no other fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public

domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Regulatory and Accreditation Programs Professional Certification or Recognition Program Quality Improvement (Internal to the specific organization)	

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

n/a

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

This is a new measure for which approval for trial use is requested.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

The Joint Commission maintains a certification program in Blood Management, which is a voluntary program for hospitals to achieve excellence in patient blood management. The measures in this set can be made available within a year for hospitals to use in fulfilling the requirements for certification. Hospitals using these measures evaluate care by these measures and submit data quarterly, either directly to The Joint Commission or through a vendor. The Joint Commission then generates reports and feeds the reports back to the certified organizations.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance

results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

n/a

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
n/a

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences identified during testing.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
n/a

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
n/a

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Tricia, Elliott, telliott@jointcommission.org, 630-792-5643-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Tricia, Elliott, telliott@jointcommission.org, 630-792-5643-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel was to provide advisory oversight in literature review, measure construct and content, review of testing results, and endorsement of draft and finalized measures, as well as to continue to provide measure content oversight and update in the future.

eCQM Blood Management Technical Advisory Panel Member List
Richard J. Benjamin, MD, PhD, FRCPath, MS
Chief Medical Officer, Biomedical Services
American Red Cross, National Headquarters
7/15/15:
Chief Medical Officer
Cerus Corporation Laurence Bilfield, MD
Orthopaedic Surgeon
Cleveland Clinic HS - Lutheran
Lawrence Tim Goodnough, MD
Director, Transfusion Service Stanford Medical Center
Associate Director, Stanford Blood Center
Stanford University Medical Center Joseph E. Kiss, MD
Associate Professor of Medicine; Dept. of Medicine; Div. of Hem/Onc
Medical Director, Hemapheresis and Blood Services, CBB/ITxM
The Institute for Transfusion Medicine
University of Pittsburgh
Harvey G. Klein, MD
Senior Investigator
Transfusion Medicine Department
National Institutes of Health

Vijay K. Maker, MD, FACCS
Chairman, Department of Surgery
Executive Director, MGH Residency in General Surgery
Advocate Illinois Masonic Hospital
John (Jeffrey) McCullough, MD
Professor, Clinical Pathology, Blood Banking
University of Minnesota
Steven Frank, MD
Medical Director, The Johns Hopkins Health System
Blood Management Program
Associate Professor, Johns Hopkins Hospital,
Department of Anesthesiology and Critical Care
Medicine, Division of Vascular, Thoracic, Transplant Anesthesia
Neil K. Shah, M.D.
Medical Director of Informatics for Transfusion Services
Medical Director of Referral (Send Out) Testing
Stanford University Medical Center
Aryeh Shander, MD, FCCM, FCCP
Executive Medical Director of The Institute for Patient Blood Management and Bloodless Medicine and Surgery
Englewood Hospital and Medical Center
Jonathan H. Waters, MD, Chair
Medical Director in the Blood Management Division of Procirca, Inc.
Chief and Professor
Magee Women's Hospital
University of Pittsburgh

The purpose of the eCQM Task Force is to engage eCQM implementers in the electronic specification process, in order to produce clear, implementable eCQM specifications. Task force membership includes both hospital and vendor representatives with expertise in clinical informatics, electronic health record (EHR) implementation, and standard terminologies, as well as content experts with experience leveraging the EHR for blood management.

ePBM Task Force Roster

Irwin Gross, MD
Medical Director of Transfusion Services
Eastern Maine Medical Center
Hugh H. Ryan, MD
Senior Director & Chief Medical Officer
Population Health Programs
Cerner Corporation

Kimberly Bodine, DNP, RN
EHR Manager, Clinical Quality Measures and Clinical Analytics
Health Corporation of America
Douglas Van Deale, MD, FACS
Chief Medical Information Officer
University of Iowa

Jason Kratz, PhD
Inpatient eCQM Development Lead
Business Intelligence Developer
Epic
Cathy Bickerstaff, RN-BC
Informatics Specialist
St. Jude's Children's Research Hospital

Andrew Higgins, RN
Patient Blood Management Coordinator
Mayo Clinic

Catherine A Shipp, RN
Transfusion Safety Officer
Loyola University Medical Center
David Krusch, MD
Chief Medical Information Officer
Professor of Surgery
University of Rochester Medical Center
Lisa Gulker, DNP, ACNP-BC
Senior Director, Applied Clinical Informatics
Tenet Healthcare

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2016

Ad.3 **Month and Year of most recent revision:** 05, 2016

Ad.4 **What is your frequency for review/update of this measure?** Annual

Ad.5 **When is the next scheduled review/update for this measure?** 05, 2017

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The measures and specifications are provided without warranty.

Ad.8 **Additional Information/Comments:**

NQF Measure Feasibility Assessment Report

Measure Title: PBM-04: Initial Transfusion Threshold

Measure Background and Overall Assessment of Measure Logic and Feasibility

The following assessment is conducted solely by the measure developer, The Joint Commission, based on our experience working with clinical experts, EHR and technical experts, and hospitals to assess feasibility throughout the measure development process. The measure was evaluated by five volunteer hospitals throughout the country during the fourth quarter of 2015.

This detailed report will provide a narrative summary of data elements found to be highly feasible, and will include verbatim scorecard responses for those data elements that were deemed to be more difficult to capture or for which there was great variability in feasibility. For complete scorecard results, please refer to the scorecard excel files attached to this submission.

Data Elements used in this Measure (in QDM format):

1. "Encounter, Performed: Encounter Inpatient" using "Encounter Inpatient SNOMEDCT Value Set (2.16.840.1.113883.3.666.5.307)"
2. "Laboratory Test, Performed: Hemoglobin blood serum plasma" using "Hemoglobin blood serum plasma Grouping Value Set (2.16.840.1.113762.1.4.1104.4)"
3. "Procedure, Performed: Blood Transfusion Administration" using "Blood Transfusion Administration SNOMEDCT Value Set (2.16.840.1.113762.1.4.1029.24)"
4. "Procedure, Performed: ECMO" using "ECMO Grouping Value Set (2.16.840.1.113762.1.4.1029.22)"
5. "Procedure, Performed: Solid Organ Transplant" using "Solid Organ Transplant Grouping Value Set (2.16.840.1.113762.1.4.1029.11)"
6. Attribute: "Diagnosis: Traumatic Injury" using "Traumatic Injury Grouping Value Set (2.16.840.1.113762.1.4.1029.10)"
7. Attribute: "Diagnosis: Sickle Cell Disease and Related Blood Disorders" using "Sickle Cell Disease and Related Blood Disorders Grouping Value Set (2.16.840.1.113762.1.4.1029.35)"

Initial Population and Denominator Data Elements

Data elements 1- "Encounter, Performed: Encounter Inpatient" and 3- "Procedure, Performed: Blood Transfusion Administration" are used to define the initial population and denominator of this measure.

On the feasibility scorecard, hospitals rated these data elements 1 and 3 as highly feasible when considering workflow, data availability, accuracy, definition, and use of standards.

Four out of five hospitals rated capture of data element 1 as highly feasible, represented as a score of 3 out of 3, for all domains of feasibility in both the current state and in the future. One site was not certain whether the data source for this data element was currently interfaced with the certified electronic health record. This site scored feasibility as a 1 for all domains in the current state, but as a 3 for future state, acknowledging that future state would be achieved much more quickly than the 3-5 year timeframe outlined in the scorecard, as the site would be interfacing this data in 2016 in order to report eCQMs.

Three out of five sites found data element 3 to be highly feasible in all domains except Data Standards- these hospitals have structured data fields for capture of transfusion data, but do not have those fields encoded in the terminology standard used in this measure, SNOMEDCT. These 3 sites reported that data capture would be highly feasible in the near term, stating that mapping this field to SNOMED would not be difficult and could be accomplished rather quickly. These three sites rated future state feasibility 3 out of 3, stating that capturing blood products in SNOMEDCT could occur in a much shorter timeframe than 3-5 years.

Two sites placed orders for blood in the EHR, but recorded blood product administration on paper. Both sites had plans to move to EHR-based barcode blood product administration, and found data element 3 to be highly feasible in the future state.

Numerator Data Element

Data element 2- "Laboratory Test, Performed: Hemoglobin blood serum plasma" is used to define the numerator for this measure. While some measures in this set require hemoglobin results recorded prior to the start of the encounter, PBM-04 evaluates hemoglobin results recorded within 45 days of the first blood transfusion. All sites have policies and practices in place that require a hemoglobin result prior to transfusion, and thus found this data element to be highly feasible, represented as a score of 3 out of 3 for all domains of feasibility.

Denominator Exclusions Data Elements

Data elements 4, 5, 6, and 7 are used to represent denominator exclusions.

Data element 4- "Procedure, Performed: ECMO," was found to be highly feasible by sites that perform ECMO. One site, a regional hospital, reported frequently using ECMO as a bridge for transport for patients requiring a higher level of care.

Data elements 6- "Diagnosis: Traumatic Injury," and 7- "Diagnosis: Sickle Cell Disease and Related Blood Disorders" both represent encounter diagnoses. All hospitals rated these data elements as highly feasible. Discussion around these data elements suggested that while missing data may occur due to clinician practice related to updating the patient problem list, the functionality to support collection of this data element is well established.

Feasibility for data element 5- "Procedure, Performed: Solid Organ Transplant" was found to be comparable to data element 6- "Procedure, Performed: Selected Elective Surgical Procedures." These data elements are found in the surgical schedule or operative record, and thus findings were similar, with the exception of sites that do not perform organ transplant, which would not use this data element.

Conclusion

Hospitals completing the feasibility scorecard reported the data elements required to calculate this measure to be highly feasibility in the current state. Of the measures in the PBM set, this measure received the highest ratings for feasibility. Approval for Trial Use status will support The Joint Commissions' efforts to further test this measure.

Appendix B

Comment ID 6370 on #1543: Postoperative Stroke or Death in Asymptomatic Patients undergoing Carotid Artery Stenting (CAS

Thank you for the opportunity to comment on the Society for Vascular Surgery's quality measure #1543 - Postoperative Stroke or Death in Asymptomatic Patients undergoing Carotid Artery Stenting (CAS). The Society for Cardiovascular Angiography and Interventions (SCAI) is a non-profit professional association representing the majority of interventional cardiologists and cardiac cath lab teams in the United States. We believe that the review of the medical literature that was discussed while this Committee was deliberating omitted significant amounts of data and that it was inaccurate in summarizing the clinical guidelines on carotid stenting. The medical literature surrounding the NQF's review should be balanced and free from bias or self-promotion. We are concerned that comments offered by SVS did not encompass this ideal completely, instead focusing on highly selected subsets of data in favor of their position, but not including the totality of substantial evidence. As noted in the discussion below, we find conflict in their recommendation of support for carotid stenting in asymptomatic high risk surgical patients via surgical cut down based on one small trial, yet recommend against carotid stenting via percutaneous access in the same patients despite an overwhelming rigorous data base.

SCAI specific position stated below incorporates the following beliefs. We believe that both carotid endarterectomy and carotid stenting offer safe and effective treatment alternatives for patients with obstructive carotid disease, provided that the procedure is performed by an experienced operator with proven results and in carefully selected patients. Furthermore, we believe that these therapies are complementary, not competitive. There are clearly patients who are at high risk for endarterectomy who are at low risk for carotid stenting and those who are high stent risk but low surgical risk, and some patients who are better served by medical therapy without either form of revascularization. Furthermore, as we believe that the two therapies are equally effective at preventing strokes and maintaining carotid patency (discussed below), and that the vast majority of strokes occur in asymptomatic patients, there is a role for both interventions in carefully selected asymptomatic patients.

SCAI's position on carotid artery stenting in asymptomatic patients: Carotid stenting in asymptomatic patients can be recommended if:
1. Patient has a documented carotid stenosis of > or = 80% (2 modalities), AND
2. Patient has a life expectancy (of quality) of > or = 5 years, AND
3. Is performed by an experienced, certified operator with documented acceptable event rates, AND
4. The patient has pre- and post- procedural independent neurological evaluation AND is enrolled in a prospective data base AND
5. That the patient's risk/benefit for CAS individualized to their medical condition and carotid anatomy AND to the operator/institution results is at least as good as that of the same patient's individualized risk/benefit ration for carotid endarterectomy or medical therapy.

Finally this statement and our response to the SVS's statements are based on extremely

rigorous data bases including all 3 North American randomized comparisons of CEA vs. CAS (SAPHIRE, CREST, and ACT 1), the 10 year follow-up data on the CREST randomized trial (all four of these published in the New England Journal of Medicine), extremely large real world carefully regulated registries (representing over 40,000 patients prospectively studied with independent neurological adjudication), multiple IDE trials, and meta-analyses.

Based on extensive data, we believe that the totality of extensive data should be included and considered. The comprehensive data shows:

1. We acknowledge statistically increased risk of minor strokes for CAS over CEA in CREST. However, that alone cannot be used as the only determinant in decisions, and must be considered in the context of all information.

a. First despite increased peri-procedural minor strokes, there is NO difference in neurological outcomes in these patients. Careful neurological assessments in CREST showed no difference between CAS and CEA patients, and multiple trials have shown return to neurologic baseline in patients with small peri-procedural strokes b. These studies were done relatively early in CAS experience AND did not utilize proximal protection. Operator experience, improved case selection, and proximal protection have all been shown to reduce peri-procedural events.

2. There is no difference in major strokes or death between the two therapies in any of the North American randomized trials.
3. There is a consistent increase in cranial nerve injuries of CEA over CAS. While some resolve, others persist. If these were included to the same extent that minor peri-procedural strokes (which resolve as above), there would be no advantage of CEA over CAS.
4. There is NO difference between the 2 therapies for preventing long term ipsilateral strokes. BOTH therapies are excellent and durable.
5. There is either no difference or a slight advantage to carotid stenting in maintaining long term carotid patency. Again, both are excellent.
6. There is significant increase for peri-procedural myocardial infartions in CEA patients over CAS.

We believe this is very important to consider and monitor, as post procedural MI's do correlate with increased mortality, and patients with carotid artery disease frequently have concomitant coronary artery disease. In summary, we believe that there is compelling evidence (including the recent ACT 1 and 10 year CREST randomized trials) that carotid artery stenting is a viable treatment option for patient with severe carotid artery disease, especially those at high risk for carotid endarterectomy. To exclude it as a treatment option based on a small increased rate of peri-procedural minor strokes in earlier trials would be every bit as unfair as rejecting endarterectomy because of an increased rate of cranial nerve injuries or peri-procedural myocardial infarctions. As above, we believe that both therapies when performed by experienced operators in appropriately selected patients are excellent complementary procedures. We advocate for all societies to strive for established standards for carotid revascularization in a collaborative fashion, emphasizing the optimal benefits for patients independent of specialty bias.

I would like to thank D. Chris Metzger, MD, FSCAI, the Chair of our Carotid Stenting Committee

for developing these public comments.
Sincerely,
/s/
Kenneth Rosenfield, MD, MHCDS, MSCAI

To: National Quality Forum (NQF) Surgery Standing Committee Members

From: Pam Owens and Mamatha Pancholi, AHRQ Quality Indicators Project
Agency for Healthcare Research and Quality
On behalf of the entire AHRQ QI Team

Re: Additional Information regarding Maintenance Measure NQF 0351: Death Rate Among Surgical Inpatients with Serious Treatable Complications (PSI 04)

Date: October 21, 2016

AHRQ appreciated the opportunity to have the maintenance measure NQF 0351: Death Rate Among Surgical Inpatients reviewed by the NQF Surgery Standing Committee. Reviews by external stakeholders allows AHRQ to inform multiple experts about measure refinements, discuss and understand concerns of the measures, and be able to address those concerns and refine to make the measures even better.

As a follow-up to the NQF Surgery Standing Committee review on August 16, 2016, AHRQ would like to share with the Committee additional information that was requested during the review and present additional proposed refinements based on reviewers' concerns.

