NATIONAL QUALITY FORUM

Memo

TO: Patient Safety Standing Committee

FR: NQF Members

RE: Voting Draft Report: *NQF Endorsed Measures for Patient Safety*

DA: October 21st, 2015

Background

Patient Safety related events due to medical errors result in tens of thousands of premature deaths each year. Currently, NQF's portfolio of safety measures spans a variety of topic areas including, but not limited to, health care associated infections, falls, pressure ulcers, surgical complications, and workforce issues. However, significant gaps remain in the measurement of patient safety and how providers approach minimizing the risk of patient safety events. There is also a recognized need to expand avoidable patient safety measures beyond the hospital setting, as well as harmonize safety measures across sites and settings of care.

On June 17-18, 2015, the 25-member Patient Safety Standing Committee evaluated four new measures and 19 maintenance measures. A total of 18 of 23 measures were recommended for endorsement, and one measure was not recommended. At the in person meeting, two measures were deferred and the Committee did not reach consensus on two measures. Following the comment period and the review of additional materials, the remaining four measures were recommended for endorsement by the Committee.

The Patient Safety Standing Committee also conducted *ad hoc* reviews of three additional measures. In two measures, definitions were changed and in one measure substantial changes were made that required a full review of all the NQF criteria. Ultimately, all three *ad hoc* review measures received continued endorsement.

The 22 measures recommended by the Standing Committee include:

- 0101: Falls: Screening, Risk Assessment, and Plan of Care to Prevent Future Falls (National Committee for Quality Assurance)
- 0141: Patient Fall Rate (American Nurses Association)
- 0202: Falls With Injury (American Nurses Association)
- 0204: Skill Mix Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN, Unlicensed Assistive Personnel [UAP], and Contract (American Nurses Association)
- 0205: Nursing Hours per Patient Day (American Nurses Association)
- 0337: Pressure Ulcer Rate (PDI 2) (Agency for Healthcare Research and Quality)
- 0347: Death Rate in Low-Mortality Diagnosis Related Groups (PSI02) (Agency for Healthcare Research and Quality)
- 0419: Documentation of Current Medications in the Medical Record (Quality Insights of Pennsylvania)
- 0537: Multifactor Fall Risk Assessment Conducted for All Patients Who Can Ambulate (Centers for Medicare & Medicaid Services)
- 0538: Pressure Ulcer Prevention and Care (Centers for Medicare & Medicaid Services)
- 0674: Percent of Residents Experience One or More Falls with Major Injury (Long Stay) (Centers for Medicare & Medicaid Services)
- 0679: Percent of High Risk Residents with Pressure Ulcers (Long Stay) (Centers for Medicare & Medicaid Services)
- 0687: Percent of Residents Who Were Physically Restrained (Long Stay) (Centers for Medicare & Medicaid Services)
- 0689: Percent of Residents Who Lose Too Much Weight (Long Stay) (Centers for Medicare & Medicaid Services)

- 2720: National Healthcare Safety Network (NHSN) Antimicrobial Use Measure (Centers for Disease Control)
- 2723: Wrong Patient Retract and Reorder (WP-RAR) (Montefiore Health System)
- 2726: Prevention of Central Venous Catheter (CVC)-Related Bloodstream Infections (American Society of Anesthesiologists)
- 2732: INR Monitoring for Individuals on Warfarin after Hospital Discharge (Centers for Medicare and Medicaid Services/Mathematica)
- 0097: Medication Reconciliation Post-Discharge (National Committee for Quality Assurance)
- 0531: Patient Safety for Selected Indicators (PSI90) (Agency for Healthcare Research and Quality)
- 0352: Failure to Rescue In-Hospital Mortality (risk adjusted) (The Children's Hospital of Philadelphia): Tabled for further discussion
- 0353: Failure to Rescue 30-Day Mortality (risk adjusted) (The Children's Hospital of Philadelphia): Tabled for further discussion

The Committee did not recommend the following measure:

- 2729: Timely Evaluation of High-Risk Individuals in the Emergency Department (Centers for Medicare and Medicaid Services/Mathematica)

The Committee conducted an *ad hoc* review and approved the changed specifications for three measures:

- 0138: National Healthcare Safety Network (NHSN) Catheter Associated Urinary Tract Infection (CAUTI) (Centers for Disease Control and Prevention)
- 0139: National Healthcare Safety Network (NHSN) Central Line Associated Bloodstream Infection (CLABSI) (Centers for Disease Control and Prevention)
- 0345: Unrecognized Abdominopelvic Accidental Puncture or Laceration Rate (PSI15) (Agency for Healthcare Research and Quality)

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 May 4th - May 22nd, 2015 for twenty-three of the twenty-six measures under review (the ad hoc review measures were not included in this comment period). A total of seven pre-evaluation comments were received, the majority of which pertained to measure *0531: Patient Safety for Selected Indicators*, and expressed concerns with the specifications of several of the component measures. All of these pre-evaluation comments were provided to the Committee prior to the June 17-18th in-person meeting.

Post-evaluation comments

The Draft Report went out for Public and Member comment August 3rd - September 3rd, 2015. During this commenting period, NQF received 282 comments from 19 member organizations and 62 members of the public:

Consumers – 8	Professional – 49
Purchasers – 6	Health Plans – 27
Providers – 29	QMRI – 24
Supplier and Industry – 3	Public & Community Health - 134

Comments and their Disposition

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

1. Level of analysis
2. Support for measures
3. Implementation issues (burden on providers, unintended consequences, general readiness)
4. Requests for changes (numerator, denominator, risk adjustment, definitions)
5. Small number of cases
6. Preference for outcome measures

Theme 1 – Level of Analysis

NQF received comments on most of the measures stating that they were not appropriate for health plan level analysis, although the commenters were generally supportive of the measures. However, only one measure, *0097: Medication Reconciliation Post-Discharge*, is specified at the health plan level, so no changes are requested. As this comment did not directly apply to most of the measures, it is not included in the measure-specific comment discussions below.

Theme 2 – Support for Measures

A number of measures, including *0139: National Healthcare Network (NHSN) Central Line Associated Bloodstream Infection (CLABSI) Outcome Measure*, *0204: Skill Mix Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN, Unlicensed Assistive Personnel [UAP], and Contract*, *0205: Nursing Hours per Patient Day*, *0537: Multifactor Fall Risk Assessment Conducted for All Patients Who Can Ambulate*, and *0538: Pressure Ulcer Prevention and Care* received all supportive comments. The comments included many reasons for support of each particular measure, but many particularly noted the importance of a particular measure for improving patient safety and for public reporting. Measure *0531: Patient Safety for Selected Indicators* received extensive support from many commenters; the details are below in the measure-specific comments section.

Theme 3 – Implementation Issues

Commenters raised a number of implementation issues on several measures. The issues raised included the burden of reporting on providers, issues with CPT II coding, potential unintended consequences of a particular measure, and the general readiness of measures for use in public reporting and payment programs. These are noted under the individual measure, as appropriate.

Theme 4 – Request for Changes to Measures

Commenters suggested changes to many measures; ranging from changes to the numerator or denominator; requests for risk adjustment for particular populations; and revising definitions to be clearer.

Theme 5- Small Number of Cases

For two measures (*0337: Pressure Ulcer Rate (PDI 2*) and *0347: Death Rate in Low-Mortality Diagnosis Related Groups (PSI02)*), that cover infrequent events, commenters were concerned that the small number of cases could make the measure challenging to use and suggested that the Committee look into this; however, they noted it would be important to catch these incidents for follow up with providers.

Theme 6 – Preference for Outcome Measures

For two measures (*0419: Documentation of Current Medications in the Medical Record* and *2723: Wrong Patient Retract and Reorder (WP-RAR)*), commenters noted a preference for outcome measures rather than the currently specified process measures.

Committee Response: In general, the Committee would prefer outcome measures rather than process or structural measures; however, when measuring the process or structure may still be useful for quality improvement or other purposes, there still can be a role for these types of measures, especially where outcomes may be difficult to measure. Measure 419 is a process measure of attestation to the documentation of a medication list. NQF does have a related endorsed measure for adverse drug events: *0709: Proportion of patients with a chronic condition that have a potentially avoidable complication during a calendar year*. In this case, the Committee thinks that attestation of medication reconciliation is still an important area that deserves endorsement because not conducting medication reconciliation can have important consequences to patients. In addition, transitions of care are a particularly high-risk time, especially after a hospitalization where medications may change. For Measure 2723, the Committee has determined that this is an outcome measure because while the error did not actually reach the patient, a wrong-patient retract and reorder in the electronic health record is still a medical error.

Measure Specific Comments

Measures Recommended for Endorsement

0101: FALLS: SCREENING, RISK-ASSESSMENT, AND PLAN OF CARE TO PREVENT FUTURE FALLS

This measure received six comments, from a variety of organizations, with mixed levels of support; some were urging the Committee to reconsider endorsement as the measure is currently specified. All comments cited the importance of measurement in this area. Two comments raised concerns with use of CPT II codes and mentioned the burden on providers because it requires data from medical charts to calculate the numerator unless a random sampling methodology is used. One comment recommended the measure be broken into three individual measures and another suggested that the measure be closely aligned to the Medicare Annual Wellness visit that includes all risk assessment and personalized health advice aimed at fall prevention. Lastly, one comment suggested removing nursing home and assisted living patients from the denominator because the process of gathering information to accurately report the measure has created an undue burden.

Developer Response: The developers acknowledged the need to harmonize with the Medicare Annual Wellness visit. Providers conducting an assessment and offering evidenced-based falls risk interventions as part of the Medicare Annual Wellness visit would meet the numerator for the rates in this measure. The three rates on this measure were combined into a single measure at the request of the NQF Patient Safety Committee when the measure was presented for re-endorsement in 2012. The developers are willing to separate the measures into their original format if the Standing Committee advises. Finally, the developers agree not all patients have the resources to attend physical therapy or exercise programs beyond those benefits covered by Medicare. However, it is important providers advise patients about the need for this type of intervention and help connect seniors to resources, such as falls risk prevention programs, in their communities.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

0141: PATIENT FALL RATE

This measure received 27 comments supporting re-endorsement (particularly with the expanded level of analysis), from a variety nursing associations and patient advocacy groups. Only two comments were opposed to endorsement; one suggested that the definition of falls is too broad and both comments raised implementation concerns because the measure relies on electronic or paper medical records rather than administrative claims.

Developer Response: Data are collected through incident reporting systems, which are electronic and already in place in most hospitals. Feasibility studies have shown this measure has a low burden for hospitals currently collecting data. Collecting injury levels happens in the medical record 24 hours after the fall, because assignment of injury level has to follow medical evaluation. Assisted falls are built into the measure through National Database of Nursing Quality Indicators NDNQI, but aren't currently included in this definition of the measure. Reason for fall has also been added to the NDNQI measure, but has not been fully tested. These are potential revisions that could be made to the measure in the future.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

0202: FALLS WITH INJURY

This measure received 30 comments supporting re-endorsement (particularly with the expanded level of analysis), from a variety of nursing associations and patient advocacy groups. One comment, while supporting the measure, requested that additional work be done to harmonize measures across settings of care. The comment also noted that all falls measures be re-evaluated after the release of the upcoming US Preventive Services Task Force (USPSTF) study on the effectiveness of falls prevention measures to ensure all endorsed measures are aligned with the best evidence.

Developer Response: The measure, as currently defined, is being proposed for acute care hospitals and their units. Currently, testing is being conducted on an expanded measure including pediatric and psychiatric units, which could be implemented in the future.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

0337: PRESSURE ULCER RATE (PDI 02)

This measure received three comments that were generally supportive, while also raising concerns. One comment supported the modification to include stage II pressure ulcers, but recommended an additional exclusion for patients who are receiving end-of-life care, as it may be too painful to move these patients or if they refuse to be repositioned. One commenter was concerned that provider-level of analysis may have small number issues. The final comment also supported the measure but raised the same concern with provider numbers being too small.

Developer Response: The developers are considering a number of important modifications to PDI 02. One of those changes is the inclusion of stage II pressure ulcers, as is consistent with several major pediatric patient safety efforts. They will be considering these changes using clinical and expert panel review and empirical analyses and changes will be implemented if deemed appropriate after this comprehensive evaluation. The developers appreciate the support for including Stage II pressure ulcers in further measure development. Given detailed data, exclusion to the indicator for actively dying patients makes sense from a clinical and patient preference perspective. With administrative data, however, it is difficult to identify patients for whom repositioning is contraindicated. During future indicator refinements, the developers will empirically test methods to exclude patients who may fit this circumstance based on data elements available.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

0347: DEATH RATE IN LOW-MORTALITY DIAGNOSIS RELATED GROUPS (PSI 02)

This measure received two comments, both stating support for the topic but raised concerns that measurement at the provider level may have issues with small numbers.

Developer Response: Hospitals with more than 205 eligible discharges, on average, have risk adjusted rates with moderate to high reliability (average signal-to-noise ratio of 0.422 to 0.840). Overall, the signal to noise ratio for this indicator is strong with a weighted mean value of 0.716. These findings were confirmed by Bernal-Delgado et al. (BMC Med Res Methodol 2012; 12:19), who analyzed data from 171-175 Spanish hospitals in 2005-2006. They estimated PSI 02 virtually unchanged (as Spain also uses ICD-9-CM for inpatient coding and MS DRGs for resource allocation). The Empirical Bayes estimator of systematic hospital-level variation in a two-stage hierarchical random effects model was 0.32, similar to the values for other NQF-endorsed AHRQ Patient Safety Indicators. Although "small number issues" may affect hospitals in the lower 20-30% of the national distribution of hospital volume, the high signal to noise ratio supports high reliability. Using more than 1 year of data may further improve the reliability of this measure.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

0419: DOCUMENTATION OF CURRENT MEDICATIONS IN THE MEDICAL RECORD

This measure received six comments, all with tepid support. Commenters agreed that accurate medication lists remain an area for improvement and that it is important information. However, all raised concerns with the measure, including:

- Information provided by patients may be inaccurate or incomplete (particularly for over the counter drugs or supplements), and it is impossible to fully validate;
- CPT II codes can be challenging for providers who do not use them regularly;
- A commenter requested the prioritization of measures of adverse drug event outcomes and noted this measure is not linked with a decrease in ADEs.

Developer Response:

- The developer agrees with this comment and recognizes the measure assesses a foundational practice and merely sets a minimum requirement that a medication review is performed. Without broader adoption of this practice, it will be difficult to realize improvement in adverse drug events (ADEs). The developer will consult with the expert work group to discuss and review approaches to addressing the issue.
- The developer cited several programs that currently use this measure such as the Physician Quality Reporting System (PQRS) program and may be reported via Claims/Registry, GPRO, and EHR.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

0679: PERCENT OF HIGH RISK RESIDENTS WITH PRESSURE ULCERS (LONG STAY)

This measure received two comments that were generally supportive. One comment also suggested the addition of wheelchair bound patients to the denominator.

Developer Response: The denominator for NQF #0679 includes all long-stay nursing home residents (length of stay is greater than 100 days) who had a target MDS assessment (OBRA, PPS, or discharge) during the selected measurement quarter and were identified as at high risk for pressure ulcer, except those meeting exclusion criteria. Residents must be high risk for pressure ulcer where high risk is defined by meeting one of the following criteria on the selected target assessment: 1. Impaired in bed mobility or transfer: This is indicated by a level of assistance reported on either item G0110A1), bed mobility (self-performance) or G0110B1 Transfer (self-performance) at the level of: extensive assistance, total dependence, activity occurred only once or twice OR activity or any part of the activities of daily living was not performed by resident or staff at all over the entire 7 day period. MDS 3.0 G0110B transfer includes how the resident moves between surfaces including to or from: bed, chair, wheelchair, standing position (excludes to/from bath/ toilet). Using the impairments in bed mobility and transfer as criteria should capture a large proportion of wheelchair bound long-stay residents in the denominator.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

0687: PERCENT OF RESIDENTS WHO WERE PHYSICALLY RESTRAINED (LONG STAY)

This measure also received two comments supporting re-endorsement. In addition, one comment requested that measurement of the utilization of restraint alternatives (chemical use vs. non-chemical alternatives) also needs to be evaluated along with physical restraint to prevent unintended consequences.

Developer Response: This measure is currently restricted to long-stay patients cared for in a nursing facility. The specifications for this measure are designed for the evaluation of the quality of nursing facility care. CMS's Nursing Home Compare also publicly reports a measurement of utilization of chemical alternatives to physical restraints: Percent of Long-Stay Residents Who Newly Received an Antipsychotic Medication, which indicates the proportion of long-stay residents without schizophrenia, Tourette's syndrome, or Huntington's disease who received an antipsychotic medication (MDS N0410A={1,2,3,4,5,6, or 7}) in the target period.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

0689: PERCENT OF RESIDENTS WHO LOSE TOO MUCH WEIGHT (LONG STAY)

The single comment on this measure raised the issue of appropriateness for health plan level measurement; however it is not specified at that level. The commenter also noted that evidence shows that nursing home patients have a higher mortality rate in the six months following a 10% loss of bodyweight.

Developer Response: NQF #0689 is an outcome measure that reports the percentage of long-stay nursing home residents with a target MDS assessment that indicates a weight loss of 5% or more in the last 30 days or 10% or more in the last 6 months, which is not a result of a physician-prescribed weight-loss regimen. Long-stay residents are identified as residents who have had at least 101 cumulative days of nursing facility care. This measure is currently restricted to long-stay patients cared for in a nursing facility. The specifications for this measure are designed for the evaluation of the quality of nursing facility care. The developer appreciates the comments on the association between weight loss and mortality among nursing home residents, and shares the same understanding. The evident higher mortality associated with excessive weight loss is one of the fundamental and most important reasons for publicly reporting this quality measure for nursing homes.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

2720: NATIONAL HEALTHCARE SAFETY NETWORK (NHSN) ANTIMICROBIAL USE MEASURE

This measure received fifteen comments. All of the comments recognized that antimicrobial resistance is a major public health issue and that antimicrobial stewardship programs can be effective in increasing appropriate use. Supportive comments noted that this is a critical need and that this measure will help establish baselines for antimicrobial use as well as develop a better understanding of the role of antimicrobial use in drug resistance.

Commenters (including both those supportive and not supportive of the measure) were concerned that, while this measure is appropriate for surveillance, it is not yet ready for public reporting and payment programs. Commenters suggested additional reliability and validity testing and suggested that the Committee consider recommending that this measure be excluded from public reporting. These comments also raised concerns that the measure has feasibility issues, noting that a standardized EMR guidance would be needed, as well as significant lead time to ensure that facilities have the necessary data mining capabilities. It was noted that this is a challenging topic to measure and additional concerns were raised about the selection of some of the drugs used in the measure. Commenters also raised concerns with the difference between utilization and appropriateness, noting that appropriateness incorporates many factors that were not fully accounted for, including geography, seasonal variation, prevalence, and patient mix, all of which could affect predicted use. Commenters noted the need for risk adjustment for cancer and transplant patients and the importance of controlling for differences between types of hospitals and the complexity of their patient population. One comment was particularly concerned with the pediatric population, noting that it is more complex than an adult population, and that the pediatric sample size was extremely small; they suggested further testing. Lastly, a commenter suggested this measure be expanded to include antifungal agents.

> **Developer Response:** The developer agrees that the measure is not yet ready for public reporting or incentive payment. However, they recommend use of the measure for quality improvement by hospitals, specifically as a benchmark that can assist efforts by antimicrobial stewardship programs to monitor antimicrobial use and foster data-driven improvements. The data used to predict antimicrobial use (AU) were reported to CDC's National Healthcare Safety Network (NHSN) in 2014 by a geographically diverse set of 60 U.S. hospitals including acute care hospitals, critical access hospitals, children's hospitals, and an oncology hospital. Each of these hospitals successfully implemented and validated the AU data reported electronically to NHSN, demonstrating the feasibility of implementation across a variety of hospital types. The summary statistics proposed for the measure are designed to provide benchmarks for antimicrobial use not appropriateness of use. As stated in the measure proposal, these summary statistics are a starting place for further analysis and possible action. Additional analyses to determine the appropriateness of antibiotic use are likely to require access to detailed, patient-level data that is beyond the scope of data collection and analysis using NHSN, e.g., clinical indications for specific antibiotics and dose and duration decisions. The developer appreciates concerns about antimicrobial agents that are not included in the antibacterial agent-patient care location categories and would be grateful to know which agents in particular have been omitted and "are often the most inappropriately used." The measure construct is extensible to additional antibacterial agent-patient care location pairings. The specific pairings included in the measure proposal are the product of extensive consultation with infectious disease physicians and pharmacists who are at the forefront of antimicrobials stewardship programs (ASPs) at their hospitals/health systems and the measure is intended for use by ASPs throughout the U.S. The developer agrees with the importance of including antifungal agents in the measure. They plan to do so when antifungal use data reported to CDC's National Healthcare Safety Network (NHSN) are sufficient to add antifungal agents to the measure.

> **Committee Response:** The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

2723: WRONG-PATIENT-RETRACT AND REORDER MEASURE

This measure received six comments, some supporting the intent but some raising concerns. The four comments that did not support the measure raised a variety of concerns, including:

- A concern that the measure could undermine the fair and just culture in hospitals;
- Concerns with the lack of exclusions, especially in cases where certain protocol orders are automated and then retracted by a physician;
- A comment noting that this measure does not focus on patient outcomes, rather, it focuses on staff errors.