The *NQF-Endorsed Measures for Surgical Procedures 2015-2017: Draft Report for Comment* (September 22, 2016) noted that the reviewers wanted additional information and/or had concerns about the following areas:

- "Risk-adjustment strategy includes patients transferred in with complications present on admission"
- "Claims data cannot accurately capture complications reliably"
- "Does not include the transfers out thus providing a potential for 'gaming'"
- "Absence of testing data that demonstrates the measure assesses what it is supposed to measure"

In the paragraphs below, we respond to these concerns and describe enhancements to the risk-adjustment approach that AHRQ is planning to implement to PSI 04 in the next release of the Patient Safety Indicators, v6.0.2. These enhancements will carry forward into the ICD-10 PSI software, as soon as an adequate period of ICD-10 coded data is available for predictive modeling.

"Risk-adjustment strategy includes patients transferred in with complications present on admission"

This question has been a subject of considerable attention and analysis over the 12 years since PSI 04 was introduced, the 14 years since Needleman and Buerhaus (*N Engl J Med* 2002;346(22):1715-22) first described this approach to operationalizing "failure to rescue," and the 24 years since Silber and colleagues (*Med Care* 1992;30(7):615-629) first described the concept of "failure to rescue".

First, it is important to recognize that NQF recently re-endorsed the related (not competing) Silber/Children's Hospital of Philadelphia (CHOP) version of this measure, with two different specifications of the outcome variable, "0353 Failure to Rescue 30-Day Mortality (risk adjusted)" and "0352 Failure to Rescue In-Hospital Mortality (risk adjusted)." Both of these specifications of the "failure to rescue" concept include ALL 30-day deaths and ALL inpatient deaths, respectively. In other words,

not only do these existing NQF-endorsed measures include deaths among patients transferred in with complications present on admission, but they even include deaths among patients who never experienced any reported complication (see http://www.qualityforum.org/QPS/0352).

In previous discussions with the Patient Safety Standing Committee, the 0352/0353 measure stewards from CHOP provided extensive support for the concept of NOT limiting the denominator to patients who experienced a complication acquired in the same hospital. There are two important conceptual arguments supporting this decision:

1. PSI 04 and other measures of "failure to rescue" focus on the progression from complication to death (and the hospital's ability to influence that progression); whether the hospital was responsible for causing the complication during the same admission, during a prior admission, or not at all (i.e., the complication originated in ambulatory surgery), is not material to that focus. PSI04 is a risk-adjusted inpatient surgical mortality measure in which the denominator is limited to patients with certain complications that often arise in association with surgical care. Prior studies have suggested that high-quality care can prevent similar proportions of these deaths, whether the complication happened to originate during the same hospital stay or not.
2. AHRQ and other measure developers attempt to design measures in a manner that minimizes unintended consequences and reduces opportunities for "gaming." Limiting the denominator for PSI 04 to patients with complications reported as hospital-acquired (i.e., not present on admission) would encourage hospitals to shift blame for complications, which is an unproductive exercise. In other words, a serious complication may lead to death if it is not recognized and treated in a timely manner; this downward trajectory can often be interrupted with aggressive treatment. The goal of PSI 04 and other measures of "failure to rescue" is to reward hospitals for providing such treatment, regardless whether the complication started within that hospital's walls or outside its walls. AHRQ does not wish to encourage hospitals to waste effort on determining exactly when a complication started, given that the evidence-based focus of this indicator is on treating complications, not preventing them.

Thus, this critique of PSI 04 was already discussed extensively, and set aside, during last year's successful re-endorsement process for "0353 Failure to Rescue 30-Day Mortality (risk adjusted)" and "0352 Failure to Rescue In-Hospital Mortality (risk adjusted)." AHRQ encourages the Surgery 2015-2017 Standing Committee to consider the previous work of the Patient Safety Standing Committee, as consistent evaluation processes benefit all of NQF's stakeholders.

Second, AHRQ is revising its PSI 04 risk-adjustment model for the next release, v6.0.2, by adding adjustors for whether the denominator-triggering complication was present on admission, and whether it was relatively mild or severe. This addition is superimposed on stratified risk-adjustment models, in which patients with different complications are allowed to have different predictors of death, and different relationships between age, gender, and transfer status and the risk of death. In other words, the v6.0.2 risk models will explicitly account for whether a denominator-triggering complication was present on admission, and whether it was mild or severe at the time of presentation. For example:

- Stratum A includes deep vein thrombosis (DVT) and pulmonary embolism (PE); the latter diagnosis is now considered to be more severe than the former.
- Stratum B includes pneumonia; staphylococcal, gram negative, anaerobic, and aspiration pneumonias are now considered to be more severe than other types of pneumonia.
- Stratum C includes sepsis; sepsis with septic shock or acute organ system dysfunction is now considered to be more severe than uncomplicated sepsis.

- Stratum D includes shock and cardiorespiratory arrest; the latter diagnosis is now considered to be more severe than the former.
- Stratum E includes all types of gastrointestinal (GI) hemorrhage; GI bleeding with perforation is now considered to be more severe than GI bleeding without perforation.

Conceptually, this change will reduce any residual bias resulting from the transfer of patients with severe complications.

AHRQ's updated models for v6.02 have significantly higher c statistics than the current v6.01 models (see details in Appendix A). Specifically, the re-estimated c statistics, based on the AHRQ 37-state reference population, increased from 0.780 to 0.797 for Stratum A, from 0.771 to 0.782 for Stratum B, from 0.726 to 0.776 for Stratum C, from 0.715 to 0.818 for Stratum D, and from 0.860 to 0.878 for Stratum E.

These revised c statistics, representing the ability of the model to discriminate between patients who survived and patients who died, are very consistent with other NQF-endorsed mortality measures, including postoperative morbidity and mortality measures based on registry data. For example, among measures reviewed in 2016 for the same project (Surgery 2015-2017), developers reported c statistics of 0.716-0.719 for the American College of Surgeons' (ACS) "Risk Adjusted Colon Surgery Outcome Measure," 0.65-0.70 for CMS' "Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)," 0.708-0.738 for the STS "Mitral Valve Repair/Replacement (MVRR) + Coronary Artery Bypass Graft (CABG) Composite Score," 0.708-0.807 for the STS "Individual Surgeon Composite Measure for Adult Cardiac Surgery," and 0.758-0.772 for the ACS "Risk Adjusted Case Mix Adjusted Elderly Surgery Outcomes Measure." (http://www.qualityforum.org/ProjectMaterials.aspx?projectID=80864)

Therefore, the argument that the risk-adjustment strategy for PSI 04 is inadequate lacks empirical foundation. AHRQ believes that the risk-adjustment upgrade to be implemented in v6.0.2 will address the remaining concerns of hospitals that receive transfers from other hospitals.

"Claims data cannot accurately capture complications reliably"

As described above, the concept of "failure to rescue" or "death given a complication" has never hinged on the precise level of coding accuracy. The intent of "failure to rescue" measures is not to "ding" hospitals for complications, but to identify hospitals that are able to recognize complications early and treat them aggressively to reduce the risk of death.

This criticism is also puzzling because there is no clear connection between inaccurate coding and PSI 04 performance. Inpatient death—the numerator or outcome event for PSI 04—is widely understood to be accurately reported. Thus, the only ways a hospital could artificially reduce its PSI 04 rate through inaccurate coding would involve manipulation of the *denominator* by either: (1) preferentially decreasing the coding of denominator complications among patients who subsequently died; or (2) preferentially increasing the coding of denominator complications among patients who did *not* subsequently die. (In epidemiologic language, clinically significant information bias would require *nondifferential* misclassification of complications, with respect to an outcome that was unknown when the complication occurred and was diagnosed by the treating physician.) Numerous factors argue against the likelihood of such manipulation: (1) when complications are diagnosed, it is frequently unpredictable which patients will go on to die as inpatients; (2) even if it were possible to predict which

patients will die, it would be implausible that providers would suppress the accurate description and treatment of a complication for the sake of PSI 04 performance; and (3) the most plausible means of gaming PSI 04—systematic over-coding of complications such that the denominator might be inflated with milder cases unlikely to result in death—would be counter-intuitive to providers and an inefficient means of influencing PSI 04 rates. For example, the most common evidence-based critique of PSI 12 – that it is susceptible to "surveillance bias" or "overdiagnosis bias" from inter-hospital variation in how postoperative patients are screened for venous thrombosis – would lead hospitals with falsely high PSI 12 rates to have falsely low PSI 04 rates. In other words, a hospital attempting to suppress its PSI 04 rate could do so by exaggerating its PSI 12 rate – an implausible scenario on its face.

Finally, it should be noted that the complications captured in the denominator of PSI 04 are serious, often life-threatening complications (i.e., DVT, PE, pneumonia or aspiration, sepsis, shock or cardiac or respiratory arrest, GI hemorrhage) that affect MS-DRG assignment in the Inpatient Prospective Payment System, so failing to code them would trigger underpayment, whereas consistently overcoding them would trigger auditing and financial penalties. The underlying ICD-coded data are widely accepted as sufficiently accurate to determine hospital payment (through MS-DRG complication or "CC/MCC" assignments), so they should logically be accepted as sufficiently accurate to identify patients who had complications that placed them at risk of death.

"Does not include the transfers out thus providing a potential for 'gaming'"

Patients transferred-out are excluded because the outcome of the hospital episode is unknown for these patients, in the absence of linked data across multiple hospitals. In other words, the numerator for PSI 04 is in-hospital death, but if the patient is appropriately transferred to a regional referral center for more advanced care, the patient's outcome is unknown to the referring hospital.

Conceptually, it may be preferable to attribute the outcomes of transferred patients proportionally to both referring hospitals and receiving hospitals, since both hospitals were involved in providing care. In some cases, referring hospitals may transfer patients "too late" for receiving hospitals to provide effective treatment, and it would be appropriate to place most of the "blame" on the referring hospital. In other cases, referring hospitals may transfer patients immediately, but receiving hospitals may either fail to deliver effective and timely treatment, or may be powerless to do so due to the patients underlying condition(s). This idea of "split attribution" merits methodologic attention, but would be difficult to operationalize. NQF has a separate process currently underway to promote more transparent, reproducible, and valid approaches to attribution. AHRQ is not aware of any currently endorsed quality measures, whether based on claims data, registry data, or electronic health records, that "split attribution" across multiple hospitals.

Given that there is currently no validated methodology for "split attribution" between referring and receiving hospitals, AHRQ continues to favor exclusion of these patients. Empirically, this exclusion has virtually no effect on the distribution of risk-adjusted, smoothed PSI 04 rates across hospitals, because the hospitals that transfer out more than 10% of their eligible patients are usually small hospitals. These small hospitals have smoothed observed/expected ratios close to one, indicating that their observed results (without smoothing) are not very reliable. Approaches of this type have become standard practice among measure developers, and have been extensively reviewed by NQF committees.

"Absence of testing data that demonstrates the measure assesses what it is supposed to measure"

This concern is demonstrably false. AHRQ believes that the NQF Surgery Standing Committee may have misinterpreted NQF's 2016 Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. This document (Algorithm 3) specifically emphasizes "validity testing computed with performance measure scores for each measure entity" over "validity testing conducted with patient-level data elements." Recommended approaches to "validity testing computed with performance measure scores" include "(1) correlation of the performance measure score on this measure and other performance measures; (2) differences in performance scores between groups known to differ on quality; and (3) other accepted method…" AHRQ presented substantial and unrefuted evidence of "construct validity" based on these approaches. Such findings are the only "testing data" requested by the NQF for measures of this type. Evidence of "patient-level data element" validity is only required – indeed, only helpful – if evidence of validity of "performance measure scores" is unavailable or conflicting, which does not apply in the case of PSI 04.

Specifically, the Evidence form under Section 1a.8.2 includes an extensive and detailed environmental scan of the literature, demonstrating that measures of "failure to rescue," including PSI 04, are consistently associated with many hospital-level measures of high quality care, including higher nurse-to-bed ratios, better nurse skill mix ratios (i.e., baccalaureate-trained nurses), higher US-trained nurse ratios, Magnet designation by the American Nurses Credentialing Center, and the Practice Environment Scale of the Nursing Work Index. Higher hospital volume was associated with lower rates of "failure to rescue," based on multiple specifications of the concept, in at least six studies. The Measure Testing form under Section 2b.2.3 reports confirmatory analyses using PSI 04, as specified by AHRQ. The analyses reported included expected differences in performance scores between teaching and non-teaching hospitals, between high-technology and low-technology hospitals, between large and small hospitals, between hospitals with high and low nurse staffing levels, and between hospitals with better and poorer nursing skill mix.

Additional evidence of face validity was also reported under Section 2b.2.3, although evidence of face validity is not required for a "high" or "moderate" rating of validity under NQF's 2016 Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement.

AHRQ has confirmed some of these findings in updated analyses after the NQF Standing Committee discussion this summer. For example, AHRQ's analysis of both the current v6.01 and proposed v6.02 risk-adjusted measures indicates that major teaching hospitals, defined as hospitals with 0.2500 to 0.5999 residents per licensed bed, have an aggregate observed-to-expected ratio of 0.969 to 0.972. In other words, major teaching hospitals have about 3% fewer deaths than expected, based on the five risk-adjustment models described above. By comparison, non-teaching hospitals (with 45.3% of all eligible patients) have aggregate observed-to-expected ratios of 1.002 to 1.007. Hospitals with unknown teaching status in the American Hospital Association's survey database (with 2.3% of all eligible patients) have aggregate observed-to-expected ratios of 1.032-1.046. In other words, non-teaching hospitals consistently demonstrate more deaths than expected, whereas major teaching hospitals consistently demonstrate fewer deaths than expected.

In summary, PSI 04 and other measures of "failure to rescue" have demonstrated validity at the hospital-score level because of their correlation with structural and process measures of quality, particularly those related to the nursing workforce. There is no evidence that these correlations are appreciably affected by the details of how denominator-triggering complications are specified. For

example, Mattke et al (*Med Care* 2004; 42(2 Suppl):II21-33) found that hospital rankings of failure-to-rescue for surgical patients (of which AHRQ's PSI 04 is a widely used specification) were not influenced appreciably by denominator definitions: "the indicator's estimate of relative hospital performance was fairly robust, even if specific patient cases were misclassified in terms of the timing of complications. This finding comports with the original FTR definition, which did not require that complications occur during the hospital stay." Silber et al's analysis (*Med Care* 2007; 45(10):918-25) also supports having as broad a denominator definition as possible, including conditions that were present on admission. Construct validity relationships were consistent across all tested specifications of failure-to-rescue, including PSI 04. Absent any empirical evidence that restricting the denominator to conditions that arose during the same hospital stay would increase the validity of the indicator, AHRQ has chosen to retain fidelity with the original concept of failure to rescue, as it was developed by Silber et al. (*Med Care* 1992; 30:615-29) and adapted by Needleman et al. (*N Engl J Med* 2002; 346(22):1715-22). As described by Needleman and Buerhaus (*Med Care* 2007; 45(10:913-5), "We have argued elsewhere that with adequate risk adjustment and use of methods that compare actual to expected rates of complications, inclusion of cases in which complications were present on admission can be controlled… FTR-N was developed with some sensitivity to these concerns… which contributed to its adaptation by AHRQ as FTR-A. As noted above, FTR-N was developed to be a nursing sensitive measure. It was constructed ex ante around complications that nursing was believed to influence and for which early identification and intervention by nurses might be central to rescue."

Additional, more recent literature is summarized in Appendix B of the document (see also the National Health Institute for Health Research review at: http://www.crd.york.ac.uk/prospero/display_record.asp?ID=CRD42013004080) . This literature again highlights that PSI 04 should be viewed principally as a measure of hospital team performance, due to its particular sensitivity to aspects of nursing skill mix and the nursing work environment. It should not be viewed as a measure of quality for individual surgeons.

Appendix A. Revised PSI 04 Risk Models

Source: Agency for Healthcare Research and Quality, Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases, 2013

Table 1. Revised Risk Model for PE/DVT Stratum (PSI 04A)
Table 2. Revised Risk Model for Pneumonia Stratum (PSI 04B)
Table 3. Revised Risk Model for Sepsis Stratum (PSI 04C)
Table 4. Revised Risk Model for Shock/Cardiac Arrest Stratum (PSI 04D)
Table 5. Revised Risk Model for GI Hemorrhage / Acute Ulcer Stratum (PSI 04E)

Table 1. PSI 04A Risk Model: Pulmonary Embolism / Deep Vein Thrombosis Stratum

		Current Software (v60rev2)							Modification 1						
		C Stat = 0.780							C Stat = 0.797						
Variable	Description	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model
Intercept	Intercept	1	-3.563444	-3.840335943	-3.286552656	636.2334	0	GEE	1	-3.86802	-4.166441187	-3.569607861	645.3993	0	GEE
M_AgeCat_1	Male | Age < 30	1	-0.927206	-1.845744996	-0.00866627	3.914288	0.0478773	GEE	1	-1.13468	-2.086159647	-0.18319945	5.463151	0.0194216	GEE
M_AgeCat_2	Male | Age < 35	1	-0.173093	-1.00452195	0.658336814	0.166495	0.6832454	GEE	1	-0.12213	-0.958506075	0.714248487	0.081908	0.7747278	GEE
M_AgeCat_3	Male | Age < 40	1	0.229203	-0.509239462	0.96764623	0.370088	0.5429566	GEE	1	0.232201	-0.513671971	0.978074151	0.372302	0.5417524	GEE
M_AgeCat_4	Male | Age < 45	1	-0.184039	-0.787758174	0.419679445	0.356984	0.5501863	GEE	1	-0.18702	-0.79638901	0.422343405	0.36185	0.5474806	GEE
M_AgeCat_5	Male | Age < 50	1	0.234726	-0.249436348	0.718889109	0.902895	0.3420066	GEE	1	0.212607	-0.276352432	0.701566034	0.726282	0.3940911	GEE
M_AgeCat_6	Male | Age < 55	1	0.010089	-0.426636274	0.44681477	0.00205	0.9638848	GEE	1	0.030812	-0.410366018	0.471989867	0.018737	0.8911224	GEE
M_AgeCat_7	Male | Age < 60	1	0.063575	-0.336060373	0.463209438	0.097216	0.755197	GEE	1	0.091324	-0.311598586	0.494246826	0.197344	0.6568734	GEE
M_AgeCat_8	Male | Age < 65	1	-0.351013	-0.727590447	0.025564453	3.337597	0.0677134	GEE	1	-0.35276	-0.732675045	0.027156055	3.311917	0.0687792	GEE
M_AgeCat_9	Male | Age >=65	1	-0.063487	-0.414630142	0.287656428	0.125573	0.7230674	GEE	1	-0.05356	-0.408248623	0.301125905	0.087601	0.7672497	GEE
M_AgeCat_10	Male | Age >=70	1	0.148922	-0.192680009	0.49052491	0.730087	0.392855	GEE	1	0.164046	-0.182234041	0.5103253	0.862128	0.353144	GEE
M_AgeCat_11	Male | Age >=75	1	0.223227	-0.118043758	0.564497446	1.643582	0.1998348	GEE	1	0.228109	-0.116963604	0.573182517	1.678652	0.1951037	GEE
M_AgeCat_12	Male | Age >=80	1	0.096707	-0.276991629	0.470405899	0.257259	0.6120098	GEE	1	0.135667	-0.2410733	0.51240791	0.498152	0.4803133	GEE
M_AgeCat_13	Male | Age >=85	1	0.396169	-0.028924023	0.821261905	3.336485	0.0677592	GEE	1	0.410781	-0.01905072	0.84061321	3.508488	0.0610552	GEE
F_AgeCat_2	Female | Age < 35	1	-0.892643	-2.009796739	0.224510577	2.452596	0.1173307	GEE	1	-0.83183	-1.962210239	0.298555924	2.080229	0.1492178	GEE
F_AgeCat_3	Female | Age < 40	1	-0.500452	-1.44188958	0.440985328	1.085521	0.2974657	GEE	1	-0.47443	-1.429518171	0.480666847	0.947854	0.330266	GEE
F_AgeCat_4	Female | Age < 45	1	-0.267878	-0.947917326	0.412161877	0.596075	0.4400795	GEE	1	-0.30112	-0.986024068	0.38378768	0.74252	0.3888547	GEE
F_AgeCat_5	Female | Age < 50	1	0.2302	-0.332398324	0.792798682	0.643147	0.4225735	GEE	1	0.312415	-0.252351905	0.877181451	1.175496	0.2782748	GEE
F_AgeCat_6	Female | Age < 55	1	-0.134632	-0.634334008	0.365070277	0.278849	0.5974565	GEE	1	-0.18666	-0.687711697	0.314398559	0.533104	0.4653049	GEE
F_AgeCat_7	Female | Age < 60	1	-0.23221	-0.646474635	0.182054377	1.206988	0.2719294	GEE	1	-0.18652	-0.603640568	0.230599187	0.768119	0.3807996	GEE
F_AgeCat_8	Female | Age < 65	1	0.070329	-0.297009656	0.437668063	0.14081	0.707477	GEE	1	0.055578	-0.315340962	0.426497317	0.086247	0.7690031	GEE
F_AgeCat_10	Female | Age >=70	1	0.101705	-0.262091685	0.465501953	0.300237	0.5837339	GEE	1	0.112962	-0.253583245	0.479507488	0.364843	0.5458286	GEE
F_AgeCat_11	Female | Age >=75	1	0.162337	-0.205094265	0.529768799	0.749859	0.3865208	GEE	1	0.157247	-0.212347169	0.526840468	0.69536	0.4043473	GEE
F_AgeCat_12	Female | Age >=80	1	0.361002	0.00835274	0.713650389	4.025589	0.044815	GEE	1	0.397913	0.042737521	0.753088323	4.821548	0.0281061	GEE
F_AgeCat_13	Female | Age >=85	1	-0.183121	-0.575453125	0.209211718	0.836879	0.3602909	GEE	1	-0.16556	-0.560553095	0.229427119	0.674917	0.4113426	GEE
TRNSFER	Transfer	1	0.500719	0.322278476	0.679158963	30.24811	3.802E-08	GEE	1	0.555854	0.374630438	0.737078413	36.1399	1.836E-09	GEE
ANEMDEF	Deficiency Anemias	1	-0.357698	-0.524211962	-0.191183174	17.72657	2.55E-05	GEE	1	-0.29133	-0.460048232	-0.122618107	11.45429	0.0007133	GEE
CHF	Congestive heart failure	1	0.375465	0.176247595	0.574682921	13.6452	0.0002208	GEE	1	0.394166	0.192793312	0.595539294	14.71813	0.0001248	GEE
COAG	Coagulopthy	1	0.556899	0.355336506	0.758462194	29.32437	6.122E-08	GEE	1	0.628221	0.422871577	0.833570692	35.95288	2.021E-09	GEE
DEPRESS	Depression	1	-0.36943	-0.616278178	-0.122581016	8.603958	0.0033543	GEE	1	-0.35421	-0.602758347	-0.105659952	7.801732	0.0052196	GEE
IMMUNE	Immune disorders	1	1.013531	0.86088331	1.166178793	169.3517	0	GEE	1	1.052148	0.897139906	1.207156754	176.9862	0	GEE
LYTES	Fluid and electrolyte disorders	1	0.617964	0.473460152	0.762468113	70.25271	0	GEE	1	0.692542	0.54545148	0.83963174	85.1571	0	GEE
MDC_4	MDC 4: Respiratory System	1	1.139684	0.901583737	1.377784437	88.01265	0	GEE	1	1.089303	0.84831973	1.330286945	78.49082	0	GEE
MDC_5	MDC 5: Circulatory System	1	0.79997	0.583907685	1.016032434	52.66055	0	GEE	1	0.776844	0.558282448	0.995404893	48.53077	3E-12	GEE
METS	Metastatic cancer	1	0.633645	0.444658246	0.822631793	43.18425	5E-11	GEE	1	0.690866	0.498367714	0.883363964	49.48011	2E-12	GEE
RENLFAIL	Renal failure	1	0.159237	-0.016560573	0.335035343	3.1518	0.0758433	GEE	1	0.24453	0.065198676	0.423861674	7.142462	0.007528	GEE
TUMOR	Solid tumor w/out metastasis	1	0.339309	0.051212224	0.627406323	5.328562	0.0209787	GEE	1	0.436592	0.145685504	0.727497759	8.652493	0.0032661	GEE
VALVE	Valvular disease	1	0.28366	0.008681109	0.558638542	4.087838	0.0431929	GEE	1	0.263734	-0.013002132	0.540469423	3.488965	0.0617792	GEE
mdrg_2406	Craniotomy for multiple significant trauma	1	2.922676	2.146766222	3.698585734	54.50484	0	GEE	1	3.016712	2.22594763	3.807477156	55.90741	0	GEE
mdrg_402	Other resp system O.R. procedures	1	-0.480927	-0.829265842	-0.132587588	7.322333	0.0068103	GEE	1	-0.38229	-0.735811294	-0.028763587	4.49199	0.034054	GEE
mdrg_502	Perc cardiovasc proc w non-drug-eluting stent	1	-0.469873	-1.079009656	0.139263614	2.285745	0.1305674	GEE	1	-0.33994	-0.956840652	0.276967674	1.166424	0.2801373	GEE
mdrg_509	Amputation for circ sys disorders	1	-0.52962	-1.244099174	0.184859723	2.11079	0.1462638	GEE	1	-0.22805	-0.954013904	0.497923813	0.379054	0.5381098	GEE
mdrg_1802	Postoperative or post-traumatic infections	1	-1.43342	-2.579978437	-0.286861161	6.004135	0.0142724	GEE	1	-1.3124	-2.461865298	-0.16292603	5.007613	0.0252361	GEE
mdrg_2104	Other O.R. procedures for injuries	1	-1.361083	-2.357080435	-0.365085221	7.173791	0.0073976	GEE	1	-1.30854	-2.308268166	-0.308807168	6.581165	0.0103063	GEE
mdrg_102	Cranio w major dev impl/acute complex CNS	1	1.334762	0.919640532	1.749882935	39.71493	2.94E-10	GEE	1	1.416502	0.99386544	1.839137575	43.15162	5.1E-11	GEE
mdrg_104	Spinal procedures	1	-0.92458	-2.078639386	0.229479076	2.465637	0.1163605	GEE	1	-0.99475	-2.155082781	0.165583752	2.823307	0.0929048	GEE
mdrg_511	Perc cardiovasc proc w drug-eluting stent	1	-0.514923	-0.967864041	-0.061982843	4.964769	0.0258688	GEE	1	-0.42435	-0.884009886	0.035318179	3.273822	0.0703934	GEE
mdrg_513	Perc cardiovasc proc w/o coronary artery stent	1	-0.762463	-1.556135021	0.031208552	3.545293	0.0597146	GEE	1	-0.5925	-1.389885965	0.204878173	2.121024	0.1452893	GEE
mdrg_514	Other vascular procedures	1	-0.909611	-1.280259463	-0.538962304	23.13571	1.51E-06	GEE	1	-0.62221	-1.001114696	-0.243300154	10.35859	0.0012887	GEE
mdrg_602	Major small & large bowel proce	1	-0.604709	-0.932234639	-0.277182456	13.09473	0.0002961	GEE	1	-0.73717	-1.067251079	-0.407092879	19.1601	1.202E-05	GEE
mdrg_701	Pancreas liver & shunt procedures	1	-0.937022	-1.945047648	0.07100277	3.319353	0.0684688	GEE	1	-1.08105	-2.089949193	-0.072159291	4.410616	0.0357159	GEE
mdrg_803	Spinal fusion except cervical	1	-0.869746	-1.684561866	-0.054930274	4.376856	0.0364301	GEE	1	-1.07014	-1.886167738	-0.25412103	6.606566	0.0101603	GEE
mdrg_806	Revision of hip or knee replacement	1	-1.438044	-2.586713385	-0.289374741	6.020749	0.0141386	GEE	1	-1.51111	-2.659960116	-0.362263338	6.646054	0.0099376	GEE
mdrg_807	Major joint replacement or reattachment	1	-1.290399	-1.661738905	-0.919058837	46.38746	1E-11	GEE	1	-1.4898	-1.862704559	-1.116895475	61.3136	0	GEE
mdrg_816	Lower extrem & humer proc	1	-0.93356	-1.824476013	-0.042644152	4.218007	0.0399971	GEE	1	-0.98051	-1.872473989	-0.088544525	4.64199	0.0311988	GEE
TPPS04A_ANY	ANY triggering complications were POA								1	-0.51781	-0.6594455	-0.376180367	51.34708	1E-12	GEE
TPPS04A_SEVERE	ANY of triggering complications were SEVERE								1	0.697766	0.556356214	0.839176596	93.53085	0	GEE