The two supportive comments raised additional concerns, including:

- A concern that 10 minutes may not be long enough and that the measure could be potentially "gamed" by waiting longer;
- One comment included multiple concerns including suggesting a longer time window, potential false positives, a suggestion that the specificity should be increased for long-term use, possible unintended consequences of deterring self-reports, and inconsistencies in the denominator.

Developer Response:

- The measure is designed to hold health systems and vendors accountable for the design and configuration of their EHRs that may increase the risk of wrong-patient errors and to test the effectiveness of interventions. It is not designed as a measure of individual provider performance.
- Once a provider realizes that they have placed an order on the wrong patient, they are highly motivated (if not anxious) to remove that order before any actions are taken as a wrong patient error is an egregious mistake. For example, in a JAMIA paper the developers report that 6,885 WP RAR events occurred in one year at one hospital, and the mean time of retraction was just 1 minute and 18 seconds. They tested a longer window, and it increased false positives (not a good option).
- Since submission, a second hospital (the VA New York Harbor Healthcare System) has replicated the measure in a different EMR, and has also replicated the validation process with near-real time phone calls. To date 45 out of 58 calls were true positives with a PPV of 77.6%. This PPV is very similar to the original PPV of 76.2% and is reassuring.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

2726-PREVENTION OF CENTRAL VENOUS CATHETER (CVC) - RELATED BLOODSTREAM INFECTIONS

This measure received five comments that all expressed support for measurement in this area but all highlighted points of concern. These concerns pointed to potential challenges documenting and reporting the measure. One comment stated that the measure may present a challenge when patients are transferred from another facility with a central line already in place. A few comments stated that a review of best practices may be more beneficial than monitoring.

Developer Response: Anesthesia providers and others who perform central line insertion influence patient outcomes because of this process of care. The healthcare industry has already seen this result in the lowered the rate of bloodstream infections (after implementation of NQF #0464 and other related measures) and there are national campaigns to drive the bloodstream infections closer to zero. The developers recognize that we cannot control what happens to the patient over their length of stay, but anesthesia providers (and their practices and those within the anesthesia care team) have the clinical responsibility to ensure that CVC-related bloodstream infections are reduced.
The developers appreciate the concern with patients who are transferred from one location to another location. They will take that under consideration as the role of quality and performance reporting continue to evolve. Previous specifications of this measure have used the CPT II Code and the developers anticipate few issues with implementing this measure. ASA is aware of the need to develop the e-specifications for this particular measure and they are open to collaboration between

interested parties to ensure that all anesthesia and other healthcare providers have the means to report this measure.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

2732- INR MONITORING FOR INDIVIDUALS ON WARARIN AFTER HOSPITAL DISCHARGE

This measure received four comments which expressed support for the concept but there were a few concerns raised. One comment questioned how the INR information will be captured because it may be burdensome. Another comment suggested making changes to the denominator definition, revising the upper bound from INR>=5 and INR>=4 and making discharged hospitals accountable for patient follow-up.

Developer Response: The developers agree that there is evidence indicating a number of different ranges to define therapeutic INR. However, this measure is designed to detect a pre-discharge INR that is more than 0.5 outside of two of the more common of these varying ranges: between 2.0 and 3.0 for most patients and from 2.5 to 3.5 for patients with mechanical valves. The range was selected by a technical expert panel to represent a conservative estimate for an event where there is no single standard, particularly with respect to the higher end where a therapeutic range can be as high as 5.0. The numerator examines whether the INR monitoring has occurred and does not require a numeric INR value. All data required to calculate the measure are obtained through a mix of administrative claims and EHR data. Feasibility tests demonstrated that all required data elements were found to be available in the EHR systems tested. Providers are not required to conduct medical record abstraction.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for endorsement.

0097- MEDICATION RECONCILIATION POST-DISCHARGE

This measure received six comments. These comments expressed support for the measure and recognized its importance in improving patient safety, but there were a few issues raised. There were concerns that the use of CPT II codes would make it challenging for providers to report this measure. Further, the measure excludes professionals that commonly perform reconciliation in primary care settings. One comment stated that the measure should not be used on the provider level as discharge information is often not communicated in a consistent manner. One comment mentioned the measure may be burdensome because it requires chart abstraction and another recommended that nursing home and assisted living patients be removed from the denominator.

Developer Response: This measure encourages team based care by allowing medication reconciliation to be conducted by a variety of professionals including any prescribing practitioner, clinical pharmacist or registered nurse. NCQA's advisory panels felt that additional professionals in the office such as a nurse's assistant would not have sufficient clinical knowledge to conduct reconciliation. This approach aligns with successful transitional care models, such as those designed by Eric Colman that suggest medication reconciliation be conducted by a registered nurse. The developers recognize the limitation that in some EHRs medication reconciliation may be a checkbox. As with any quality measure collected in the EHR, it is possible providers may document processes they are not conducting. However, given the low performance on this measure the developers do not believe this is a widespread problem. This measure continues to highlight a significant quality gap. The developers also recognize the challenges that providers face in communicating with hospitals about discharge, however they believe measures of care coordination should drive providers and health care systems to improve communication and thus improve care for the patient. The developers also understand the burden this measure places on health plans for those who choose to report through the hybrid methodology, health plans do have the option of reporting this measure administratively through the use of three billing codes. Currently, only 5% of health plans are choosing to report this measure

administratively. Furthermore, the provider level measure is restricted to patients who are seen by the provider within 30 days of discharge. Therefore patients who do not have a post-discharge follow-up with their provider are not included in the denominator of the provider level measure.

Committee Response: The Committee agreed the developer sufficiently addressed the concerns raised and those voiced in the public comments. However, some members reiterated their concern that the measure does not indicate that actual medications were reconciled in a way that is accurate and correct. Another remaining point of concern is that registered nurses are included as one of the professionals eligible to conduct medication reconciliation. Some members expressed that this task should be completed or authorized by a physician. The Committee re-voted on this measure and recommended the measure for continued endorsement.

0531- PATIENT SAFETY AND ADVERSE EVENTS COMPOSITE (PSI 90)

During the comment period, there were a total of 60 comments submitted on measure 0531 (PSI-90). The majority of comments were supportive, specifically those from individual patients, patient advocate groups, and payers. However, several comments noting concerns with PSI-90 were submitted, primarily by physicians and hospital groups. There were comments specifically around the harmonization of the reporting of central-line associated blood stream infections, which are also reported via NHSN data and endorsed under a separate NQF-endorsed measure. The comment suggested better measure alignment because the NHSN data may be more accurate as it is based on case-report rather than claims data. There were also concerns that some of the events that are captured in administrative claims and reported as adverse events may not be preventable due to limitations in claims data. These data do not suggest a cause for the adverse event, only that it was coded in the chart. Other concerns were raised over the validity of the measure, specifically noting that many of the underlying components of PSI-90 may not be valid, and some have high rates of misclassification when the claims data are compared to chart review.

There were also specific concerns about PSI-12: Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate (DVT), which is included in the PSI-90 composite. The comment asked the developer to consider excluding trauma patients from "hospital acquired" DVT. The rationale was that trauma patients are at high risk for DVT, even when aggressive preventative measures are taken. In addition, trauma centers are vigilant in the detection of DVT by routinely screening patients. As a result of patients being high-risk and aggressive screening, there are high rates of DVT due to early identification of calf vein thrombosis. This could result in unfairly penalizing trauma centers, as compared to other centers, which do not screen for DVT as aggressively.

Developer Response: After reviewing the comments, AHRQ is proposing the removal of one component, PSI 07, of the measure; a new title for the measure; and one change to the component measure, PSI 12, Perioperative Deep Vein Thrombosis and Pulmonary Embolism.

1. *PSI-90 has been modified to not include PSI 07 (Central Venous Catheter-Related Bloodstream Infection Rate) due to the comments and concerns around the NHSN measures which are competing with this component of PSI 90.* Users of the AHRQ QI software will now have a choice between using the full version of PSI 90, containing PSI 07, or the modified version of PSI 90, without PSI 07. For the purpose of endorsement considerations, AHRQ recommends that the Committee consider only PSI 90 without the inclusion of PSI 07. In addition, the new version of PSI 90 has been re-weighted appropriately. This directly addresses the comments raised during the comment period. In addition, the developer conducted an analysis of the impact of this they found that it would not negatively impact the reliability of this measure. Additional detail is provided in a detailed memo from the developer (Appendix B).

2) *The name of PSI 90 will change.* PSI 90, version 6.0, will be changed from Patient Safety Composite for Selected Indicators, to Patient Safety and Adverse Events Composite. The developer stated that this was done in response to comments that raised concerns over the preventability of some of the coded adverse events included in the measure. The developer noted that the name better reflects the fact that some of the component indicators capture adverse events occurring during hospital care, and there is room for discussion and disagreement about the exact percentage of those events that are preventable given current knowledge.

3) *The definition of PSI 12 (Perioperative Deep Vein Thrombosis and Pulmonary Embolus) – a component of PSI 90 -- will now exclude patients with any diagnosis of major cranial and spinal trauma from the denominator.* While the public comment suggested excluding all trauma patients, the developer reasoned that exclusion specifically of major cranial and spinal trauma was reasonable because it may not be safe for physicians to prescribe thromboprophylaxis in these patients because of the increased risk of bleeding and potential catastrophic consequences of that bleeding. In addition, the developer noted that patients with major cranial and spinal trauma are clustered at major trauma centers. Initial analysis revealed that there would be no changes to the reliability and validity of the measure based upon this change.

Committee Response: The Committee agreed the developer's response sufficiently address the concerns raised and those voiced in the public comments. They commended the developers on the great level of effort taken to improve the measure. The Committee discussed the appropriateness of claims data for use in this kind of measure. One member voiced concerns about whether the measure demonstrates an adequate degree of validity. The Committee re-voted on this measure and recommended the measure for continued endorsement.

0352- FAILURE TO RESCUE IN-HOSPITAL MORTALITY (RISK ADJUSTED)

This measure received four comments. Each comment was in favor of the Committee's initial decision to defer the measure until more information is provided. One comment stated that the measure should not be endorsed because, for provider level measurement, the values would be very low. Another comment stated that failure to rescue does not always result in death and the measure may be too general.

Developer Response: The developers have shown in the Measure Testing form that their risk adjustment models are valid and reliable for the index population. It must be remembered that (1) surgeons did decide to perform surgery; (2) they are asking whether the patient survives a complication, NOT whether they develop a complication, and (3) the group of patients who develop a complication are far sicker than the general population of patients undergoing surgery. The developers will consider incorporating other data elements in future versions of the FTR measure, as they do when new data become available from literature or coding systems, but considering the strong reliability of the measure and predictive ability of the current risk-adjustment model, the developers do not believe these minor changes would merit changing the entire algorithm at this point, and they have no evidence that the changes suggested would in any substantive way change the ranking or rating of hospitals.