Table 2. PSI 04B Risk Model: Pneumonia Stratum

		Current Software (v60Rev2)							Modification 1						
		C Stat = .771							C Stat = 0.782						
Variable	Description	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model
Intercept	Intercept	1	-3.619139	-3.815702504	-3.422575347	1302.269	0	GEE	1	-3.796784	-3.995713751	-3.597854181	1399.356	0	GEE
M_AgeCat_1	Male | Age < 30	1	-0.653407	-1.00976066	-0.297054151	12.91525	0.0003259	GEE	1	-0.673636	-1.031337188	-0.315935076	13.62406	0.0002233	GEE
M_AgeCat_2	Male | Age < 35	1	0.118139	-0.262630093	0.498908153	0.369794	0.5431167	GEE	1	0.160177	-0.223767188	0.544121445	0.668591	0.4135432	GEE
M_AgeCat_3	Male | Age < 40	1	0.058311	-0.293807329	0.410428902	0.105346	0.7455064	GEE	1	0.020696	-0.335032729	0.376423753	0.013002	0.9092168	GEE
M_AgeCat_4	Male | Age < 45	1	-0.125891	-0.422642411	0.170860559	0.691352	0.4057049	GEE	1	-0.108738	-0.406664496	0.189188753	0.511728	0.4743918	GEE
M_AgeCat_5	Male | Age < 50	1	-0.044679	-0.279406031	0.190047659	0.139181	0.7090962	GEE	1	-0.063541	-0.299068741	0.171986932	0.279588	0.5969714	GEE
M_AgeCat_6	Male | Age < 55	1	-0.101844	-0.299800205	0.096111469	1.016795	0.3132803	GEE	1	-0.092604	-0.291323725	0.106116703	0.834194	0.3610626	GEE
M_AgeCat_7	Male | Age < 60	1	-0.206487	-0.380272166	-0.032702799	5.423263	0.0198702	GEE	1	-0.19856	-0.373196172	-0.023924269	4.966076	0.0258492	GEE
M_AgeCat_8	Male | Age < 65	1	0.027315	-0.149301248	0.203931413	0.091884	0.7617957	GEE	1	-0.014762	-0.192615492	0.163091459	0.026464	0.8707712	GEE
M_AgeCat_9	Male | Age >=65	1	0.035183	-0.138388009	0.2087538	0.157836	0.6911569	GEE	1	0.000125	-0.17454116	0.174790925	1.96E-06	0.9988819	GEE
M_AgeCat_10	Male | Age >=70	1	0.176656	0.025526052	0.32778677	5.248698	0.0219632	GEE	1	0.182946	0.031099551	0.334793001	5.576125	0.018207	GEE
M_AgeCat_11	Male | Age >=75	1	0.109076	-0.036870319	0.255021891	2.145694	0.1429703	GEE	1	0.101128	-0.04552358	0.247780364	1.826697	0.1765182	GEE
M_AgeCat_12	Male | Age >=80	1	0.19842	0.054877031	0.341963663	7.340111	0.0067433	GEE	1	0.185501	0.041115389	0.329886968	6.34075	0.0117995	GEE
M_AgeCat_13	Male | Age >=85	1	0.236833	0.087493771	0.386172202	9.661239	0.001882	GEE	1	0.223664	0.073492746	0.373834459	8.521489	0.0035098	GEE
F_AgeCat_1	Female | Age < 30	1	0.225588	-0.357158204	0.808333891	0.575664	0.448017	GEE	1	0.209532	-0.375531777	0.794595578	0.492709	0.4827215	GEE
F_AgeCat_2	Female | Age < 35	1	-0.075817	-0.747165093	0.595530985	0.048993	0.8248248	GEE	1	-0.061806	-0.73438212	0.610769484	0.03244	0.8570656	GEE
F_AgeCat_3	Female | Age < 40	1	-0.466474	-1.022908623	0.089959741	2.699755	0.1003637	GEE	1	-0.465971	-1.023163757	0.091220941	2.686609	0.101195	GEE
F_AgeCat_4	Female | Age < 45	1	-0.170673	-0.571919685	0.230573219	0.695032	0.4044582	GEE	1	-0.165769	-0.568343256	0.236805789	0.651342	0.4196334	GEE
F_AgeCat_5	Female | Age < 50	1	-0.012104	-0.328429142	0.304221149	0.005625	0.9402172	GEE	1	-0.009426	-0.326685882	0.307833348	0.003391	0.9535626	GEE
F_AgeCat_6	Female | Age < 55	1	-0.246965	-0.507836685	0.013906059	3.442827	0.0635269	GEE	1	-0.247024	-0.508969348	0.014922225	3.416254	0.0645574	GEE
F_AgeCat_7	Female | Age < 60	1	-0.046844	-0.272798946	0.179110973	0.165105	0.6844993	GEE	1	-0.041505	-0.26836227	0.185351821	0.128587	0.7199016	GEE
F_AgeCat_8	Female | Age < 65	1	-0.15849	-0.362185601	0.045205089	2.325622	0.1272594	GEE	1	-0.156739	-0.361443491	0.047965478	2.252141	0.1334296	GEE
F_AgeCat_10	Female | Age >=70	1	0.059057	-0.126034463	0.244148502	0.391081	0.5317319	GEE	1	0.068673	-0.117373751	0.254720023	0.52339	0.4693993	GEE
F_AgeCat_11	Female | Age >=75	1	0.27119	0.098457568	0.443922187	9.468822	0.0020899	GEE	1	0.267785	0.094371356	0.44119796	9.160153	0.0024734	GEE
F_AgeCat_12	Female | Age >=80	1	0.107594	-0.051967868	0.267156849	1.746688	0.1862936	GEE	1	0.096398	-0.063898117	0.25669399	1.389268	0.2385288	GEE
F_AgeCat_13	Female | Age >=85	1	0.299016	0.145649165	0.452382286	14.60235	0.0001327	GEE	1	0.298165	0.144098188	0.452231636	14.38771	0.0001488	GEE
TRNSFER	Transfer	1	0.178395	0.094374219	0.262415038	17.3177	3.162E-05	GEE	1	0.187826	0.103438821	0.272212802	19.03074	1.286E-05	GEE
ALCOHOL	Alcohol abuse	1	-0.158132	-0.280686318	-0.035577581	6.395549	0.0114407	GEE	1	-0.204239	-0.327317387	-0.081159731	10.57803	0.0011444	GEE
ANEMDEF	Deficiency Anemias	1	-0.23985	-0.315655255	-0.16404521	38.45745	5.6E-10	GEE	1	-0.224483	-0.30067166	-0.148293824	33.34867	7.703E-09	GEE
BLDLOSS	Chronic blood loss anemia	1	-0.354551	-0.63043751	-0.078665259	6.344462	0.0117748	GEE	1	-0.34224	-0.619328961	-0.065151964	5.860328	0.0154859	GEE
CHF	Congestive heart failure	1	0.251489	0.166878547	0.336099881	33.93788	5.69E-09	GEE	1	0.242564	0.157503156	0.327624935	31.23843	2.282E-08	GEE
CHRNLUNG	Chronic pulmonary disease	1	-0.089826	-0.156264059	-0.023387284	7.02197	0.0080516	GEE	1	-0.071803	-0.138731087	-0.004874328	4.421374	0.0354915	GEE
COAG	Coagulopthy	1	0.531704	0.437603591	0.625804141	122.6462	0	GEE	1	0.512776	0.418131218	0.607420257	112.7613	0	GEE
DEPRESS	Depression	1	-0.180529	-0.284370414	-0.076686598	11.61027	0.0006559	GEE	1	-0.188929	-0.293221817	-0.084635482	12.60608	0.0003845	GEE
DMCX	Diabetes w/ chronic complications	1	-0.245449	-0.383799943	-0.10709888	12.0909	0.0005067	GEE	1	-0.222074	-0.360988945	-0.083159614	9.817401	0.0017287	GEE
HTN_C	HTN_C	1	-0.324442	-0.386580976	-0.262302506	104.722	0	GEE	1	-0.310368	-0.372854537	-0.247881324	94.77111	0	GEE
HYPOTHY	Hypothyroidism	1	-0.178882	-0.271028009	-0.086735673	14.47685	0.0001419	GEE	1	-0.181269	-0.273755313	-0.088782104	14.75653	0.0001223	GEE
LIVER	Liver disease	1	0.431868	0.285002893	0.578733388	33.217	8.243E-09	GEE	1	0.451789	0.304217829	0.599360921	36.00501	1.968E-09	GEE
MDC_24	MDC 24: Multiple Significant Trauma	1	1.460484	1.195011195	1.725956736	116.2652	0	GEE	1	1.326494	1.056312724	1.59667623	92.5966	0	GEE
MDC_5	MDC 5: Circulatory System	1	1.395215	1.230330906	1.56009828	275.0571	0	GEE	1	1.342854	1.177300823	1.508407009	252.743	0	GEE
MDC_6	MDC 6: Digestive System	1	0.929168	0.655703008	1.202633082	44.34872	2.7E-11	GEE	1	0.836563	0.562095469	1.111030149	35.68716	2.317E-09	GEE
MDC_7	MDC 7: Hepatobiliary System And Pancreas	1	0.651056	0.319978326	0.9821336	14.85502	0.0001161	GEE	1	0.656581	0.324617094	0.988544544	15.02766	0.0001059	GEE
MDC_18	MDC 18: Infectious and Parasitic	1	1.460015	1.305990073	1.614039658	345.1672	0	GEE	1	1.431132	1.275381283	1.586883363	324.335	0	GEE
MDC_21	MDC 21: Injuries Poison And Toxic	1	0.424699	0.080380863	0.769016379	5.844388	0.0156268	GEE	1	0.362692	0.017569466	0.70781399	4.242536	0.0394232	GEE
LYTES	Fluid and electrolyte disorders	1	0.462466	0.399015595	0.525915909	204.0751	0	GEE	1	0.451989	0.387795905	0.516182628	190.4461	0	GEE
MDC_1	MDC 1: Nervous System	1	0.978738	0.779245819	1.178229358	92.46529	0	GEE	1	0.800859	0.599782453	1.001934843	60.93778	0	GEE
MDC_8	MDC 8: Musculoskeletal And Connective	1	0.403258	0.231376895	0.575139064	21.14492	4.258E-06	GEE	1	0.381529	0.209355652	0.553702569	18.86336	1.404E-05	GEE
METS	Metastatic cancer	1	0.530914	0.39478423	0.66704455	58.43008	0	GEE	1	0.570123	0.433208677	0.707037566	66.60931	0	GEE
NEURO	Neurological disease	1	0.262856	0.169463533	0.356249073	30.43032	3.461E-08	GEE	1	0.207215	0.113164614	0.301266263	18.64727	1.573E-05	GEE
OBESE	Obesity	1	-0.215551	-0.309566005	-0.121535471	20.19289	7.001E-06	GEE	1	-0.211031	-0.305451057	-0.116610662	19.18926	1.184E-05	GEE
PARA	Paralysis	1	-0.281142	-0.396179318	-0.166105088	22.94417	1.668E-06	GEE	1	-0.331565	-0.446907278	-0.216222729	31.74363	1.759E-08	GEE
PERIVASC	Peripheral vascular disease	1	0.202123	0.117001229	0.287244579	21.65942	3.256E-06	GEE	1	0.206033	0.120473451	0.291592549	22.27577	2.362E-06	GEE
PSYCH	Psychoses	1	-0.322709	-0.488842308	-0.156574835	14.49444	0.0001406	GEE	1	-0.334395	-0.501073708	-0.167716781	15.46171	8.419E-05	GEE
PULMCIRC	Pulmonary circulation disease	1	0.211973	0.07688234	0.347063543	9.458162	0.0021021	GEE	1	0.199482	0.063711519	0.335251773	8.292655	0.0039806	GEE
RENLFAIL	Renal failure	1	0.192228	0.110389664	0.274066624	21.1941	4.15E-06	GEE	1	0.204224	0.121951488	0.286496189	23.67016	1.143E-06	GEE

TUMOR	Solid tumor w/out metastasis	1	0.378217	0.227238272	0.529195652	24.10721	9.112E-07	GEE	1	0.406827	0.254877418	0.558777131	27.53694	1.541E-07	GEE
WGHTLOSS	Weight loss	1	0.155636	0.059766004	0.251506297	10.12396	0.0014635	GEE	1	0.125505	0.029314544	0.221694742	6.539658	0.0105496	GEE
mdrg_101	Intracranial vascular procedures	1	1.362385	1.129989906	1.594780945	132.0203	0	GEE	1	1.360221	1.126418772	1.594023256	130.0223	0	GEE
mdrg_102	Cranio w major dev impl/acute complex CNS	1	1.931539	1.753411845	2.109665718	451.6935	0	GEE	1	1.919738	1.74029652	2.099180055	439.6764	0	GEE
mdrg_103	Craniotomy & endovascular intracranial procedures	1	1.071057	0.886460708	1.255653296	129.3229	0	GEE	1	1.080617	0.894753262	1.26648125	129.8522	0	GEE
mdrg_107	Extracranial procedures	1	-0.518564	-0.994401429	-0.042726783	4.562303	0.0326831	GEE	1	-0.445434	-0.922613498	0.03174564	3.347334	0.0673139	GEE
mdrg_1802	Postoperative or post-traumatic infections	1	-0.664615	-1.043942926	-0.285287074	11.79253	0.0005947	GEE	1	-0.657048	-1.038308063	-0.275787732	11.40899	0.0007309	GEE
mdrg_2406	Craniotomy for multiple significant trauma	1	2.093755	1.766518836	2.420990765	157.2627	0	GEE	1	2.022114	1.689638079	2.354590714	142.0974	0	GEE
mdrg_2407	Limb reattachment hip & femur proc	1	-0.547866	-0.976673493	-0.119057968	6.270746	0.0122747	GEE	1	-0.473888	-0.906739229	-0.041036277	4.604361	0.0318907	GEE
mdrg_503	Cardiac valve & oth maj cardiothoracic proc	1	-0.193612	-0.344865256	-0.042358435	6.294332	0.0121125	GEE	1	-0.160507	-0.313341509	-0.007673014	4.236861	0.0395552	GEE
mdrg_504	Cardiac defibrillator implant	1	-1.486684	-1.952871933	-1.020495319	39.06704	4.09E-10	GEE	1	-1.531045	-1.997776257	-1.064313229	41.33688	1.28E-10	GEE
mdrg_507	Coronary bypass w cardiac cath	1	-1.193725	-1.434939287	-0.952510476	94.08019	0	GEE	1	-1.126696	-1.368919773	-0.884472341	83.11443	0	GEE
mdrg_510	Permanent cardiac pacemaker implant	1	-1.160269	-1.689028753	-0.631509769	18.49683	1.702E-05	GEE	1	-1.09456	-1.624119971	-0.56500064	16.41141	5.098E-05	GEE
mdrg_511	Perc cardiovasc proc w drug-eluting stent	1	-0.469223	-0.620414601	-0.318030428	36.99949	1.182E-09	GEE	1	-0.400138	-0.552675188	-0.247601011	26.43413	2.727E-07	GEE
mdrg_514	Other vascular procedures	1	-0.417536	-0.66570603	-0.169365825	10.87389	0.0009753	GEE	1	-0.360881	-0.610161101	-0.111600982	8.050989	0.0045479	GEE
mdrg_519	Other circulatory system O.R. procedures	1	-0.605744	-1.101290082	-0.110198658	5.739942	0.0165833	GEE	1	-0.49158	-0.988504656	0.005345092	3.759263	0.0525157	GEE
mdrg_601	Stomach esophageal & duodenal	1	-0.537577	-0.853458166	-0.221696135	11.12579	0.0008514	GEE	1	-0.511684	-0.828725012	-0.194642741	10.00618	0.0015602	GEE
mdrg_602	Major small & large bowel proce	1	-0.523683	-0.794484054	-0.252882616	14.36594	0.0001505	GEE	1	-0.447199	-0.719405445	-0.174993249	10.36819	0.0012821	GEE
mdrg_603	Rectal resection w MCC	1	-1.181119	-2.028357421	-0.333880007	7.46572	0.0062885	GEE	1	-1.085171	-1.933943478	-0.236397531	6.279269	0.0122159	GEE
mdrg_604	Peritoneal adhesiolysis w MCC	1	-0.982971	-1.384660775	-0.581280742	23.00357	1.617E-06	GEE	1	-0.935292	-1.338263458	-0.532320196	20.69386	5.389E-06	GEE
mdrg_605	Appendectomy w complicated principal diag	1	-1.201667	-1.904802015	-0.498532616	11.21986	0.0008093	GEE	1	-1.063218	-1.767546355	-0.358889209	8.753667	0.0030898	GEE
mdrg_606	Appendectomy w/o complicated principal diag	1	-1.767613	-2.931481794	-0.603744517	8.860591	0.0029139	GEE	1	-1.615769	-2.780838236	-0.450699312	7.38841	0.0065646	GEE
mdrg_607	Minor small & large bowel procedures	1	-1.198335	-2.219548154	-0.1771219	5.289564	0.0214536	GEE	1	-1.116078	-2.13975727	-0.092399475	4.566233	0.0326082	GEE
mdrg_608	Anal & stomal procedures	1	-1.210192	-2.23207225	-0.188311103	5.38771	0.0202791	GEE	1	-1.093521	-2.116333916	-0.070707286	4.390939	0.0361304	GEE
mdrg_609	Inguinal & femoral hernia procedures	1	-1.344465	-2.050734861	-0.638195882	13.92049	0.0001907	GEE	1	-1.276837	-1.984413665	-0.569260993	12.50895	0.000405	GEE
mdrg_610	Hernia procedures except inguinal & femoral	1	-1.62454	-2.266616693	-0.982462548	24.59136	7.087E-07	GEE	1	-1.553662	-2.196827389	-0.910496537	22.4163	2.195E-06	GEE
mdrg_701	Pancreas liver & shunt procedures	1	-0.776802	-1.348272331	-0.205331275	7.097885	0.0077175	GEE	1	-0.799455	-1.37256197	-0.226347802	7.475023	0.0062561	GEE
mdrg_705	Laparoscopic cholecystectomy	1	-1.187563	-1.650107493	-0.725019186	25.32231	4.851E-07	GEE	1	-1.135088	-1.598614679	-0.671560613	23.03588	1.59E-06	GEE
mdrg_801	Combined anterior/posterior spinal fusion	1	-1.272059	-2.263708625	-0.280408718	6.321115	0.0119309	GEE	1	-1.353701	-2.34634492	-0.361057307	7.144221	0.0075206	GEE
mdrg_803	Spinal fusion except cervical	1	-1.533449	-2.19822238	-0.868674924	20.44029	6.152E-06	GEE	1	-1.488482	-2.153634858	-0.823328299	19.2371	1.154E-05	GEE
mdrg_807	Major joint replacement or reattachment	1	-0.340605	-0.515551532	-0.16565897	14.56094	0.0001357	GEE	1	-0.327881	-0.503296023	-0.152465088	13.42121	0.0002488	GEE
mdrg_808	Cervical spinal fusion w MCC	1	-0.414011	-0.922392765	0.094369843	2.547662	0.1104576	GEE	1	-0.499261	-1.008730815	0.010209675	3.689037	0.0547712	GEE
mdrg_815	Back & neck proc exc spinal fusion	1	-0.766959	-1.436489083	-0.097428784	5.040808	0.0247569	GEE	1	-0.773812	-1.444666411	-0.102958561	5.111069	0.0237737	GEE
mdrg_816	Lower extrem & humer proc	1	-0.395008	-0.773131417	-0.016884103	4.192178	0.0406109	GEE	1	-0.391006	-0.769870658	-0.012142144	4.091633	0.043096	GEE
TPPS04B_ANY	ANY triggering complications were POA								1	-0.011666	-0.076113899	0.052781216	0.125879	0.7227439	GEE
TPPS04B_SEVERE	ANY of triggering complications were SEVERE								1	0.617668	0.558219933	0.677115175	414.7026	0	GEE

Table 3. PSI 04C Risk Model: Sepsis Stratum

		Current Software (v60rev2)							Modification 1						
		C Stat = .726							C Stat = .776						
Variable	Description	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model
Intercept	Intercept	1	-2.035927	-2.198589515	-1.87326387	601.7884	0	GEE	1	-2.746307	-2.921833613	-2.570780615	940.3922	0	GEE
M_AgeCat_1	Male | Age < 30	1	0.211794	-0.311455049	0.735043418	0.629371	0.427586	GEE	1	0.183208	-0.353638415	0.72005517	0.44739	0.5035771	GEE
M_AgeCat_2	Male | Age < 35	1	-0.376697	-0.971094968	0.217700935	1.542858	0.2141929	GEE	1	-0.382274	-0.9902011	0.225652229	1.518953	0.217778	GEE
M_AgeCat_3	Male | Age < 40	1	-0.313073	-0.791325441	0.165179184	1.646166	0.1994816	GEE	1	-0.283524	-0.772666842	0.205618752	1.290639	0.2559303	GEE
M_AgeCat_4	Male | Age < 45	1	-0.145247	-0.490616108	0.20012308	0.679422	0.4097859	GEE	1	-0.064809	-0.420039568	0.290420937	0.127865	0.7206562	GEE
M_AgeCat_5	Male | Age < 50	1	-0.121926	-0.389988335	0.14613705	0.794719	0.372677	GEE	1	-0.117071	-0.393642308	0.159499677	0.68831	0.4067399	GEE
M_AgeCat_6	Male | Age < 55	1	-0.134897	-0.352857405	0.08306381	1.471443	0.2251182	GEE	1	-0.104575	-0.32990789	0.120758838	0.827365	0.3630357	GEE
M_AgeCat_7	Male | Age < 60	1	-0.106795	-0.296104768	0.082513929	1.222524	0.2688659	GEE	1	-0.103329	-0.298736279	0.09207811	1.074138	0.3000125	GEE
M_AgeCat_8	Male | Age < 65	1	-0.15996	-0.342880425	0.022960117	2.937624	0.086538	GEE	1	-0.104108	-0.29246953	0.084253655	1.173489	0.2786854	GEE
M_AgeCat_9	Male | Age >=65	1	0.028397	-0.146900089	0.203693349	0.100805	0.7508657	GEE	1	0.055373	-0.125397777	0.236143209	0.36044	0.5482621	GEE
M_AgeCat_10	Male | Age >=70	1	0.15099	-0.006921245	0.308901499	3.512097	0.0609223	GEE	1	0.153962	-0.009277322	0.317201053	3.417226	0.0645194	GEE
M_AgeCat_11	Male | Age >=75	1	0.095582	-0.063919936	0.255083387	1.37948	0.2401897	GEE	1	0.110373	-0.054183049	0.274928093	1.728196	0.1886415	GEE
M_AgeCat_12	Male | Age >=80	1	0.216736	0.049372466	0.384100526	6.442227	0.0111439	GEE	1	0.200541	0.02785251	0.373228818	5.180551	0.0228411	GEE
M_AgeCat_13	Male | Age >=85	1	0.069821	-0.123190182	0.26283318	0.502698	0.4783169	GEE	1	0.106735	-0.093138481	0.306607674	1.095464	0.2952638	GEE
F_AgeCat_1	Female | Age < 30	1	-0.29615	-0.976263886	0.383964384	0.728375	0.3934103	GEE	1	-0.238735	-0.945616724	0.468146642	0.438162	0.5080107	GEE
F_AgeCat_2	Female | Age < 35	1	-0.20904	-0.876794051	0.458714291	0.376462	0.5395028	GEE	1	-0.207041	-0.903997284	0.48991445	0.339	0.5604068	GEE
F_AgeCat_3	Female | Age < 40	1	0.050183	-0.505533558	0.605900128	0.031326	0.8595147	GEE	1	0.034478	-0.534661098	0.603616792	0.014097	0.9054872	GEE
F_AgeCat_4	Female | Age < 45	1	-0.166236	-0.63119054	0.298718738	0.491049	0.4834598	GEE	1	-0.127442	-0.602597405	0.347713852	0.276343	0.5991084	GEE
F_AgeCat_5	Female | Age < 50	1	-0.157475	-0.528978278	0.214028454	0.690229	0.4060864	GEE	1	-0.141502	-0.521005746	0.23800085	0.534064	0.4649034	GEE
F_AgeCat_6	Female | Age < 55	1	-0.404646	-0.689667371	-0.119623722	7.742647	0.0053932	GEE	1	-0.345959	-0.63807384	-0.053844548	5.388145	0.020274	GEE
F_AgeCat_7	Female | Age < 60	1	-0.116019	-0.347664725	0.11562586	0.963631	0.3262738	GEE	1	-0.10064	-0.338422404	0.137143252	0.688134	0.4067999	GEE
F_AgeCat_8	Female | Age < 65	1	-0.225065	-0.431219818	-0.01890987	4.578507	0.0323754	GEE	1	-0.191312	-0.403339794	0.020716234	3.12747	0.0769831	GEE
F_AgeCat_10	Female | Age >=70	1	-0.015623	-0.205980631	0.174733874	0.025877	0.872202	GEE	1	0.023548	-0.172645109	0.219741078	0.05534	0.8140195	GEE
F_AgeCat_11	Female | Age >=75	1	0.399913	0.219445647	0.580380337	18.86383	1.404E-05	GEE	1	0.380821	0.194376545	0.567264724	16.02654	6.246E-05	GEE
F_AgeCat_12	Female | Age >=80	1	-0.018419	-0.195520982	0.158683147	0.041551	0.8384788	GEE	1	0.001155	-0.181979888	0.184290232	0.000153	0.990136	GEE
F_AgeCat_13	Female | Age >=85	1	0.154887	-0.042874694	0.35264782	2.356358	0.1247732	GEE	1	0.209587	0.004400418	0.414773749	4.007993	0.045285	GEE
TRNSFER	Transfer	1	0.318417	0.217880966	0.418953342	38.53402	5.38E-10	GEE	1	0.364613	0.260061978	0.469163729	46.72021	8E-12	GEE
ANEMDEF	Deficiency Anemias	1	-0.181983	-0.271105892	-0.092859519	16.01677	6.278E-05	GEE	1	-0.085143	-0.177431541	0.007145575	3.269627	0.0705737	GEE
BLDLOSS	Chronic blood loss anemia	1	-0.632978	-0.970339027	-0.295616506	13.5233	0.0002356	GEE	1	-0.506315	-0.849849739	-0.162780111	8.34442	0.0038687	GEE
CHF	Congestive heart failure	1	0.161369	0.063000619	0.259737251	10.33774	0.0013034	GEE	1	0.104151	0.002953557	0.205347555	4.068966	0.043678	GEE
CHRNLUNG	Chronic pulmonary disease	1	0.104327	0.026736696	0.181916944	6.945048	0.0084052	GEE	1	0.092933	0.012788685	0.173076593	5.165248	0.0230432	GEE
COAG	Coagulopthy	1	0.442321	0.330105151	0.55453745	59.68445	0	GEE	1	0.461413	0.344791968	0.578033639	60.13443	0	GEE
DEPRESS	Depression	1	-0.254199	-0.381327948	-0.127070632	15.35882	8.89E-05	GEE	1	-0.155573	-0.286709167	-0.024436102	5.406488	0.020062	GEE
DM	Diabetes w/o chronic complications	1	-0.116695	-0.200516842	-0.032872528	7.445274	0.0063603	GEE	1	-0.034566	-0.121003677	0.051870977	0.614328	0.4331627	GEE
DMCX	Diabetes w/ chronic complications	1	-0.31215	-0.468590089	-0.155710738	15.29432	9.199E-05	GEE	1	-0.177697	-0.339738375	-0.015655661	4.619604	0.0316085	GEE
HTN_C	HTN_C	1	-0.23896	-0.307701456	-0.170219378	46.42115	1E-11	GEE	1	-0.186132	-0.25728141	-0.11498354	26.29079	2.937E-07	GEE
LIVER	Liver disease	1	0.779	0.639485442	0.918515029	119.7653	0	GEE	1	0.719624	0.574818392	0.864430226	94.87138	0	GEE
LYTES	Fluid and electrolyte disorders	1	0.198495	0.124931939	0.272058919	27.96872	1.233E-07	GEE	1	0.287385	0.209691825	0.365078264	52.56035	0	GEE
MDC_4	MDC 4: Respiratory System	1	1.05803	0.88577158	1.230289047	144.9205	0	GEE	1	0.991516	0.812934351	1.170096933	118.4195	0	GEE
MDC_5	MDC 5: Circulatory System	1	0.991357	0.869608889	1.113105426	254.7011	0	GEE	1	0.877698	0.752033075	1.003362854	187.3952	0	GEE
MDC_6	MDC 6: Digestive System	1	0.453829	0.330423235	0.577234759	51.9529	1E-12	GEE	1	0.357138	0.230122866	0.484152848	30.37092	3.568E-08	GEE
MDC_7	MDC 7: Hepatobiliary System And Pancreas	1	1.298304	0.986955246	1.609652638	66.79667	0	GEE	1	1.148743	0.824156956	1.473328122	48.11529	4E-12	GEE
MDC_22	MDC 22: Burns	1	1.979577	1.567809936	2.391343484	88.78473	0	GEE	1	1.80504	1.376477334	2.233602322	68.14617	0	GEE
MDC_24	MDC 24: Multiple Significant Trauma	1	2.655623	2.076939879	3.234306889	80.89971	0	GEE	1	2.78388	2.185598879	3.382161916	83.17373	0	GEE
MDC_1	MDC 1: Nervous System	1	0.310091	0.115060625	0.505121553	9.711132	0.0018316	GEE	1	0.303677	0.102685972	0.504667254	8.769346	0.0030634	GEE
MDC_17	MDC 17: Myeloproliferative DDs	1	0.856401	0.561961194	1.150840983	32.49805	1.193E-08	GEE	1	0.916762	0.608663168	1.224860027	34.01183	5.478E-09	GEE
METS	Metastatic cancer	1	0.517485	0.399389541	0.635579598	73.76116	0	GEE	1	0.564877	0.442803741	0.686950869	82.25479	0	GEE
PARA	Paralysis	1	-0.24843	-0.415937834	-0.080921466	8.449499	0.0036514	GEE	1	-0.222339	-0.394563336	-0.050114311	6.402312	0.0113972	GEE
PERIVASC	Peripheral vascular disease	1	0.132564	0.032396815	0.232731065	6.728164	0.0094902	GEE	1	0.108363	0.004773977	0.211951987	4.203691	0.0403361	GEE
PSYCH	Psychoses	1	-0.270878	-0.461591612	-0.080164182	7.749603	0.0053724	GEE	1	-0.238279	-0.433584426	-0.042974488	5.717982	0.016792	GEE
PULMCIRC	Pulmonary circulation disease	1	0.463084	0.311672638	0.614494537	35.93357	2.042E-09	GEE	1	0.4414	0.285121885	0.597678205	30.64533	3.098E-08	GEE
RENLFAIL	Renal failure	1	0.207919	0.116633722	0.299204679	19.92884	8.038E-06	GEE	1	0.133352	0.039191452	0.227512465	7.70473	0.0055076	GEE
TUMOR	Solid tumor w/out metastasis	1	0.208085	0.032177603	0.383992625	5.375383	0.0204229	GEE	1	0.241379	0.05945404	0.423303321	6.762557	0.009309	GEE
mdrg_1002	Amputat of lower limb for endocrine	1	-1.365165	-2.185348093	-0.544982769	10.64255	0.0011052	GEE	1	-0.940644	-1.769877828	-0.111410871	4.943032	0.026196	GEE
mdrg_101	Intracranial vascular procedures	1	0.746079	0.3815561	1.110601148	16.09225	6.033E-05	GEE	1	0.752328	0.376894501	1.127761417	15.42569	8.581E-05	GEE

mdrg_102	Cranio w major dev impl/acute complex CNS	1	1.262516	0.985964612	1.539067955	80.06037	0	GEE	1	1.388972	1.102566688	1.675377563	90.34845	0	GEE
mdrg_103	Craniotomy & endovascular intracranial procedures	1	0.794051	0.523838409	1.064263935	33.17275	8.432E-09	GEE	1	0.826159	0.54530517	1.10701239	33.24017	8.145E-09	GEE
mdrg_1104	Kidney & ureter procedures for non-neoplasm	1	-1.157874	-1.588675692	-0.727073076	27.75013	1.38E-07	GEE	1	-1.052566	-1.48911751	-0.616013882	22.33176	2.294E-06	GEE
mdrg_1107	Transurethral procedures	1	-1.053989	-1.583820761	-0.524157588	15.20172	9.662E-05	GEE	1	-0.993865	-1.531638516	-0.456091963	13.12057	0.0002921	GEE
mdrg_1109	Other kidney & urinary tract procedures	1	0.77604	0.136621715	1.415457377	5.65841	0.0173719	GEE	1	0.882162	0.215655354	1.548667723	6.729519	0.009483	GEE
mdrg_2104	Other O.R. procedures for injuries	1	-0.161406	-0.432632495	0.109820286	1.360417	0.2434651	GEE	1	-0.05452	-0.334289346	0.225249176	0.145885	0.7024997	GEE
mdrg_2407	Limb reattachment hip & femur proc	1	-1.873364	-2.572409077	-1.174319849	27.58868	1.501E-07	GEE	1	-2.00243	-2.724697608	-1.280161812	29.52665	5.515E-08	GEE
mdrg_2408	Other O.R. procedures for multiple sig tr	1	-1.657191	-2.281691645	-1.032691265	27.05057	1.982E-07	GEE	1	-1.879107	-2.525354966	-1.232859471	32.47889	1.205E-08	GEE
mdrg_401	Major chest procedures	1	0.380317	0.143669206	0.616965013	9.92162	0.0016335	GEE	1	0.34971	0.10314506	0.59627474	7.727691	0.005438	GEE
mdrg_503	Cardiac valve & oth maj cardiothoracic proc	1	0.497144	0.338905766	0.655381577	37.91744	7.38E-10	GEE	1	0.463	0.298962456	0.62703765	30.60353	3.165E-08	GEE
mdrg_504	Cardiac defibrillator implant	1	-1.408579	-2.070405262	-0.746752699	17.40085	3.027E-05	GEE	1	-1.36493	-2.039890041	-0.689968962	15.70941	7.386E-05	GEE
mdrg_509	Amputation for circ sys disorders	1	-1.170614	-1.487203946	-0.854024209	52.5206	0	GEE	1	-0.883096	-1.209744116	-0.556447766	28.07705	1.166E-07	GEE
mdrg_511	Perc cardiovasc proc w drug-eluting stent	1	-0.443206	-0.683656918	-0.202754324	13.05126	0.0003031	GEE	1	-0.305	-0.552840622	-0.057159624	5.817705	0.0158656	GEE
mdrg_514	Other vascular procedures	1	-0.401685	-0.661534332	-0.141836274	9.179645	0.0024472	GEE	1	-0.223384	-0.49429782	0.047529303	2.611799	0.1060713	GEE
mdrg_515	Upper limb & toe amputation	1	-2.230971	-3.648175851	-0.813766503	9.519605	0.0020329	GEE	1	-1.866815	-3.309160034	-0.424469432	6.435172	0.0111882	GEE
mdrg_602	Major small & large bowel proce	1	-0.165016	-0.288514639	-0.041516393	6.858324	0.0088229	GEE	1	-0.16479	-0.291728478	-0.037851201	6.47395	0.0109467	GEE
mdrg_603	Rectal resection w MCC	1	-0.986774	-1.684491836	-0.289056642	7.683736	0.0055721	GEE	1	-0.980142	-1.691918655	-0.268366009	7.284293	0.006956	GEE
mdrg_604	Peritoneal adhesiolysis w MCC	1	-0.460816	-0.743885425	-0.177746109	10.18038	0.0014194	GEE	1	-0.412969	-0.70360728	-0.122331129	7.755798	0.005354	GEE
mdrg_608	Anal & stomal procedures	1	-1.566953	-2.989461886	-0.144443468	4.6612	0.0308516	GEE	1	-1.500178	-2.939573908	-0.060781153	4.172734	0.0410795	GEE
mdrg_701	Pancreas liver & shunt procedures	1	-0.985019	-1.366186595	-0.603851454	25.65391	4.085E-07	GEE	1	-0.931603	-1.327578727	-0.535628247	21.26289	4.004E-06	GEE
mdrg_702	Biliary tract proc except only cholecyst	1	-1.268788	-2.200539174	-0.337036722	7.123187	0.0076093	GEE	1	-1.303967	-2.266776931	-0.341157157	7.046092	0.0079438	GEE
mdrg_704	Cholecystectomy except by laparoscope	1	-0.721941	-1.439201665	-0.004680157	3.891754	0.0485237	GEE	1	-0.492491	-1.235526358	0.250544132	1.687618	0.1939153	GEE
mdrg_705	Laparoscopic cholecystectomy	1	-1.432112	-2.073223826	-0.790999561	19.16823	1.197E-05	GEE	1	-1.2584	-1.918033486	-0.59876726	13.9807	0.0001847	GEE
mdrg_7701	Heart transplant or implant heart assist sys	1	0.782416	0.328400747	1.236430343	11.40857	0.0007311	GEE	1	0.781186	0.308405642	1.253966675	10.48784	0.0012016	GEE
mdrg_803	Spinal fusion except cervical	1	-0.438685	-0.892844379	0.01547478	3.584135	0.0583338	GEE	1	-0.38858	-0.85051962	0.073359063	2.71824	0.0992074	GEE
mdrg_8899	Non-Extensive O.R. Proc Unrelated to PDX	1	-0.156754	-0.345896261	0.032388198	2.638497	0.1043022	GEE	1	-0.031538	-0.227937579	0.16486221	0.099054	0.7529675	GEE
TPPS04C_ANY	ANY triggering complications were POA								1	-0.674529	-0.771642055	-0.577415459	185.3274	0	GEE
TPPS04C_SEVERE	ANY of triggering complications were SEVERE								1	1.2232	1.150852225	1.295546855	1098.11	0	GEE