Committee Response: The Committee agreed the evidence is strong and there is a performance gap. There were questions about the types of test-retest that was used for reliability and the dataset used for the analysis. There were also concerns that the developer used a Medicare dataset for validity testing while the measure also applies to individuals under the age of 65 years. The developer noted that those who use this measure will need to risk adjust for a younger population. The developer stated that the measure utilizes data from Medicare claims which makes it feasible to implement and the Committee agreed. One member of the Committee expressed a concern that the measure was not currently in broad use despite having been endorsed several times by NQF. The Committee reviewed the response submitted by the developer (Appendix C), voted on this measure and recommended the measure for continued endorsement, agreeing that the measure met the criteria for NQF endorsement.

0353- FAILURE TO RESCUE 30-DAY MORTALITY (RISK ADJUSTED)

This measure received two comments. They were in favor of the Committee's initial decision to defer the measure until more information is provided. One comment stated that current risk methodology does not adequately account for risk of patients with cancer.

Developer Response: The developers have shown in the Measure Testing form that the risk adjustment models are valid and reliable for the index population. The developers note that (1) surgeons did decide to perform surgery; (2) they are asking whether the patient survives a

complication, NOT whether they develop a complication and (3) the group of patients who develop a complication are far sicker than the general population of patients undergoing surgery. The developers will consider incorporating other data elements in future versions of the FTR measure, as they do when new data become available from literature or coding systems, but, considering the strong reliability of the measure and predictive ability of the current risk-adjustment model, they do not believe these minor changes would merit changing the entire algorithm at this point, and they have no evidence that the changes suggested would in any substantive way change the ranking or rating of hospitals.

Committee Response: The Committee agreed the developer sufficiently addressed the concerns raised and those of voiced in the public comments. There is a great deal of similarity between this measure and 0352. Many of the questions that arose during the discussion of 0352 covered concerns about this measure. There was one question related to the reliability testing which was not done using an approach that is commonly by measure developers; however, the developer explained the methodology and the Committee was reassured of the reliability results. One Committee member requested clarification on whether the measure was tested on one large sample or multiple smaller samples. Another Committee member requested that patients who are over the age of 90 be included in the denominator. The Committee reviewed the response submitted by the developer (Appendix D), voted on this measure and recommended the measure for continued endorsement, agreeing that the measure met the criteria for NQF endorsement.

Ad Hoc Reviews

0138- NATIONAL HEALTHCARE SAFETY NEXTWORK (NHSN) CATHETER-ASSOCIATED URINARY TRACT INFECTION (CAUTI) OUTCOME MEASURE

This measure received six comments. Most comments were in support of the Committee's approval of the changes made. Another comment stressed that the measure should not be applied to the spinal cord injury (SCI) population, there needs to be meaningful monitoring of unintended adverse consequences, and the Committee should align its decision with a previous decision on a 2010 Nursing Home measure, in which the Committee decided that patients with neurogenic bladder should be exempt due to concerns over their safety.

Developer Response: These are important concerns about indiscriminate removal of indwelling urinary catheters from patients with spinal cord injuries (SCIs) treated in non-specialty hospitals. While the frequency and extent of this problem are not known, the developers agree that concerted efforts are warranted to close performance gaps and protect at-risk SCI patients. To that end, in January 2015 CDC proposed in a letter to the President of the American Spinal Cord Injury Association (ASIA) a collaborative ASIA-CDC initiative aimed at promoting safe and appropriate use of indwelling urinary catheters in the SCI patient population, particularly at non-specialty hospitals. That offer still stands and could include joint development and testing of a clinical quality measure of bladder function management of SCI patients. The CDC Healthcare Infection Control Practices Advisory Committee (HICPAC) Guideline for Prevention of Catheter-associated Urinary Tract Infections (CAUTI) includes a recommendation for use of intermittent urinary catheterization preferentially over indwelling urinary catheters in patients with bladder emptying dysfunction. This specific recommendation refers to patients with impaired bladder function, not all of whom are SCI patients, and the recommendation should be placed in the context in which it is presented in the guidelines, namely that practitioners should "consider using alternatives to indwelling urethral catheterization in selected patients when appropriate." The HICPAC guideline specifically recommends consideration of alternatives to chronic indwelling catheters, such as intermittent catheterization, in SCI patients, but the guidelines do not strongly recommend use of alternatives for these patients. Lastly, this measure is intended to be used in inpatient locations and facilities as it uses urinary catheter days as the denominator for calculating the standardized infection ratio. The data generated by the measurement may be useful by health plans in their assessment of quality of care.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for continued endorsement.

0345-UNRECOGNIZED ABDOMINOPELVIC ACCIDENTAL PUNCTURE OR LACERATION RATE (PSI15)

This measure received two comments. One comment suggested that PSI15 should not be recorded if the "injury" was minor and had no subsequent consequence and that it should not be recorded if the laceration or puncture was due to the following:

- Infection/inflammation
- Cancer
- Adhesions
- Radiation damage

Developer Response: It is true that cancer patients may be at higher risk of PSI 15 (Unrecognized Accidental Abdominopelvic Puncture or Laceration) than patients without cancer, but this difference is accounted for in AHRQ's risk-adjustment model (http://qualityindicators.ahrq.gov/Downloads/Modules/PSI/V50/Parameter_Estimates_PSI_50pdf). For example, MDC 17 (Myeloproliferative Diseases and Disorders and Poorly Differentiated Neoplasms) is associated with 2.94 times higher adjusted odds of PSI 15. Some MS DRGs within MDC 17, such as 820-822 (MDRG 1707, Leukemia and Lymphoma with Major OR Procedure), 826-828 (MDRG 1709, Myeloproliferative Disorder or Poorly Differentiated Neoplasm with Major OR Procedure), 303 (MDRG 1103, Kidney and Ureter Procedures for Neoplasm), and 357 (MDRG 1302, Uterine and Adnexa Procedure for Ovarian or Adnexal Malignancy) are associated with even higher adjusted odds of 66-144. This risk-adjustment model has very high discrimination of c=0.921, indicating that it assigns a higher probability of PSI 15 to patients who actually experienced the event (among randomly selected pairs) 92.1% of the time.

The comment is related to Version 5 specification of PSI 15, when the Version 6 specification is now under review by NQF. Inconsequential or "minor" events are no longer included, because a second operation (at least one day after the first operation) is now required to trigger the numerator of PSI 15. The adjective "unrecognized" is proposed for the title of PSI 15 because return to the operating room for repair of an "accidental puncture or laceration" after abdominopelvic surgery implies that the injury was not recognized when it occurred (or else it would have been repaired at that time), or that the initial repair failed. Although AHRQ has not implemented "automatic exclusions" for infection, inflammation, adhesions, or radiation damage, most of these factors are included in the risk-adjustment model. In addition, the American College of Surgeons' bulletin highlights that "according to explicit guidance from the [American Hospital Association's] Coding Clinic for ICD-9-CM (Second Quarter, 2007 and First Quarter, 2010), 'expected' enterotomies are not coded with code 998.2. By definition, this code is limited to 'accidental' punctures and lacerations that are not 'intrinsic' or 'inherent' in a major procedure. Although (this) guidance is straightforward, the ACS has received comments from Fellows indicating that some hospital quality reporting departments continue to misunderstand how to correctly report PSI-15. This column provides more background and coding guidance to assist surgeons in working with their hospital staff on reporting PSI-15."" AHRQ supports efforts of this type to improve coding practice and promote dialogue between surgeons and coding professionals. For another example of these efforts, see Utter GH et al. in JAMA Surg. 2015 May; 150(5):388-9.

Committee Response: The Committee agrees with the developer response and maintains their decision to recommend this measure for continued endorsement.

NQF Member Voting

Information for electronic voting has been sent to NQF Member organization primary contacts. Accompanying comments must be submitted via the online voting tool.

Please note that voting concludes on November 4th, 2015 at 6:00 pm ET – no exceptions.

Appendix A: 0097- Medication Reconciliation Memo

TO: National Quality Forum Patient Safety Steering Committee

FROM: Erin Giovannetti and Bob Rehm, NCQA

DATE: September 10, 2015

RE: Endorsement Maintenance of NQF #0097 Medication Reconciliation Post-Discharge

NCQA appreciates the opportunity to clarify how the Medication Reconciliation Post-Discharge measure is collected and the validity data presented on the measure.

This measure is collected through either medical record review or administrative claims data. We worked closely with medical record reviewers and measure auditors to identify the minimum set of documentation necessary in the medical record to demonstrate that a discharge medication list was reconciled with an outpatient medication list. Specifically the measure specification instructs the medical record reviewer to look for the following documentation in the medical chart notes:

- Documentation in the medical record notes that the provider reconciled the current and discharge medications.
- Documentation of the current medications with a notation that references the discharge medications (e.g., no changes in medications since discharge, same medications at discharge, discontinue all discharge medications).
- Documentation of the member's current medications with a notation that the discharge medications were reviewed.
- Documentation of a current medication list, a discharge medication list and notation that both lists were reviewed on the same date of service.
- Notation in the medical record notes that no medications were prescribed or ordered upon discharge.

The administrative claims that can be used to satisfy this measure include administrative billing codes for transitional care visits that include medication decision making post discharge and a claims code specific to medication reconciliation.

- CPT 99495: Transitional care management services with the following required elements: (1) communication (direct contact, telephone, electronic) with the patient and/or caregiver within 2 business days of discharge, (2) medical decision making of at least moderate complexity during the service period and (3) face-to-face visit, within 14 calendar days of discharge.
- CPT 99496: Transitional care management services with the following required elements: (1) communication (direct contact, telephone, electronic) with the patient and/or caregiver within 2 business days of discharge, (2) medical decision making of high complexity during the service period and (3) face-to-face visit, within 7 calendar days of discharge.
- CPT Category II 1111F: Discharge medications reconciled with the current medication list in the outpatient medication record

After much discussion with our Measurement Advisory Panel, we determined this is the best approach to measuring a process that is not routinely documented. However, we appreciate the Steering Committee's comments that some providers may be conducting medication reconciliation and not noting anywhere in the chart that the discharge medication list was reviewed. As with all quality measures, we are limited to measuring what is documented. Therefore we propose revising the name of the measure to "*Documentation of Medication Reconciliation Post-Discharge."* This name will match the description of the measure which already specifies the measure is looking for documentation of reconciliation.