Table 4. PSI 04D Risk Model: Shock/Cardiac Arrest Stratum

		Current Software (v60rev2)							Modification 1						
		C Stat = .715							C Stat = .818						
Variable	Description	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model
Intercept	Intercept	1	-1.130567	-1.269654825	-0.99148004	253.81284	0	GEE	1	-2.074965	-2.233830898	-1.916099229	655.325	0	GEE
M_AgeCat_1	Male | Age < 30	1	0.037829	-0.352640577	0.428298098	0.036055	0.8494018	GEE	1	0.249635	-0.178078995	0.677348132	1.308578	0.2526523	GEE
M_AgeCat_2	Male | Age < 35	1	0.255037	-0.165458981	0.675532736	1.413119	0.2345393	GEE	1	-0.017048	-0.474588549	0.440493151	0.005333	0.9417846	GEE
M_AgeCat_3	Male | Age < 40	1	-0.323801	-0.703202914	0.055601451	2.798029	0.0943803	GEE	1	-0.281734	-0.6995851	0.136116203	1.746361	0.1863349	GEE
M_AgeCat_4	Male | Age < 45	1	0.153034	-0.131581714	0.437650349	1.110593	0.291953	GEE	1	0.204223	-0.112823398	0.521268875	1.593894	0.2067707	GEE
M_AgeCat_5	Male | Age < 50	1	-0.145743	-0.370842303	0.079357025	1.610348	0.2044429	GEE	1	-0.218778	-0.467343865	0.029786958	2.975938	0.0845111	GEE
M_AgeCat_6	Male | Age < 55	1	-0.011699	-0.195682691	0.172284845	0.015532	0.9008183	GEE	1	-0.00803	-0.211331013	0.195271606	0.005993	0.9382959	GEE
M_AgeCat_7	Male | Age < 60	1	0.09771	-0.065916423	0.261336764	1.369834	0.2418404	GEE	1	0.069414	-0.111135507	0.24996406	0.567806	0.4511321	GEE
M_AgeCat_8	Male | Age < 65	1	0.021262	-0.137795203	0.180318712	0.068642	0.7933246	GEE	1	-0.04769	-0.222079925	0.126700357	0.287278	0.5919699	GEE
M_AgeCat_9	Male | Age >=65	1	0.147334	-0.007082125	0.301749501	3.49717	0.0614738	GEE	1	0.034049	-0.134866082	0.202964654	0.15609	0.6927826	GEE
M_AgeCat_10	Male | Age >=70	1	0.192072	0.041822157	0.34232144	6.277641	0.0122271	GEE	1	0.214824	0.049850585	0.379797456	6.513791	0.0107041	GEE
M_AgeCat_11	Male | Age >=75	1	0.056729	-0.098269496	0.21172773	0.51458	0.4731631	GEE	1	0.041824	-0.129198227	0.212846377	0.229744	0.631714	GEE
M_AgeCat_12	Male | Age >=80	1	0.280517	0.113551766	0.447481876	10.84333	0.0009915	GEE	1	0.33862	0.153557965	0.523683013	12.86132	0.0003354	GEE
M_AgeCat_13	Male | Age >=85	1	0.174822	-0.031065016	0.380708642	2.769684	0.0960651	GEE	1	0.200267	-0.027707256	0.428241128	2.964444	0.0851137	GEE
F_AgeCat_1	Female | Age < 30	1	0.149435	-0.17077762	0.469647371	0.836611	0.3603678	GEE	1	0.193497	-0.153899465	0.540893056	1.191773	0.2749723	GEE
F_AgeCat_2	Female | Age < 35	1	-0.129072	-0.468465474	0.210321636	0.555588	0.4560436	GEE	1	-0.148278	-0.517291813	0.220736249	0.620244	0.4309566	GEE
F_AgeCat_3	Female | Age < 40	1	-0.21632	-0.537224592	0.104584818	1.745569	0.1864348	GEE	1	-0.078569	-0.430645564	0.273507944	0.191303	0.6618343	GEE
F_AgeCat_4	Female | Age < 45	1	-0.048369	-0.340529754	0.243791517	0.10529	0.7455709	GEE	1	-0.005036	-0.326044236	0.315972861	0.000945	0.975472	GEE
F_AgeCat_5	Female | Age < 50	1	-0.089042	-0.330277325	0.152194003	0.523359	0.4694124	GEE	1	-0.159615	-0.424993213	0.105762443	1.389685	0.2384583	GEE
F_AgeCat_6	Female | Age < 55	1	-0.019916	-0.223037358	0.183204991	0.036932	0.8476044	GEE	1	-0.02786	-0.251807168	0.196086706	0.059453	0.8073622	GEE
F_AgeCat_7	Female | Age < 60	1	-0.153743	-0.33598788	0.028502402	2.733843	0.0982427	GEE	1	-0.16676	-0.367939283	0.034418493	2.639473	0.1042381	GEE
F_AgeCat_8	Female | Age < 65	1	0.050887	-0.115769826	0.217543622	0.358148	0.5495366	GEE	1	0.006556	-0.17696526	0.190077907	0.004903	0.9441778	GEE
F_AgeCat_10	Female | Age >=70	1	0.232622	0.069300465	0.395943818	7.793107	0.0052446	GEE	1	0.246165	0.066332848	0.425996757	7.198042	0.0072983	GEE
F_AgeCat_11	Female | Age >=75	1	0.049172	-0.119854613	0.218199137	0.325106	0.5685551	GEE	1	0.021611	-0.166015115	0.209237674	0.050965	0.8213931	GEE
F_AgeCat_12	Female | Age >=80	1	0.151184	-0.024021606	0.326389694	2.860302	0.0907905	GEE	1	0.162309	-0.032306348	0.356924026	2.671942	0.1021313	GEE
F_AgeCat_13	Female | Age >=85	1	0.298818	0.096677048	0.50095948	8.394614	0.0037633	GEE	1	0.280238	0.056048056	0.504427212	6.002308	0.0142872	GEE
TRNSFER	Transfer	1	0.145162	0.046438018	0.243886928	8.305302	0.003953	GEE	1	0.349486	0.240026422	0.458946043	39.16038	3.9E-10	GEE
BLDLOSS	Chronic blood loss anemia	1	-0.492731	-0.71101709	-0.274445527	19.57337	9.681E-06	GEE	1	-0.374235	-0.616412219	-0.132057228	9.173114	0.002456	GEE
CHF	Congestive heart failure	1	-0.110224	-0.187971694	-0.032476977	7.721112	0.0054579	GEE	1	-0.10334	-0.188796706	-0.017883941	5.617548	0.0177815	GEE
COAG	Coagulopthy	1	0.325468	0.236708347	0.414226846	51.65168	1E-12	GEE	1	0.550131	0.449908579	0.650353462	115.744	0	GEE
DEPRESS	Depression	1	-0.221914	-0.326371452	-0.117455587	17.33728	3.13E-05	GEE	1	-0.224298	-0.339196507	-0.109399613	14.63925	0.0001302	GEE
DM	Diabetes w/o chronic complications	1	0.070596	0.001326713	0.139864696	3.990019	0.0457705	GEE	1	0.042054	-0.034278361	0.118386377	1.165986	0.2802275	GEE
HYPOTHY	Hypothyroidism	1	-0.17137	-0.262312199	-0.080427976	13.64068	0.0002213	GEE	1	-0.133569	-0.233758384	-0.033379981	6.827581	0.0089761	GEE
IMMUNE	Immune disorders	1	0.18849	0.107701424	0.269277666	20.91104	4.811E-06	GEE	1	0.130313	0.040798961	0.219826375	8.141237	0.004327	GEE
LIVER	Liver disease	1	0.414937	0.301138711	0.528735674	51.0726	1E-12	GEE	1	0.760707	0.633878795	0.887534642	138.1977	0	GEE
LYTES	Fluid and electrolyte disorders	1	0.289672	0.224517702	0.354827288	75.9308	0	GEE	1	0.383662	0.309863812	0.457461182	103.824	0	GEE
METS	Metastatic cancer	1	0.355521	0.242117722	0.468924642	37.755	8.02E-10	GEE	1	0.462978	0.336661575	0.589293683	51.60587	1E-12	GEE
PARA	Paralysis	1	-0.135251	-0.282973393	0.012470908	3.220234	0.0727332	GEE	1	-0.234794	-0.398024686	-0.071563673	7.948189	0.0048136	GEE
PERIVASC	Peripheral vascular disease	1	0.343849	0.256605396	0.431092007	59.67147	0	GEE	1	0.307147	0.210558118	0.403735183	38.84513	4.59E-10	GEE
RENLFAIL	Renal failure	1	0.067668	-0.008668515	0.144004317	3.018548	0.0823171	GEE	1	-0.006822	-0.09140879	0.077764213	0.024989	0.8743937	GEE
WGHTLOSS	Weight loss	1	-0.333196	-0.422724153	-0.243667875	53.20797	0	GEE	1	-0.245002	-0.343742843	-0.1462617	23.65076	1.155E-06	GEE
MDC_1	MDC 1: Nervous System	1	-0.2864	-0.465720907	-0.107080013	9.79905	0.001746	GEE	1	-0.710923	-0.903668339	-0.518177004	52.26027	0	GEE
MDC_6	MDC 6: Digestive System	1	-0.452667	-0.627719352	-0.277614369	25.68721	4.015E-07	GEE	1	-0.630785	-0.821397729	-0.440172179	42.06829	8.8E-11	GEE
MDC_8	MDC 8: Musculoskeletal And Connective	1	-0.379936	-0.498109934	-0.261761272	39.70725	2.95E-10	GEE	1	-0.477477	-0.606698156	-0.348256753	52.44906	0	GEE
MDC_10	MDC 10: Endocrine Nutritional And Metabolic	1	-0.317145	-0.518239689	-0.116049734	9.554516	0.0019946	GEE	1	-0.565119	-0.783516721	-0.346722176	25.72067	3.946E-07	GEE
MDC_11	MDC 11: Kidney And Urinary Tract	1	-0.231784	-0.369614492	-0.093954075	10.86364	0.0009807	GEE	1	-0.207426	-0.359563793	-0.055288773	7.14087	0.0075347	GEE
MDC_13	MDC 13: Female Reproductive System	1	-0.551529	-0.808682385	-0.294375049	17.67043	2.627E-05	GEE	1	-0.340332	-0.620291076	-0.060372972	5.67692	0.0171895	GEE
MDC_21	MDC 21: Injuries Poison And Toxic	1	-0.220211	-0.39097095	-0.049450976	6.388543	0.0114859	GEE	1	-0.087455	-0.276558126	0.10164887	0.821604	0.3647116	GEE
mdrg_101	Intracranial vascular procedures	1	1.324325	1.016958974	1.631690229	71.31395	0	GEE	1	2.052292	1.713470269	2.391112757	140.9396	0	GEE
mdrg_102	Cranio w major dev impl/acute complex CNS	1	2.640205	2.37666961	2.903740055	385.5619	0	GEE	1	3.277607	2.99084992	3.564363521	501.8598	0	GEE
mdrg_103	Craniotomy & endovascular intracranial procedures	1	1.147988	0.880382495	1.41559342	70.69373	0	GEE	1	1.394267	1.100728669	1.68780492	86.66803	0	GEE
mdrg_105	Ventricular shunt procedures	1	1.065674	0.394236899	1.737111621	9.676853	0.001866	GEE	1	1.168499	0.440439562	1.896558839	9.895075	0.0016572	GEE
mdrg_106	Carotid artery stent procedure	1	-0.793014	-1.433221588	-0.152805425	5.894067	0.015192	GEE	1	-0.555109	-1.22696781	0.116749693	2.622387	0.1053658	GEE
mdrg_1101	Kidney transplant	1	-0.564185	-1.044936653	-0.083434281	5.290531	0.0214417	GEE	1	-0.628565	-1.142881739	-0.11424846	5.73767	0.0166048	GEE
mdrg_1304	Uterine & adnexa proc for non-malignancy	1	-0.613502	-1.075171656	-0.151832617	6.783689	0.0091995	GEE	1	-0.684816	-1.173917485	-0.195714704	7.53089	0.006065	GEE

mdrg_1401	Cesarean Section W CC	1	-0.858418	-1.142653458	-0.574182673	35.03786	3.234E-09	GEE	1	-0.766546	-1.070543989	-0.462547245	24.42472	7.727E-07	GEE
mdrg_1402	Vaginal Delivery W Sterilization OR DnC	1	-1.351955	-2.078612125	-0.625298724	13.29727	0.0002658	GEE	1	-0.958923	-1.70854853	-0.209296626	6.285995	0.0121696	GEE
mdrg_1708	Lymphoma & non-acute leukemia	1	0.885916	0.431749628	1.340083125	14.61676	0.0001317	GEE	1	0.959567	0.453120949	1.46601403	13.7905	0.0002044	GEE
mdrg_1801	Infectious & parasitic diseases w procedure	1	0.769189	0.669298669	0.869079011	227.7805	0	GEE	1	0.680946	0.569179152	0.792712691	142.5923	0	GEE
mdrg_601	Stomach esophageal & duodenal	1	0.651257	0.438250292	0.864264156	35.90987	2.067E-09	GEE	1	0.897713	0.66380075	1.131625795	56.58021	0	GEE
mdrg_602	Major small & large bowel proce	1	0.806589	0.628067397	0.985111045	78.41847	0	GEE	1	1.05825	0.863066104	1.25343414	112.9235	0	GEE
mdrg_604	Peritoneal adhesiolysis w MCC	1	0.80273	0.539313161	1.066145948	35.67378	2.333E-09	GEE	1	0.926629	0.636776351	1.216481156	39.26033	3.71E-10	GEE
mdrg_605	Appendectomy w complicated principal diag	1	0.817157	0.326624263	1.307690679	10.66034	0.0010946	GEE	1	0.779389	0.232144563	1.326634105	7.791873	0.0052482	GEE
mdrg_607	Minor small & large bowel procedures	1	0.931097	0.426420167	1.435774133	13.07553	0.0002992	GEE	1	1.171946	0.607478062	1.736413348	16.55896	4.716E-05	GEE
mdrg_610	Hernia procedures except inguinal & femoral	1	0.416176	0.101701468	0.730650158	6.727898	0.0094916	GEE	1	0.512224	0.170153159	0.854295419	8.61358	0.0033367	GEE
mdrg_611	Other digestive system O.R. procedures	1	1.188852	0.912225714	1.46547848	70.952	0	GEE	1	1.558697	1.249753709	1.867641013	97.78252	0	GEE
mdrg_706	Hepatobiliary diagnostic procedures	1	1.12746	0.685898374	1.569020748	25.04473	5.602E-07	GEE	1	1.401649	0.913529129	1.889769565	31.67534	1.822E-08	GEE
mdrg_707	Other hepatobiliary or pancreas procedures	1	0.69298	0.188986812	1.196972746	7.262537	0.0070408	GEE	1	1.009257	0.453710067	1.564803625	12.67822	0.0003699	GEE
mdrg_7702	Liver transplant w MCC	1	-1.49167	-1.888528068	-1.094812182	54.27147	0	GEE	1	-1.443246	-1.866972951	-1.01951877	44.56607	2.5E-11	GEE
mdrg_801	Combined anterior/posterior spinal fusion	1	-0.475213	-0.882881026	-0.067544638	5.21986	0.0223304	GEE	1	-0.316507	-0.753506896	0.120491913	2.015127	0.1557382	GEE
mdrg_802	Spinal fus exc cerv w spinal curv/malig/infec	1	-0.37162	-0.76886652	0.025625656	3.361833	0.0667238	GEE	1	0.065286	-0.363692372	0.49426422	0.088974	0.7654852	GEE
mdrg_806	Revision of hip or knee replacement	1	-0.606434	-0.923329752	-0.289537356	14.06788	0.0001763	GEE	1	-0.180816	-0.519362626	0.157729717	1.095813	0.2951869	GEE
mdrg_808	Cervical spinal fusion w MCC	1	0.610595	0.253372381	0.967817134	11.22341	0.0008077	GEE	1	0.144503	-0.231405255	0.520410582	0.567656	0.451192	GEE
mdrg_809	Amputation for musculoskeletal sys	1	0.728028	0.306293689	1.149762675	11.4476	0.0007159	GEE	1	0.450773	-0.003897632	0.905442655	3.77588	0.0519965	GEE
mdrg_8899	Non-Extensive O.R. Proc Unrelated to PDX	1	0.341109	0.194609575	0.487608389	20.82623	5.029E-06	GEE	1	0.454425	0.292732525	0.616117572	30.34176	3.622E-08	GEE
TPPS04D_ANY	ANY triggering complications were POA								1	-0.279828	-0.366820588	-0.192835427	39.74789	2.89E-10	GEE
TPPS04D_SEVERE	ANY of triggering complications were SEVERE								1	1.966517	1.898741307	2.034292625	3234.03	0	GEE

Table 5. PSI 04ERisk Model: GI Hemorrhage / Acute Ulcer Stratum

		Current Software (v60rev2)							Modification 1						
		C Stat = .860							C Stat = .878						
Variable	Description	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model	DF	Estimate	LowerWaldCL	UpperWaldCL	ChiSq	ProbChiSq	Model
Intercept	Intercept	1	-4.44257	-4.788397771	-4.096742036	633.935	0	GEE	1	-3.74193	-4.099962643	-3.383896876	419.6059	0	GEE
M_AgeCat_1	Male | Age < 30	1	1.137914	-1.071004218	3.346831413	1.019426	0.3126552	GEE	1	0.789786	-1.47754153	3.05711394	0.466108	0.4947831	GEE
M_AgeCat_2	Male | Age < 35	1	-1.342939	-3.517934031	0.832055047	1.464511	0.2262137	GEE	1	-1.331937	-3.509654939	0.845781667	1.43701	0.2306237	GEE
M_AgeCat_3	Male | Age < 40	1	-0.33853	-1.351129848	0.674069971	0.429353	0.5123066	GEE	1	-0.302259	-1.322852373	0.718334792	0.336937	0.5616027	GEE
M_AgeCat_4	Male | Age < 45	1	-0.066776	-0.762336647	0.62878494	0.035405	0.8507495	GEE	1	-0.099299	-0.808579488	0.609982386	0.075291	0.7837828	GEE
M_AgeCat_5	Male | Age < 50	1	-0.200627	-0.730481249	0.329227809	0.550758	0.4580083	GEE	1	-0.265174	-0.805380456	0.275031597	0.925637	0.3359993	GEE
M_AgeCat_6	Male | Age < 55	1	-0.234476	-0.618619736	0.149667725	1.431217	0.2315659	GEE	1	-0.194221	-0.583898116	0.195456111	0.954286	0.3286306	GEE
M_AgeCat_7	Male | Age < 60	1	0.081011	-0.250781657	0.412803616	0.229008	0.6322604	GEE	1	0.082415	-0.254150168	0.41898	0.23034	0.6312717	GEE
M_AgeCat_8	Male | Age < 65	1	0.101315	-0.262247131	0.464877836	0.298324	0.5849349	GEE	1	0.078897	-0.288535226	0.446328807	0.177117	0.6738624	GEE
M_AgeCat_9	Male | Age >=65	1	0.166279	-0.194687443	0.52724512	0.815149	0.3666024	GEE	1	0.133516	-0.231829616	0.498860619	0.513042	0.4738249	GEE
M_AgeCat_10	Male | Age >=70	1	-0.144669	-0.465817997	0.176480067	0.779532	0.3772844	GEE	1	-0.15967	-0.486306012	0.166966799	0.917934	0.338018	GEE
M_AgeCat_11	Male | Age >=75	1	0.203364	-0.11723903	0.523967438	1.545644	0.2137797	GEE	1	0.213954	-0.111155492	0.539063752	1.663711	0.197103	GEE
M_AgeCat_12	Male | Age >=80	1	-0.088393	-0.445538877	0.268752728	0.235311	0.6276137	GEE	1	-0.083488	-0.443287	0.276310427	0.206837	0.649258	GEE
M_AgeCat_13	Male | Age >=85	1	0.219501	-0.216249326	0.655251147	0.974751	0.3234982	GEE	1	0.206207	-0.232845507	0.645259785	0.847365	0.3572989	GEE
F_AgeCat_1	Female | Age < 30	1	0.085328	-1.468429649	1.639086501	0.011586	0.9142843	GEE	1	-0.228578	-1.890918734	1.433762464	0.072632	0.787543	GEE
F_AgeCat_2	Female | Age < 35	1	0.063528	-1.445265535	1.572321671	0.00681	0.9342295	GEE	1	0.124545	-1.386916871	1.636007038	0.026083	0.8716982	GEE
F_AgeCat_3	Female | Age < 40	1	-0.377263	-1.529737991	0.77521188	0.411644	0.5211363	GEE	1	-0.510425	-1.668770762	0.647921671	0.745903	0.3877764	GEE
F_AgeCat_4	Female | Age < 45	1	-0.588447	-1.392002867	0.215108959	2.060054	0.1512048	GEE	1	-0.503925	-1.316500187	0.308649631	1.477412	0.2241799	GEE
F_AgeCat_5	Female | Age < 50	1	0.380378	-0.2699745	1.030729623	1.314102	0.2516534	GEE	1	0.394564	-0.262927718	1.05205663	1.38341	0.2395211	GEE
F_AgeCat_6	Female | Age < 55	1	-0.159194	-0.732841717	0.414454599	0.295839	0.5865026	GEE	1	-0.187245	-0.767268121	0.392778024	0.400338	0.526915	GEE
F_AgeCat_7	Female | Age < 60	1	-0.405058	-0.865679259	0.055562856	2.970595	0.0847907	GEE	1	-0.403567	-0.871097477	0.063963245	2.862253	0.0906804	GEE
F_AgeCat_8	Female | Age < 65	1	0.13481	-0.273130937	0.542751648	0.419515	0.5171792	GEE	1	0.093719	-0.31972626	0.507164503	0.197386	0.6568391	GEE
F_AgeCat_10	Female | Age >=70	1	0.35511	-0.020176044	0.730395705	3.439514	0.0636544	GEE	1	0.30047	-0.079770359	0.68071057	2.398734	0.1214335	GEE
F_AgeCat_11	Female | Age >=75	1	-0.204926	-0.568872962	0.159021306	1.217904	0.2697724	GEE	1	-0.213709	-0.582816245	0.155398326	1.287764	0.2564604	GEE
F_AgeCat_12	Female | Age >=80	1	0.199606	-0.181453982	0.580666684	1.054041	0.3045783	GEE	1	0.219484	-0.166007761	0.604975484	1.245293	0.2644534	GEE
F_AgeCat_13	Female | Age >=85	1	-0.109075	-0.548330288	0.330179939	0.236872	0.6264743	GEE	1	-0.11434	-0.560249717	0.331570174	0.252579	0.6152652	GEE
TRNSFER	Transfer	1	0.489438	0.306343374	0.672533576	27.44979	1.612E-07	GEE	1	0.471291	0.284472018	0.658110261	24.44737	7.637E-07	GEE
ANEMDEF	Deficiency Anemias	1	-0.441698	-0.601677876	-0.281717523	29.2829	6.255E-08	GEE	1	-0.318719	-0.48091518	-0.156523542	14.83316	0.0001175	GEE
BLDLOSS	Chronic blood loss anemia	1	-0.540293	-0.917990042	-0.162596522	7.860837	0.0050517	GEE	1	-0.356319	-0.738872464	0.026235281	3.332636	0.0679179	GEE
CHF	Congestive heart failure	1	0.338979	0.143651283	0.534306878	11.56948	0.0006704	GEE	1	0.290676	0.092764324	0.48858794	8.286507	0.0039941	GEE
COAG	Coagulopthy	1	0.755406	0.570194511	0.940616501	63.90325	0	GEE	1	0.794898	0.606503531	0.98329306	68.38828	0	GEE
DEPRESS	Depression	1	-0.561478	-0.829993732	-0.292962146	16.79662	4.161E-05	GEE	1	-0.495715	-0.766089639	-0.225341049	12.91309	0.0003263	GEE
DM	Diabetes w/o chronic complications	1	-0.235175	-0.402501305	-0.067849264	7.588441	0.0058744	GEE	1	-0.218744	-0.3882854	-0.049203076	6.394684	0.0114463	GEE
DMCX	Diabetes w/ chronic complications	1	-0.440992	-0.692851711	-0.189133014	11.7772	0.0005996	GEE	1	-0.442256	-0.696457713	-0.188054207	11.6275	0.0006498	GEE
HTN_C	HTN_C	1	-0.306454	-0.44704543	-0.165863466	18.25211	1.935E-05	GEE	1	-0.228892	-0.371990682	-0.085792414	9.828374	0.0017184	GEE
IMMUNE	Immune disorders	1	1.052056	0.906054817	1.198056449	199.4633	0	GEE	1	0.87665	0.727832232	1.025466769	133.3035	0	GEE
LIVER	Liver disease	1	0.364703	0.137276721	0.592128452	9.878572	0.0016721	GEE	1	0.324471	0.084422882	0.564518836	7.018616	0.0080667	GEE
LYTES	Fluid and electrolyte disorders	1	0.776294	0.63220898	0.920379716	111.5088	0	GEE	1	0.825381	0.67894354	0.971818916	122.0393	0	GEE
NEURO	Neurological disease	1	0.289627	0.042768066	0.536485689	5.28782	0.0214751	GEE	1	0.338619	0.088989107	0.588248973	7.068473	0.0078452	GEE
OBESE	Obesity	1	-0.318324	-0.537863031	-0.098784803	8.076271	0.0044849	GEE	1	-0.354202	-0.577052747	-0.131351848	9.704472	0.0018382	GEE
PERIVASC	Peripheral vascular disease	1	0.27101	0.101022391	0.440996993	9.764115	0.0017795	GEE	1	0.293953	0.121462142	0.466444673	11.15627	0.0008375	GEE
PSYCH	Psychoses	1	-0.758137	-1.209540369	-0.306733805	10.83583	0.0009955	GEE	1	-0.756669	-1.213624781	-0.299713318	10.53319	0.0011725	GEE
MDC_4	MDC 4: Respiratory System	1	2.100987	1.767219106	2.434754382	152.2139	0	GEE	1	1.915819	1.577899428	2.253738388	123.4749	0	GEE
MDC_5	MDC 5: Circulatory System	1	2.002518	1.764300753	2.240734639	271.4586	0	GEE	1	1.804204	1.562309183	2.046099374	213.7044	0	GEE
MDC_17	MDC 17: Myeloproliferative DDs	1	1.409854	0.95030567	1.869401969	36.15616	1.821E-09	GEE	1	1.27207	0.785997229	1.758143513	26.30966	2.908E-07	GEE
MDC_18	MDC 18: Infectious and Parasitic	1	1.971259	1.74050323	2.202014059	280.3365	0	GEE	1	1.895693	1.661729354	2.129656184	252.1948	0	GEE
mdrg_101	Intracranial vascular procedures	1	2.884443	2.102374906	3.666510237	52.25542	0	GEE	1	2.689377	1.902085347	3.4766693	44.8258	2.2E-11	GEE
mdrg_102	Cranio w major dev impl/acute complex CNS	1	3.047674	2.650165975	3.445181453	225.8094	0	GEE	1	2.890134	2.48354244	3.296724903	194.0957	0	GEE
mdrg_1802	Postoperative or post-traumatic infections	1	-0.744009	-1.387506532	-0.100511403	5.135216	0.0234452	GEE	1	-0.707556	-1.356054093	-0.059056962	4.572979	0.03248	GEE
mdrg_401	Major chest procedures	1	-0.784301	-1.283357962	-0.285243692	9.487701	0.0020685	GEE	1	-0.686133	-1.191143613	-0.181121946	7.091059	0.0077469	GEE
mdrg_402	Other resp system O.R. procedures	1	-0.344545	-0.783625658	0.09453592	2.365369	0.1240546	GEE	1	-0.162447	-0.607649409	0.282754906	0.511454	0.4745102	GEE
mdrg_502	Perc cardiovasc proc w non-drug-eluting stent	1	-0.655705	-0.996228298	-0.315181869	14.2436	0.0001606	GEE	1	-0.537246	-0.882783782	-0.191708968	9.286521	0.0023085	GEE
mdrg_507	Coronary bypass w cardiac cath	1	-1.154328	-1.604899584	-0.703757062	25.2132	5.133E-07	GEE	1	-1.059487	-1.514316721	-0.604656901	20.84439	4.981E-06	GEE
mdrg_509	Amputation for circ sys disorders	1	-0.62799	-1.213039639	-0.042940222	4.426047	0.0353944	GEE	1	-0.656102	-1.246024579	-0.066179281	4.751696	0.0292694	GEE
mdrg_510	Permanent cardiac pacemaker implant	1	-2.083787	-3.50253203	-0.665042921	8.286948	0.0039931	GEE	1	-1.897814	-3.316101748	-0.479526722	6.878204	0.0087253	GEE

mdrg_511	Perc cardiovasc proc w drug-eluting stent	1	-1.032303	-1.327318967	-0.737286738	47.03476	7E-12	GEE	1	-0.967141	-1.265805351	-0.668475771	40.28166	2.2E-10	GEE
mdrg_513	Perc cardiovasc proc w/o coronary artery stent	1	-0.461578	-0.868009296	-0.055146698	4.95464	0.0260207	GEE	1	-0.326564	-0.738630437	0.085502826	2.412667	0.1203572	GEE
mdrg_514	Other vascular procedures	1	-0.946246	-1.266164871	-0.626326679	33.60654	6.747E-09	GEE	1	-0.852278	-1.175987608	-0.528568614	26.62858	2.466E-07	GEE
mdrg_515	Upper limb & toe amputation	1	-1.04204	-2.265903784	0.181824144	2.784829	0.0951609	GEE	1	-0.904198	-2.142971561	0.334575871	2.046629	0.1525436	GEE
mdrg_519	Other circulatory system O.R. procedures	1	-0.754046	-1.250251625	-0.257840334	8.870915	0.0028975	GEE	1	-0.567302	-1.072443791	-0.062159269	4.845032	0.0277259	GEE
mdrg_807	Major joint replacement or reattachment	1	-1.986636	-2.884263255	-1.089009734	18.81661	1.439E-05	GEE	1	-1.963179	-2.861471095	-1.064887394	18.34768	1.84E-05	GEE
TPPS04E_ANY	ANY triggering complications were POA								1	-1.043441	-1.180253383	-0.90662867	223.4505	0	GEE
TPPS04E_SEVERE	ANY of triggering complications were SEVERE								1	0.706614	0.349440205	1.063787705	15.03493	0.0001055	GEE

Appendix B. Literature Review – Preventing Failure to Rescue

Recent research targeting preventing failure to rescue (FTR) events in the surgical population continues to demonstrate a strong association with structure and processes of care, which are often intermingled.