In our measure submission we presented testing data demonstrating that two medical record reviewers when looking at the same gold-standard medical chart had high agreement as to whether there was documentation of medication reconciliation. We also presented testing demonstrating the overall construct validity of the measure score, as it is highly correlated with another measure of medication management. We presented data showing that performance scores could demonstrate meaningful difference in performance as demonstrated by a statistically significant t-test between health plans at the 25th and 75th percentile. Finally, we presented the initial face validity panel review results, which indicated high agreement among panel members that the measure score reflects quality. During the Steering Committee meeting we were additionally asked if the same panel would continue to support the face validity of the measure. In 2014 we re-evaluated this measure with our Geriatric Measurement Advisory Panel and Committee on Performance Measurement. Both panels unanimously supported the measure's face validity and its expansion from Special Needs Plans to all of Medicare Advantage. We believe these findings together demonstrate the validity of the measure.

Finally, we appreciate the Steering Committee's comments that this measure does not fully address the range of activities that constitute a high quality medication reconciliation. We recognize these limitations of the measure. Many elements of medication reconciliation are not documented such as assessing patient knowledge of their medications, providing education to patients about their medications, assessing the indication of each medication and determining where medications can be eliminated, and coordinating medication lists across providers. These elements of medication reconciliation may be better addressed through patient reported measures, such as the items in the Medicare CAHPS survey assessing knowledge of medication and education of providers and care teams about medication reconciliation.

While this measure does not capture the full array of elements necessary to conduct high quality medication reconciliation, it captures a minimum bar for health plans and providers that can be feasibly assessed in a medical record. Given the low performance on this measure across health plans (37% of discharged patient records showing documentation of medication reconciliation) it is clear there is still room for improvement.

Appendix B: 0531- PSI 90, Patient Safety Indicator Composite for Selected Conditions Memos

Memo
To: Suzanne Theberge, National Quality Forum
From: AHRQ
Re: Proposed Additional Changes to PSI 90, Patient Safety Indicator Composite for Selected Conditions
Date: September 16, 2015

Dear Ms. Theberge,

Following careful review of the Patient Safety Committee comments and public comments regarding PSI 07, the AHRQ is proposing one additional change to PSI 90 and one change to the component measure, PSI 12, Perioperative Deep Vein Thrombosis and Pulmonary Embolism. We believe that these changes address two specific concerns raised for PSI 90, namely the conceptual overlap of PSI 07 with CDC NHSN measures of CLABSI and the inclusion of cases where prophylaxis is contraindicated in the component measure PSI 12. We outline the changes below.

Modified Version without PSI 07

AHRQ supports the goal of measure alignment across programs and service lines. For this reason, AHRQ has developed an alternative specification of PSI 90 that gives zero weight to PSI 07 (Central Venous Catheter-Related Bloodstream Infection Rate) and redistributes that weight proportionately across all of the other component measures in PSI 90. Users of the AHRQ QI software will now have a choice between using the full version of PSI 90, containing PSI 07, or the modified version of PSI 90, without PSI 07. However, for the purpose of endorsement considerations, AHRQ recommends that consider only PSI 90 without the inclusion of PSI 07.

A modified version without PSI 07 is being offered because PSI 07 overlaps with the NHSN measure of Central Line Associated Bloodstream Infection. Because the National Healthcare Safety Network (NHSN) measure is required and utilized in many of the same programs as PSI 90, an indicator without PSI 07 will address potential overlap in indicators used in the same program. Further, concern was raised regarding the usefulness of administrative data to capture these events. Although AHRQ believes that recent coding changes are likely to have vastly improved the usefulness of

administrative data in capturing such events, further validation work is required before the measure can be definitively assessed. AHRQ prefers to remove PSI 07 from the NQF-endorsed specification of PSI 90 until such work is completed.

Analysis of the unique contribution of component indicators to the reliability of PSI 90 suggests that the removal of PSI 07 is not expected to negatively impact the reliability of the measure. Specifically, the Spearman rank order correlation between PSI 90 scores with and without PSI 07 is 0.993, with a kappa score of 0.950. Out of 3,171 hospitals in the AHRQ 36-state all-payer reference population, only 210 (6.2%) shift across ranked performance quartiles when PSI 07 is zero weighted, and only two of those hospitals (<0.1%) shift more than one quartile. Further analytic work on the performance of PSI 90 with zero weighting of PSI 07 will be submitted to NQF by September 22, but preliminary analyses indicate no significant changes to reliability or validity.
The full model of PSI 90, that includes PSI 07, will continue to be available to users in the AHRQ QI software as a non-NQF endorsed measure. PSI 07 in the full model will continue to be risk-adjusted, accounting for 14 comorbidities, transfer status, 16 Major Diagnostic Categories, and nearly 60 Modified DRGs. This model has very high discrimination of c=0.891, indicating that it assigns a higher probability of PSI 07 to patients who actually experienced the event - among randomly selected pairs - 89.1% of the time), and remains a viable indicator for use.

Note: Based on feedback from NQF, the name for PSI 90, version 6.0, has been changed from Patient Safety Composite for Selected Indicators, to Patient Safety and Adverse Events Composite. This name better reflects the fact that some of the component indicators capture adverse events occurring during hospital care, and there is room for discussion and disagreement about the exact percentage of those events that are preventable given current knowledge. However, as AHRQ has recently reported, "Preliminary estimates for 2013 show a further 9 percent decline in the rate of hospital-acquired conditions (HACs) from 2012 to 2013, and a 17 percent decline, from 145 to 121 HACs per 1,000 discharges, from 2010 to 2013. A cumulative total of 1.3 million fewer HACs were experienced by hospital patients over the 3 years relative to the number of HACs that would have occurred if rates had remained steady at the 2010 level. We estimate that approximately 50,000 fewer patients died in the hospital as a result of the reduction in HACs, and approximately $12 billion in health care costs were saved from 2010 to 2013. Although the precise causes of the decline in patient harm are not fully understood, the increase in safety has occurred during a period of concerted attention by hospitals throughout the country to reduce adverse events, spurred in part by Medicare payment incentives and catalyzed by the U.S. Department of Health and Human Services (HHS) Partnership for Patients…" The great majority of the measures used in that report were from the Medicare Patient Safety Monitoring System and the National Healthcare Safety Network (and are thus immune to any efforts to manipulate PSI reporting), but PSI 90 is an integral component of HHS' efforts to promote continuing improvements in patient safety and adverse events.

PSI 12 Perioperative Deep Vein Thrombosis and Pulmonary Embolus

Following public comments, AHRQ will be implementing an additional change in Version 6.0 of PSI 12. This new change will exclude patients with any diagnosis of major cranial and spinal trauma from the denominator. These conditions are excluded because it may not be safe for physicians to prescribe thromboprophylaxis in this clinical setting (due to the increased risk of bleeding and potential catastrophic consequences of that bleeding). This change is targeted at excluding events that may be less preventable and that are beyond the control of the provider. This clinical scenario is uncommon overall, but patients with major cranial and spinal trauma are clustered at major trauma centers. Further analytic work on the performance of PSI 12 and PSI 90 with this specification change will be submitted to NQF by September 22, but preliminary analyses indicate no significant changes to reliability or validity. Because less than 0.1% of patients in the PSI 12 denominator qualify for this exclusion, AHRQ does not consider this to be a material change requiring ad hoc review of PSI 12 by NQF.

To: Suzanne Theberge, National Quality Forum

Re: Empirical Performance, PSI 90 without PSI 07, Modified Patient Safety Indicator Composite for Selected Conditions (Renamed Patient Safety and Adverse Events Composite)

Date: September 23, 2015

As stated in our September 16, 2015 memo, based on feedback during the NQF Consensus Development Process, AHRQ has developed and tested an alternative specification of PSI 90 (hereafter referred to as the modified version of PSI 90) that omits PSI 07 (Central Venous Catheter-Related Bloodstream Infection Rate) from the component indicators and redistributes that weight proportionately across all of the other component measures in PSI 90. Starting in Version 6.0, to be released in early 2016, users of the AHRQ QI software will have a choice between using the full version of PSI 90, containing PSI 07, or this modified version

of PSI 90, without PSI 07. The full model of PSI 90, including PSI 07, will continue to be available to users in the AHRQ QI software as a non-NQF endorsed measure.

AHRQ believes that central line associated bloodstream infections are an important component of overall patient safety, and will continue to examine the method by which central line blood stream infections can be best measured and integrated into an overall measure of patient safety in the future. Because validation work has not yet been completed on improvements to coding of central line blood stream infection (i.e., adoption of a very specific ICD-9-CM code 999.32 – bloodstream infection due to central venous catheter – in 2011), and how PSI 07 compares with NHSN's measure of central line associated blood stream infection, AHRQ is currently requesting NQF endorsement of the revised specification of PSI 90 with zero weight for PSI 07.

New Weights for PSI 90 without PSI 07

The new weighted average uses the same methodology as described in the original NQF submission. It is a weighted average of the reliability-adjusted, indirectly standardized, observed-to-expected ratios for the component indicators. Removing PSI 07, Central Venous Catheter-Related Blood Stream Infection Rate, leaves the following components in the PSI 90 composite: PSI 03 Pressure Ulcer Rate, PSI06 Iatrogenic Pneumothorax Rate, PSI 08 Postoperative Hip Fracture Rate, PSI09 Postoperative Hemorrhage or Hematoma, PSI10 Postoperative Physiologic and Metabolic Derangement, PSI 11 Postoperative Respiratory Failure, PSI 12 Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate, PSI 13 Postoperative Sepsis Rate, PSI 14 Postoperative Wound Dehiscence Rate, and PSI 15 Accidental Puncture or Laceration Rate.

The composite measure is a weighted average of the smoothed rates of the component indicators. The final weight for each component is the product of harm weights and volume weights (numerator weights). Harm weights are calculated by multiplying empirical estimates of excess harms associated with the patient safety event by utility weights linked to each of the harms. Excess harms are estimated using statistical models comparing patients with a safety-related event to those without that safety-related event in a CMS Medicare fee-for-service sample that allowed up to one year of follow-up from the discharge date of the hospital stay associated with the index event. Volume weights, the second part of the final weight, are calculated on the basis of the number of safety-related events for the component indicators in the all-payer reference population.

Population rate and distribution

The population rates and distribution of hospital performance between the original PSI 90 and modified version of PSI 90 (without PSI 07) is similar (Table 1). There is slightly more variation in the modified version (evidenced by the slightly wider range and higher standard deviation) than in the original PSI 90, however the differences between mean and median are minimal.

Table 1. Reference Population Rate and Distribution of Hospital Performance PSI 90 Patient Safety Composite for Selected Indicators Distribution of Hospital-level Observed Rates in Reference Population for FY 2012 between PSI 90 without PSI 07 compared to PSI 90 with PSI 07.[1]

				Percentile				
Version	**Number of Hospitals**	**Mean**	**SD**	**5th**	**25th**	**Median**	**75th**	**95th**
Modified Version of PSI 90 Without PSI 07	3,380	0.980	0.146	0.775	0.900	0.979	1.023	1.240
Original PSI 90 With PSI 07	3,380	0.981	0.127	0.796	0.911	0.982	1.021	1.206

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

[1] The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals included in the dataset, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile.

Reliability

Using the same reference population (2012 data), the average hospital signal-to-noise ratio (SNR) for the overall reweighted modified version of PSI 90 without PSI 07 using all-payer discharge data is 0.768, only slightly lower than the corresponding estimate for PSI 90 with PSI 07 (0.785). The reliability of the indicator remains strong (SNR = 0.768) when PSI 07 is removed.

Table 2. Signal-to-Noise Ratio by Size Decile: PSI 90 without PSI 07

Size Decile	Number of Hospitals	Ave. Number of Discharges per Hospital in Decile	Ave. Signal-to-Noise Ratio for Hospitals in Decile
1	321	74	0.539
2	322	130	0.554
3	321	266	0.572
4	322	556	0.596
5	321	1,099	0.624
6	322	1,490	0.650
7	322	1,327	0.690
8	321	2,043	0.738
9	322	3,067	0.795
10	321	5,794	0.875
Overall	**3,215**	**1,584**	**0.768**

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

Performance Discrimination

Similar to previously presented evidence of PSI 90 including PSI 07, the modified version of PSI 90 without PSI 07 can detect statistically meaningful differences across hospitals in the upper 5-7 deciles of the hospital volume distribution. The number of hospitals that could not be statistically classified as better or worse than the benchmark (i.e., the value that separates the bottom 80% of hospitals from the upper 20%) or the threshold (i.e., the value that separates the bottom 20% of hospitals from the upper 80%) remains stable as shown in Table 3.

Table 3: Performance Categories by Hospital Size Decile for PSI 90 without PSI 07

			Benchmark (80th percentile)			Threshold (20th percentile)		
Size Decile	Number of Hospitals	Average Number of Patients Per Hospital	Proportion Better	Proportion Worse	Proportion Unclassified[1]	Proportion Better	Proportion Worse	Proportion Unclassified[1]
1	338	20	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000
2	338	82	0.0000	0.0063	0.9937	0.0063	0.0000	0.9937
3	338	192	0.0000	0.0159	0.9841	0.0540	0.0000	0.9460
4	338	374	0.0000	0.0280	0.9720	0.2857	0.0000	0.7143
5	338	660	0.0000	0.1038	0.8962	0.3522	0.0000	0.6478
6	338	1,040	0.0000	0.2215	0.7785	0.4000	0.0123	0.5877
7	338	1,547	0.0123	0.2708	0.7169	0.4123	0.0123	0.5754
8	338	2,256	0.0124	0.3137	0.6739	0.4969	0.0311	0.4720
9	338	3,269	0.0183	0.4329	0.5488	0.4756	0.0610	0.4634

10	338	6,165	0.0334	0.6018	0.3647	0.4590	0.0942	0.4468
Overall	**3,380**	**1,561**	**0.0078**	**0.2022**	**0.7900**	**0.2967**	**0.0215**	**0.6818**

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

[1]Proportion Unclassified refers to Neither Better Nor Worse.

Construct Validity

Removing PSI 07 also does not materially affect the classification of hospital performance. Specifically, the Spearman rank order correlation between hospital-level PSI 90 scores with (original) and without (modified version) PSI 07 is 0.993, with a kappa score of 0.950. Out of 3,171 hospitals in the AHRQ 36-state all-payer reference population, only 210 (6.2%) shift across ranked performance quartiles when PSI 07 is zero weighted, and only two of those hospitals (<0.1%) shift more than one quartile. These estimates indicate a very high level of agreement between PSI 90 with and without PSI 07.

Summary

In summary, to address the concerns of committee reviewers and to avoid confusion between two widely used measures of healthcare associated infections, AHRQ has developed and tested an alternative specification of PSI 90 that places zero weight on PSI 07 (central line associated bloodstream infection rate), effectively removing the indicator. This alternative specification demonstrates distributional properties, signal-to-noise reliability, and construct validity essentially equivalent to those of the previously submitted specification of PSI 90 (with nonzero weight on PSI 07).

Attachment 1: Empirical Performance of Original PSI 90 (with PSI 07)

The following information was provided in the Composite Measure Testing Form submitted to the NQF Patient Safety Committee and reviewed in June 2015. This information is provided for reference in interpreting the empirical performance of the modified PSI 90 (without PSI 07) presented in the body of this memo. Table 1a provides the 2011 and 2012 hospital rate distribution in the AHRQ QI Reference Population.

Table 1a. Reference Population Rate and Distribution of Hospital Performance for Original PSI90 (with PSI 07)

Distribution of Hospital-level Observed Rates in Reference Population								
Year	Number of Hospitals	Distribution of Observed Hospital-level Rates per 1000 (p=percentile)[1]						
		Mean	SD	p5	p25	Median	p75	p95
2012	3380	0.981	0.1271	0.7961	0.9112	0.9818	1.0209	1.2064
2011	3214	1.003	0.1359	0.8166	0.9309	0.9908	1.0459	1.2568

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2011-2012. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0 alpha)

[1]The distribution of hospital rates reports the mean and standard deviation (SD) of the observed rates for all hospitals included in the dataset, as well as the observed rate for hospitals in the 5th, 25th, 50th (median), 75th, and 95th percentile.

Reliability

PSI90 has strong reliability as measured by signal-to-noise ratios and intraclass correlation of split samples. The reliability was sufficient across all size deciles, although reliability increases with hospital size.

Table 2a provides the signal-to-noise ratio by hospital decile. The signal-to-noise ratio refers to the entire population of US hospitals, comparing the degree to which rates are different from hospital to hospital (the signal) to how stable the rates are within hospitals (the noise). This metric is a stringent measure of reliability that takes into account the observed distribution of rates within a reference population.

Table 2a. Signal-to-Noise Ratio by Size Decile for Original PSI90 (with PSI 07)

Size Decile	Number of Hospitals	Ave. Number of Discharges per Hospital in Decile	Ave. Signal-to-Noise Ratio for Hospitals in Decile
1	321	91.7	0.54608
2	322	136.7	0.55549
3	321	290.7	0.57013
4	322	556	0.59319
5	321	932.4	0.61971
6	322	1347.4	0.64668
7	322	2165.3	0.67874
8	321	2911.5	0.71418
9	322	3790	0.76308
10	321	6791.1	0.84219
Overall	**3215**	**1900.8**	**0.74845**

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

Performance Discrimination

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).

Table 3b reports the proportion of hospitals above and below the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of hospitals not classified as either better or worse have rates that fall within the 95% confidence interval.

Table 3b. Performance Categories by Hospital Size Decile Original PSI90 (with PSI 07)

			Benchmark			Threshold		
Size Decile	Number of Hospitals	Average Number of Denominator Discharges Per Hospital	Proportion Better	Proportion Worse	Proportion Unclassified	Proportion Better	Proportion Worse	Proportion Unclassified
1	338	30	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000
2	338	110	0.0000	0.0064	0.9936	0.0000	0.0000	1.0000
3	338	249	0.0000	0.0095	0.9905	0.0316	0.0000	0.9684
4	338	464	0.0000	0.0248	0.9752	0.2267	0.0000	0.7733
5	338	816	0.0000	0.0969	0.9031	0.3438	0.0000	0.6563
6	338	1,281	0.0031	0.1569	0.8400	0.3723	0.0092	0.6185
7	338	1,890	0.0062	0.2438	0.7500	0.4660	0.0031	0.5309
8	338	2,749	0.0124	0.2950	0.6925	0.5280	0.0248	0.4472
9	338	3,931	0.0182	0.4012	0.5805	0.4742	0.0547	0.4711
10	338	7,183	0.0242	0.6182	0.3576	0.4182	0.0970	0.4848
Overall	**3,380**	**1,870**	**0.0065**	**0.1882**	**0.8053**	**0.2890**	**0.0193**	**0.6918**

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

The indicator can detect statistically meaningful differences across hospitals. The number of statistically significant outliers is limited by the relatively tight distribution of performance. In other words, when 11 separate indicators are combined into a single composite, hospitals that are random outliers on one or two of the component indicators are no longer outliers on the overall composite. These findings are expected but they do suggest that, for low-volume hospitals, PSI 90 should be used cautiously – or multiple years of data should be combined.

To: Suzanne Theberge, National Quality Forum

From: Agency for Healthcare Research and Quality and Contractors (Patrick S. Romano, MD MPH on behalf of Stanford University and its subcontractors)

Re: Additional Responses to Reviewer and Public Comments Regarding the Preventability and Surveillance of components of PSI 90 Patient Safety and Adverse Events Composite (NQF 0531) and Component Measures

Date: September 25, 2015

AHRQ would like to address concerns verbally expressed by the NQF Patient Safety Standing Committee reviewers and during the public comment period regarding the level of preventability and impact of surveillance on risk-adjusted outcome measures, such as AHRQ's Patient Safety and Adverse Events Composite (NQF 0531, PSI 90). The topic of preventability and surveillance has been a recurring topic of discussion in NQF reviews, academic critiques, and media coverage of the Patient Safety and Adverse Events Composite (NQF 0531, PSI 90). In particular, NQF reviewers and public commenters expressed concern about whether limited preventability and variations in surveillance practices may invalidate one component of the Patient Safety and Adverse Events Composite (PSI 90, NQF 0531): PSI 12, Perioperative Pulmonary Embolism (PE) or Deep Vein Thrombosis (DVT). In this memo, we summarize several lines of evidence that support viewing PSI 12 as a potentially preventable event that is suitable for inclusion in the Patient Safety and Adverse Events Composite. This information was not provided in the original PSI 90 submission, because PSI 12 is an endorsed measure that previously passed the evidence requirements for endorsement under the Patient Safety Standing Committee. In the future, AHRQ will provide this same information in the NQF evidence form for PSI 12, when PSI 12 is considered for maintenance of endorsement.