Structures and processes of care

Residents and hospital

The first set of articles looked at the impact of residents. Use of residents can represent a process of care, but can also be a surrogate for structural measures or components. Performing a retrospective cohort analysis using the National Surgical Quality Improvement Program (NSQIP) Participant User Files for 2005-2009, Castleberry et al. (2013) found that resident trainee participation in complex, oncologic surgery was associated with significantly higher rates of 30-day postoperative complications in NSQIP-participating hospitals; however, this effect was countered by overall lower 30-day mortality and improved rescue rate in preventing death among patients suffering complications (5.9 vs. 7.6%, AOR 0.79, 95% CI 0.68-0.90). Navathe, A. S., et al. (2013) found no differences in FTR in teaching hospitals when comparing rates prior to and after ACGME duty hour reform, suggesting that resident fatigue was not a major factor contributing to mortality given complications.

Ferraris et al (2014), also using the NSQIP database to evaluate failure to rescue among 200 hospitals over 5-years, found that FTR was lower with resident involvement (9.4% vs 12.4% for attending alone; $P < .001$), despite significantly increased operative morbidity (11.4% vs 7.8% with attending only; $P < .001$) and prolonged operative time (127 minutes vs 93 minutes for attending only; $P < .001$). The most serious complications occurred 5 to 10 days before death, suggesting that there is a window for intervention to rescue patients with early aggressive treatments (especially of the initial sentinel complication).

Hospital level variables

A systemic review of 42 relevant papers (1980-2012) by Johnston et al found that the overall incidence of FTR ranged between 8.0 and 16.9%. Two studies demonstrated that FTR discriminated high- and low-volume hospitals better than morbidity measures. Greater hospital volume was associated with lower FTR rates in four studies. Two studies that analyzed the effect of the Safe Practices of the National Quality Forum revealed that FTR was less frequent in those with greater compliance. A higher level of nurse staffing was associated with lower FTR rates in two studies, with no significant association in another study. One study analyzed several hospital characteristics and found that the following were associated with lower FTR rates: teaching status, hospital size >200 beds, daily census >50%, increased nurse-to-patient ratios, and use of technology. Lower FTR rates were associated with patient age <70 years, absence of malignancy, and gastrointestinal complications. Higher FTR rates were found in patients with medical complications (compared with surgical complications), surgical site infections, deep vein thrombosis, pneumonia, sepsis, and non-white ethnicity. Mortality rates were higher among patients with escalation delay (defined as the lack of timely recognition or action when caring for

a deteriorating patient) compared with no delay in three studies. One study found that a rapid transfer to the ICU decreased mortality.

Using the 2007 to 2011 Nationwide Inpatient Sample (NIS), Hyder et. al., looked at high-mortality hospitals compared with low-mortality hospitals and found a difference in FTR rates (22.4% vs 20.2%, $p = 0.0020$). They then used Monte Carlo models to estimate the potential overall mortality reduction that could be achieved by focusing on 5 target subpopulations, assuming that these target subpopulations in high-mortality hospitals could achieve the same mortality rates as analogous patients in low-mortality hospitals. Approximately 50% absolute reduction in baseline total mortality (5.22%) was demonstrated when targeting FTR (2.73%; 95% CI 2.61 to 2.87) and reducing the FTR mortality gap by nearly 75%.

A Japanese study using PSI4 (v4.2) (Kitazaqa et al, 2014), found that low-volume hospitals had more deaths among surgical inpatients with serious treatable complications (38.5%, 95% CI, 33.7% to 43.2%) than high-volume hospitals (21.4%, 95% CI, 19.0% to 23.9%). In multiple linear regression analysis, each additional surgical patient per month was associated with 0.2 fewer cases ($p < 0.01$) of PSI04.

Nursing and nurse staffing

FTR is a considered a nurse sensitive indicator. Many of the recent research articles looked at the association between nursing staffing and FTR with mixed results.

Mark et al studied the impact of nurse staffing regulations in CA on FTR using PSI04. Hospitals were divided into quartiles based on how their pre-regulation RN ratios compared to other States. California quartile 1 hospitals had lower RN staffing compared to other states whereas quartiles 3 and 4 hospitals had higher staffing. Although there was a 35% post-regulation increase in RN staffing in quartile 1 hospitals, the associated change in FTR was not significantly different than the change in the comparison hospitals. However, the researchers saw a statistically significant differential decrease in FTR in California Quartile 4 hospitals, where the differential increase in RNs was slightly greater than in any other quartile. These mixed findings suggest other factors, such as culture or other staffing enhancements, influence FTR rates. Besides staffing, the type of nurse has been shown to make a difference. Using robust regression with clustering by UHC hospital type, Blegen et al (2013) found that as RN education increased in University affiliated hospitals, FTR decreased ($r = -0.399$; $p < 0.05$). This is supported by work by Kutney-Lee et al. (2015), who found that between 1999 and 2007, 11 Magnet recognized hospitals in the State of Pennsylvania had 6.1 fewer FTR deaths per 1000 patients ($P=0.02$) than the 125 non-Magnet Pennsylvania comparison hospitals.

Supporting the preventability of FTR events, Brooke, et al. (2012) found that hospitals that complied fully with the 27 National Quality Forum (NQF) safe practices had an increased likelihood of diagnosing a complication after any of six high-risk operations (odds ratio [OR], 1.13; 95% confidence interval [CI], 1.03-1.25), but had a decreased likelihood of failure to rescue (OR, 0.82; 95% CI, 0.71-0.96), and a decreased odds of mortality (OR, 0.80; 95% CI, 0.71-0.91).

Park et al examined the relationship between RN staffing and FTR and evaluated the effect of patient turnover on that relationship in 42 UHC hospitals. In general, they found that more RN hours per patient day were associated with lower rates of FTR, controlling for non-RN staffing and hospital characteristics. They also found that patient turnover rates differed by unit type, with higher turnover on non-ICUs (56.1%) than ICUs (45.4%), but with no direct effect of patient turnover on FTR rates in either non-ICUs or ICUs. There was an interaction effect between patient turnover and RN staffing on non-ICUs, however, indicating that the association between RN staffing and FTR differed significantly depending on the level of patient turnover. They also found that higher technological complexity in hospitals was related to lower FTR ratios (Mark & Harless, 2010; Mark et al., 2004).

Risk adjustment

Ferris et al reported that more than two-thirds of patients with failure to rescue have multiple complications. A risk-scoring system based on preoperative variables predicted patients in the highest-risk category of failure to rescue with good accuracy. Cardiac events were associated with the highest failure-to-rescue rates, but stroke, renal failure, and pulmonary failure had nearly as high FTR risk. The following table includes the percent mortality with and without the complication using propensity matching on preoperative variables.

Sheetz and colleagues evaluated whether increased hospital care intensity (HCI) is associated with improved outcomes following seven major cardiovascular, orthopedic, or general surgical operations in the Medicare population. High-HCI hospitals had greater rates of major complications than low-HCI centers (risk ratio, 1.04; 95% CI, 1.03-1.05) and there was a decrease in failure to rescue at high compared with low-HCI hospitals (risk ratio, 0.95; 95% CI, 0.94-0.97). Using multilevel-models, HCI reduced the variation in failure-to-rescue rates between hospitals by 2.7% after accounting for patient comorbidities and hospital resources.

Table 2. Effect of Complication Type on Failure-to-Rescue Rates

Type of Complication	No. of Complications	Mortality Rate, %	
		Without Complication	With Complication
Cardiac	13 629	1.2	45.5
Renal	13 338	1.3	31.0
CNS	6130	1.5	28.0
Pulmonary	56 546	0.8	24.5
Sepsis/SIRS	45 464	1.1	17.8
Reoperation for bleeding	102 034	1.1	8.7
DVT/PE	17 631	1.5	8.0
Wound	41 844	1.5	4.8
Any serious complication	207 236	0.5	10.1

Abbreviations: CNS, central nervous system; DVT, deep vein thrombosis; PE, pulmonary embolism; SIRS, systemic inflammatory response syndrome.

Healthcare Disparities

Silber et al using Medicare data from six States found that when matching on age, sex, year, state, and the exact type of procedure, blacks had higher failure-to-rescue rates (6.1% vs. 5.1%, $P<0.001$) than whites. When preoperative medical risk factors were added to this matching algorithm, there was no significant racial difference in FTR. These findings confirm that race/ethnicity should not be included in PSI04 risk-adjustment.

Indicator specifications

Needleman et al, using PSI04 v 3.1, examined whether the accuracy of PSI04 (v3.1) could be improved by testing three exclusion rules using California HCUP Data. All POA-informed specifications of exclusion rules improved the C-statistic of the failure-to-rescue measure and its sensitivity, with modest losses of specificity. For the entire failure-to-rescue pool, the mortality rate was 22% among patients with hospital-acquired complications compared to 13% for patients with POA complications, and the risk-adjusted difference in mortality was 6.6% (from 1.9% for pneumonia to 7.1% for sepsis). Lengths of stay were longer for patients with hospital-acquired complications than for patients with POA complications: 14.1 versus 8.4 days, with a risk-adjusted difference of 4.3 days (from 2.2 days for shock/cardiac arrest to 5.4 days for sepsis and GI bleeding). For all tested specifications, higher licensed hours and proportions of registered nurses were statistically significant predictors of lower FTR rates.

Summary

1. Better processes and structures of care are associated with lower FTR rates.
2. Nursing type, higher nurse staffing, and lower unit turn-over are associated with lower FTR rates.
3. Use of residents is associated with higher complication rates, but lower mortality and better rescue efforts.
4. There is no new evidence to support treating academic hospitals differently than non-academic facilities in risk adjustment models.
5. Based on the work by Needleman et al, a patients' risk for death and extended (and more expensive) treatment appear to be higher when their secondary diagnoses are hospital-acquired complications versus being POA.
6. The work by Silber emphasis the need to risk-adjust for preoperative medical factors to account for healthcare disparities.

Appendix C. References

Blegen MA, Goode CJ, Park SH, Vaughn T, Spetz J. Baccalaureate education in nursing and patient outcomes. *J Nurs Adm.* 2013;43(2):89-94.

Blegen MA, Goode CJ, Park SH, Vaughn T, Spetz J. Baccalaureate education in nursing and patient outcomes. *J Nurs Adm.* 2013;43(2):89-94.

Brooke BS, Dominici F, Pronovost PJ, Makary MA, Schneider E, Pawlik TM. Variations in surgical outcomes associated with hospital compliance with safety practices. *Surgery.* 2012;151(5):651-659.

Brooke BS, Dominici F, Pronovost PJ, Makary MA, Schneider E, Pawlik TM. Variations in surgical outcomes associated with hospital compliance with safety practices. *Surgery.* 2012;151(5):651-659.

Brooke BS, Dominici F, Pronovost PJ, Makary MA, Schneider E, Pawlik TM. Variations in surgical outcomes associated with hospital compliance with safety practices. *Surgery.* 2012;151(5):651-659.

Castleberry AW, Clary BM, Migaly J, et al. Resident education in the era of patient safety: a nationwide analysis of outcomes and complications in resident-assisted oncologic surgery. *Ann Surg Oncol.* 2013;20(12):3715-3724.

Ferraris VA, Bolanos M, Martin JT, Mahan A, Saha SP. Identification of patients with postoperative complications who are at risk for failure to rescue. *JAMA Surg.* 2014;149(11):1103-1108.

Hyder JA, Wakeam E, Adler JT, DeBord Smith A, Lipsitz SR, Nguyen LL. Comparing Preoperative Targets to Failure-to-Rescue for Surgical Mortality Improvement. *J Am Coll Surg.* 2015;220(6):1096-1106.

Johnston MJ, Arora S, King D, et al. A systematic review to identify the factors that affect failure to rescue and escalation of care in surgery. *Surgery.* 2015;157(4):752-763.

Kutney-Lee A, Stimpfel AW, Sloane DM, Cimiotti JP, Quinn LW, Aiken LH. Changes in patient and nurse outcomes associated with magnet hospital recognition. *Med Care.* 2015;53(6):550-557.

Mark BA, Harless DW, McCue M, Xu Y. A longitudinal examination of hosptial registered nurse staffing and quality of care. Health Serv Res. 2004; 39(5):1629.

Mark BA, Harless DW, Spetz J, Reiter KL, Pink GH. California's minimum nurse staffing legislation: results from a natural experiment. *Health Serv Res.* 2013;48(2 Pt 1):435-454.

Mark BA, Harless DW, Spetz J, Reiter KL, Pink GH. California's minimum nurse staffing legislation: results from a natural experiment. *Health Serv Res.* 2013;48(2 Pt 1):435-454.

Mark BA, Harless DW. Nurse staffing and post-surgical complicatins using the present on admission indicator. Res Nurs Health. 2010; 33(1):35-47.

Mattke S, Needleman J, Buerhaus P, Stewart M, Zelevinsky K. Evaluating the role of patient sample defintiions for quality indicators senstive to nurse staffing patterns. Med Care. 2004; 42(2 Suppl):II21-II33.

Navathe AS, Silber JH, Small DS, et al. Teaching hospital financial status and patient outcomes following ACGME duty hour reform. *Health Serv Res.* 2013;48(2 Pt 1):476-498.

Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. *N Engl J Med.* 2002;346(22):1715-1722.

Needleman J, Buerhaus PI, Vanderboom C, Harris M. Using present-on-admission coding to improve exclusion rules for quality metrics: the case of failure-to-rescue. *Med Care.* 2013;51(8):722-730.
Needleman J, Buerhaus PI. Failure-to-rescue: Comparing definitions to measure quality of care. *Medical Care.* 2007;45(10):913-915.
Park SH, Blegen MA, Spetz J, Chapman SA, De Groot H. Patient turnover and the relationship between nurse staffing and patient outcomes. *Res Nurs Health.* 2012;35(3):277-288.
Sheetz KH, Dimick JB, Ghaferi AA. The association between hospital care intensity and surgical outcomes in medicare patients. *JAMA Surg.* 2014;149(12):1254-1259.
Sheetz KH, Dimick JB, Ghaferi AA. The association between hospital care intensity and surgical outcomes in medicare patients. *JAMA Surg.* 2014;149(12):1254-1259.
Silber JH, Romano PS, Rosen AK, Wang Y, Even-Shoshan O, Volpp KG. Failure-to-rescue: comparing definitions to measure quality of care. *Med Care.* 2007;45(10):918-925.
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):615-629.