Evidence from clinical trials on preventability

There is very strong evidence from randomized controlled trials that pharmacologic and mechanical prophylaxis are both effective at reducing the incidence of perioperative venous thromboembolism (VTE) among appropriately selected patients. The American College of Chest Physicians (ACCP)'s clinical practice guideline *Approach to Outcome Measurement in the Prevention of Thrombosis in Surgical and Medical Patients: Antithrombotic Therapy and Prevention of Thrombosis* (2012) states that:

"Although some studies have limitations of lack of concealment and blinding, evidence from meta- analyses of randomized controlled trials (RCTs) strongly suggests that prophylaxis with an anticoagulant or aspirin reduces symptomatic VTE and fatal PE in medical and surgical patients. In patients undergoing orthopedic, general, or urological surgery, unfractionated heparin (UFH) reduces the risk of fatal PE by about two-thirds; in patients undergoing hip or knee arthroplasty or hip fracture surgery, vitamin K antagonists (VKAs) reduce the risk of symptomatic VTE by about four-fifths; in patients undergoing hip or knee arthroplasty, extended-duration low-molecular-weight heparin (LMWH) or warfarin reduces the risk of symptomatic VTE by about three-fifths; in medical patients at highest risk, UFH, LMWH, danaparoid, or fondaparinux reduces the risk of PE by about two- to three-fifths; and in patients undergoing abdominal or pelvic surgery, LMWH reduces the risk of symptomatic VTE by about four- fifths. Antiplatelet therapy also is effective for the prevention of VTE in the highest-risk surgical or medical patients, reducing the risk of PE by about one-half and DVT by about three-fifths. Similar relative risk reductions are seen in trials comparing the efficacy of anticoagulant prophylaxis with placebo or no treatment based on a surrogate outcome; compared with placebo or no treatment, prophylactic anticoagulants reduce the relative incidence of silent DVT diagnosed through screening venography by 30% to 70%... Collectively, the meta-analysis data indicate that prophylactic anticoagulants are effective for the prevention of patient-important VTE and that the benefit-risk trade- off justifies their use in patients who are at sufficiently high risk of symptomatic VTE."[1]

Unfortunately, there is very limited evidence regarding the comparative effectiveness of different prophylactic modalities. However, the same ACCP guidelines note that "the compelling evidence of a decrease in fatal PE that exists for anticoagulants and for aspirin does not exist for mechanical methods."

Similar guidelines supporting the routine use of pharmacologic prophylaxis in selected populations have been published by several professional organizations:

1. For "all patients with malignant disease undergoing major surgical intervention," by the American Society of Clinical Oncology.[2]
2. For "patients undergoing elective hip and knee arthroplasty," by the American Academy of Orthopedic Surgeons.[3]
3. For "management of hip fractures in the elderly," by the American Academy of Orthopedic Surgeons.[4]
4. For "high risk patients undergoing gynecologic surgery" (defined as "surgery lasting less than 30 minutes in patients older than 60 years or with additional risk factors; major surgery in patients older than 40 years or with additional risk factors"), by the American College of Obstetricians and Gynecologists.[5]

Evidence of other modifiable risk factors suggesting preventability

Several studies suggest that delayed postoperative ambulation is an important, modifiable risk factor for postoperative VTE. For example, AHRQ recently supported a case-control study of patients undergoing total knee arthroplasty (TKA) in 15 teaching hospitals between October 2008 and March 2010. Cases were screened using PSI 12 and had objectively documented acute VTE within 9 days of surgery; controls were randomly selected TKA patients from the same hospital. Among 130 cases and 463 controls, **all** patients received thromboprophylaxis (pharmacologic in 80%, mechanical alone in 20%) but only 68% ambulated on day 1 or 2 after surgery. Factors significantly associated with VTE (after adjusting for age, sex, history of VTE, and BMI) were bilateral TKA (OR=4.2; 95% CI: 1.9-9.1), receipt of pharmacological prophylaxis (OR=0.5; 95% CI: 0.3-0.8), and ambulation by postoperative day 2 (OR=0.3; 95% CI: 0.1-0.9).[6] In an earlier case control study based on a sampling frame with 25,388 Medicare fee- for-service beneficiaries 65 years of age or older who underwent unilateral total hip arthroplasty (THA) in any nonfederal hospital in California between January 1993 and December 1996, White et al. compared processes of care between 297 randomly selected cases with VTE within 3 months after surgery and 592 randomly selected controls. Factors independently associated with VTE included initial ambulation before day 2 after surgery (OR=0.7; 95% CI 0.5–0.9), use of pneumatic compression (among patients with body-mass index <25; OR=0.3; 95% CI 0.2–0.6), and use of warfarin after discharge (OR=0.6; 95% CI 0.4–1.0).[7] These studies suggest a population fraction of post-arthroplasty VTE attributable to delayed ambulation of at least 10% and perhaps over 40%.

Two groups of authors have reported single-center results of prospectively implementing postoperative care protocols emphasizing early ambulation. Chandrasekaran et al. found that getting patients out of bed or walking for at least 15–30 minutes twice on the first day after TKA significantly reduced the odds of asymptomatic or symptomatic VTE (OR=0.35; 95% CI: 0.13-0.94) compared with the previous practice of confining patients to bed on that day.[8] Similarly, Pearse et al. implemented a treatment protocol that involved showering and walking up to 30 meters within 24 hours after TKA, and observed a substantial reduction in the odds of asymptomatic or symptomatic DVT (OR=0.04; 95% CI 0.004-0.30).[9] These findings are supported by several cohort studies summarized in a recent structured review.[10]

Evidence of gaps in processes of care consistent with preventability

AHRQ's Evidence-based Practice Review on Patient Safety recently summarized the state of the field as follows: "Even though high quality evidence exists for safe and effective strategies to reduce the risk of VTE, studies continue to show that many hospitalized patients are not given risk-appropriate VTE prophylaxis. One recent study across 32 countries found that only 59% of at-risk surgical and 40% of at- risk medical patients received guideline-recommended VTE prophylaxis, and a United States registry study found that only 42% of patients diagnosed with DVT during a hospitalization had received prophylaxis…"[11] Similar findings have been reported from Europe[12] and from 28 Veterans Health Administration hospitals (where "accounting for contraindications and early VTE occurrence, a total of 78% of cases [with PSI 12] and 80% of controls [without PSI 12] were appropriately managed").[13]

Several recent reports from Johns Hopkins Medical Institutions (JHMI) shed further light on this problem. In 2004, the JHMI Center for Innovations in Quality Patient Care assembled a multidisciplinary VTE-prevention

team to develop an education program for health-care providers, design evidence-based risk-appropriate prophylaxis strategies, establish mechanisms to assess performance, provide feedback to staff, and implement order sets or forms to guide clinicians through the risk-stratification process.

Among surgical patients, use of risk-appropriate VTE prophylaxis increased from 26% (42 of 161) at baseline to 68% (178 of 262) within 12 months, and to 85% after implementation of computer-based "smart order sets."[14] However, a retrospective review of 92 patients diagnosed with hospital-acquired VTE between July 2010 and June 2011 found that only 43 (47%) received defect-free care, while 49 (53%) had potentially preventable VTE: "13 (27%) were not prescribed risk-appropriate VTE prophylaxis, and 36 (73%) missed at least one dose of appropriately prescribed prophylaxis."[15] The same group reviewed use of VTE prophylaxis on their trauma service between July 2012 and June 2013. "Over half

of the residents (42 of 75 [56.0%]) prescribed optimal, risk-appropriate VTE prophylaxis for every patient whom they admitted, while 7 residents (9.3%) did not prescribe optimal prophylaxis to any patient.

There was no difference among the 8 attending physicians (median compliance rate, 74.2%; interquartile range, 72.6%-77.3%)," indicating that resident practice variation may be an important contributor to VTE events at teaching hospitals like JHMMI.[16]

One recent study from the Washington State Surgical Care and Outcomes Assessment Program demonstrated no significant decrease over time in VTE rates after colorectal surgery despite increased use of perioperative chemoprophylaxis, from 32% in 2006 to 86% in 2011, and postoperative prophylaxis, from 60% to 91%.[17] However, there is substantial countervailing evidence of continuing opportunities for improvement. For example, Ang et al. found that reporting real-time outcomes data on surgical outcomes in 2012 was associated with significant improvement in observed-to-expected (O/E) ratios for PSI 12 at the University of Florida.[18] Heslin et al. reported on the outcomes of a surgeon- led mortality, PSI, and hospital acquired conditions review involving 12 surgical services in a single institution; PSI 12 accounted for 25% of all events and the most common contributing factor was "failure to follow protocol."[19] Hussey et al. tested an alpha version of the AHRQ QI Toolkit in a one-year quality improvement initiative at an academic hospital with approximately 500 beds. The electronic medical record was revised so that deep vein thrombosis prophylaxis would be a mandatory part of the order set; observed PSI12 rates decreased from 20.7 pre-intervention to 15.9 post-intervention.[20] The University of California recently reported that a five-campus collaborative effort to improve VTE risk stratification and prophylaxis achieved a 23.8% relative reduction in the incidence of PSI 12 in 2014, relative to 2011, translating to 140 averted events in 2013 and 170 in 2014.[21]

This recent literature suggests that postoperative VTE is a partially preventable complication. Moreover, significant gaps in the processes of care are still being observed, and efforts to bridge these gaps appear to be "paying off" through substantial reductions in the local incidence of PSI 12. Indeed, AHRQ recently reported an 18% national reduction in the incidence of postoperative VTE between 2010 and 2013,

Based not on ICD-coded data, but on review of 18,000-33,000 medical records annually from the Medicare Patient Safety Monitoring System (MPSMS). [22] The MPSMS data come from a system in which a sample of Medicare inpatient records are reviewed by trained abstractors who use a structured protocol and software tool to determine whether any of 21 specific measures of adverse events occurred during the hospital stay, with high interrater reliability.[23] This 18% reduction, which is probably at least partially attributable to attention from CMS and other payers, translates to 5,000 averted events, 520 averted deaths, and a projected cost savings of $40 million in 2013.

Are hospitals with higher rates just doing better surveillance, suggesting lack of preventability?

Another concern raised by reviewers and the public regarding PSI 12 is that higher rates are a result of "increased vigilance in detection" at some hospitals.[24] Following this argument, high rates may be nonpreventable – even desirable – because perioperative PEs and DVTs are being diagnosed early (i.e., before symptoms develop) and treated aggressively at these "high surveillance" hospitals. Proponents of this argument cite Medicare claims data showing that "postoperative VTE imaging rates ranged from

> 85.26 per 1000 discharges in the lowest quartile of hospitals… to 168.86 in the highest quartile... drivers of high imaging rates at the 90th quantile were high resident-to-bed ratio (coefficient=51.35, p<0.01), Joint Commission accreditation (coefficient=19.05, p<0.01), presence of other hospitals in the same market with high imaging rates (coefficient=15.29, p<0.01), case severity (coefficient=11.97, p<0.01)…" (suggesting that

more imaging is associated with higher quality hospitals).[25]

Bilimoria et al. examined 2010 data from Hospital Compare and the American Hospital Association and 2009-2010 Medicare claims data; they reported that greater hospital adherence to VTE prophylaxis was very weakly associated with worse risk-adjusted VTE rates (r^2=4.2%, p=0.03) but risk-adjusted VTE rates increased concordantly with VTE imaging use rates (p<0.001). Ju et al. similarly used NSQuIP data to identify VTE events and Medicare claims data to obtain information about use of VTE imaging;[26] mean risk-adjusted VTE rates (within 30 days after surgery) were significantly lower in hospitals in the lowest quartile of VTE imaging use (1.13%) than in hospitals in the highest quartile (1.92%, p<0.001). Similarly, Pierce et al. showed in the National Trauma Data Bank, with 147 hospitals from 2001-2005, that "hospitals with an ultrasound rate of 2% or greater had a 1.07% (95% CI: 1.05-1.09%) increase in reported DVT rate for every 1% increase in ultrasound rate."[27] Admission to a "screening trauma center" that performed vascular ultrasound on at least 2% of admitted trauma patients was independently associated with 2.2 (95% CI 1.1-4.3) times higher odds of DVT, after adjusting for age, injury type, injury severity, need for major surgery, and ventilator days.[28]

The critical question, however, is whether more venous imaging, and hence more diagnosis of VTE, is actually better for patients. Overdiagnosis of VTE among asymptomatic or minimally symptomatic patients may lead to overtreatment, with the known adverse effects of anticoagulation and/or IVC device placement. Evidence-based guidelines note that "although distal DVT may be present in patients with a normal proximal ultrasound, it is seldom if ever associated with important clinical sequelae."[29] With respect to treatment, the American College of Chest Physicians also states, "in patients with acute isolated distal DVT of the leg and without severe symptoms or risk factors for extension… we suggest serial imaging of the deep veins for 2 weeks over initial anticoagulation (Grade 2C)."[30]

To explore this problem, White et al. (personal communication with the author) undertook a local root-cause analysis of all hospital-acquired VTEs at one NQF member organization that had a relatively high PSI 12 rate. They found that some surgical house staff routinely order venous imaging in all febrile patients because they believe that DVT causes postoperative fever. The hospital's vascular laboratory then routinely scans calf veins and reports the presence of DVT in soleal or gastrocnemius muscular branches, despite evidence that sonography limited to proximal veins is equally safe.[31] Indeed, the American College of Radiology's Appropriateness Criteria for Suspected Lower-Extremity Deep Vein Thrombosis specifically advise radiologists (with the maximum appropriateness rating of 9) that "the use of this procedure [ultrasound with Doppler) is limited to between the inguinal ligament and knee."[32] The American Academy of Orthopedic Surgeons also recommends "against routine post-operative duplex ultrasonography screening of patients who undergo elective hip or knee arthroplasty."[33]

In a similar way, pulmonary embolism is now being over diagnosed because small sub segmental filling defects are being read as pulmonary emboli (rather than as "small sub-segmental filling defects of undetermined significance", which is a more appropriate term).[33] This problem of over diagnosis and overtreatment (labeled as "surveillance bias" by some authors) has received increasing attention in the clinical and epidemiologic literature.[34] [35] The three key hallmarks of over diagnosis are: (1) increasing incidence over time; (2) decreasing case fatality over time; and (3) no change in overall attributable mortality over time. Public accountability for PSI 12 rates may have the salutary effect of reducing this epidemic of over diagnosis. Given the negative economic and health consequences of being labeled as having VTE, reducing over diagnosis may improve the overall health of the population.

Conclusions

1. There is irrefutable evidence from randomized controlled trials that pharmacologic and mechanical prophylaxis reduces the incidence of postoperative VTE in appropriately selected patients. This evidence has been thoroughly reviewed and summarized in guidelines from the American College of Chest Physicians and multiple surgical specialty organizations.
2. Based on observational data from case control studies and longitudinal intervention studies, delayed ambulation is an independent risk factor for postoperative VTE, even after adjusting or

restricting the analysis to patients who received appropriate pharmacologic prophylaxis.

3. There is continuing evidence of gaps in relevant processes of care, including missed doses of anticoagulants, failure to start prophylaxis in a timely manner, premature discontinuation of prophylaxis, variation in prescribing practices across physicians-in-training, and use of mechanical prophylaxis alone in untested clinical situations.
4. Therefore, the proportion of potentially preventable PSI 12 events is highly likely to exceed the proportion of potentially preventable inpatient deaths (estimated at 6%[36] to 27%[37]) in measures of 30-day mortality, the proportion of potentially preventable readmissions in measures of 30- day readmissions (estimated at 23%[38] to 27%[39]), and is comparable to the 26% to 54% of surgical site infections that are considered potentially preventable.[40]
5. Although more aggressive use of ultrasound screening appears to be related to higher PSI 12 rates at the hospital level, there is no published evidence that higher screening rates actually lead to better patient outcomes. There is significant evidence of over diagnosis of isolated distal (calf vein) DVT and small sub segmental filling defects in the pulmonary circulation. For this reason, isolated calf vein DVT has been removed from the numerator specification for PSI 12.

In conclusion, a key advantage of risk-adjusted outcome measures, such as PSI 12, is that they allow health care providers to identify previously unrecognized opportunities for prevention, and they encourage providers to allocate resources to approaches that will achieve the greatest benefit at a reasonable cost. This understanding has led to a 49% nationwide reduction in the incidence of CLABSI[22] (with sustained reductions as great as 70-80% in some hospitals and regional and national collaborative).[41 42 43] Continued progress in preventing VTE requires continued measurement of these events over time, as well as continued research to improve prophylactic approaches and continued quality improvement efforts to bring those approaches into widespread practice.

[1] Guyatt GH, Eikelboom JW, Gould MK, et al. Approach to Outcome Measurement in the Prevention of Thrombosis in Surgical and Medical Patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141(2_suppl):e185S-e194S.

[2] Lyman GH, Khorana AA, Kuderer NM, et al. American Society of Clinical Oncology Clinical Practice. Venous thromboembolism prophylaxis and treatment in patients with cancer: American Society of Clinical Oncology clinical practice guideline update. J Clin Oncol. 2013 Jun 10;31(17):2189-204.

[3] http://www.aaos.org/Research/guidelines/VTE/VTE_full_guideline.pdf

[4] http://www.aaos.org/Research/guidelines/HipFxSummaryofRecommendations.pdf

[5] http://www.guideline.gov/content.aspx?id=11429

[6] Sadeghi B, Romano PS, Maynard G, et al. Mechanical and suboptimal pharmacologic prophylaxis and delayed mobilization but not morbid obesity are associated with venous thromboembolism after total knee arthroplasty: a case-control study. J Hosp Med 2012;7(9):665-71.

[7] White RH, Gettner S, Newman JM, Trauner KB, Romano PS. Predictors of rehospitalization for symptomatic venous thromboembolism after total hip arthroplasty. N Engl J Med. 2000 Dec 14;343(24):1758-64.

[8] Chandrasekaran S, Ariaretnam SK, Tsung J, Dickison D. Early mobilization after total knee replacement reduces the incidence of deep venous thrombosis. Aust N Z J Surg. 2009;79(7–8):526–529.

[9] Pearse EO, Caldwell BF, Lockwood RJ, Hollard J. Early mobilisation after conventional knee replacement may reduce the risk of postoperative venous thromboembolism. J Bone Joint Surg Br. 2007; 89(3):316-22.

[10] Zhang ZH, Shen B, Yang J, Zhou ZK, Kang PD, Pei FX. Risk factors for venous thromboembolism of total hip arthroplasty and total knee arthroplasty: a systematic review of evidences in ten years. BMC Musculoskelet Disord. 2015 Feb 10;16:24.

[11] Agency for Healthcare Research and Quality (AHRQ). Making Health Care Safer II: An Updated Critical Analysis of the Evidence for Patient Safety Practices. Rockville, MD: AHRQ Publication No. 13-E001-EF; 2013. http://www.ahrq.gov/research/findings/evidence-based-reports/services/quality/patientsftyupdate/ptsafetyIIchap28.pdf

[12] Markovic-Denic L, Zivkovic K, Lesic A, Bumbasirevic V, Dubljanin-Raspopovic E, Bumbasirevic M. Risk factors and distribution of symptomatic venous thromboembolism in total hip and knee replacements: prospective study. Int Orthop. 2012 Jun;36(6):1299-305.

[13] Borzecki AM, Cowan AJ, Cevasco M, et al. Is development of postoperative venous thromboembolism related to

thromboprophylaxis use? A case-control study in the Veterans Health Administration. Jt Comm J Qual Patient Saf. 2012 Aug;38(8):348-58.

[14] http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6309a3.htm

[15] Haut ER, Lau BD, Kraus PS, et al. Preventability of hospital-acquired venous thromboembolism. JAMA Surg. 2015 Sep 1;150(9):912-5.

[16] Lau BD, Streiff MB, Pronovost PJ, Haider AH, Efron DT, Haut ER. Attending physician performance measure scores and resident physicians' ordering practices. JAMA Surg. 2015 Aug 1;150(8):813-4.

[17] Colorectal Writing Group for the Surgical Care and Outcomes Assessment Program–Comparative Effectiveness Research Translation Network (SCOAP-CERTAIN) Collaborative. Thromboembolic Complications and Prophylaxis Patterns in Colorectal Surgery. JAMA Surg 2015 Jun 10.[Epub ahead of print]

[18] Ang DN, Behrns KE. Using a relational database to improve mortality and length of stay for a department of surgery: a comparative review of 5200 patients. Am Surg 2013;79(7):706-710.

[19] Heslin MJ, Taylor B, Hawn MT, et al. A 100% departmental mortality review improves observed-to-expected mortality ratios and University HealthSystem Consortium rankings. J Am Coll Surg 2014;218(4):554-562.

[20] Hussey PS, Burns RM, Weinick RM, et al. Using a hospital quality improvement toolkit to improve performance on the AHRQ quality indicators. Jt Comm J Qual Saf 2013;39(4):177-184.

[21] http://www.universityofcalifornia.edu/sites/default/files/uc_chqi_report_8.27.14.pdf

[22] http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2013.html

[23] Wang Y, Eldridge N, Metersky ML, et al. National trends in patient safety for four common conditions, 2005-2011. N Engl J Med.2014 Jan 23;370(4):341-51.

[24] Rajaram R, Barnard C, Bilimoria KY. Concerns about using the patient safety indicator-90 composite in pay-for-performance programs. JAMA. 2015 Mar 3;313(9):897-8.

[25] Chung JW, Ju MH, Kinnier CV, Haut ER, Baker DW, Bilimoria KY. Evaluation of hospital factors associated with hospital postoperative venous thromboembolism imaging utilisation practices. BMJ Qual Saf. 2014;23(11):947-56.

[26] Ju MH, Chung JW, Kinnier CV, et al. Association between hospital imaging use and venous thromboembolism events rates based on clinical data. Ann Surg 2014;260(3):558-564.

[27] Pierce CA, Haut ER, Kardooni S, et al. Surveillance bias and deep vein thrombosis in the national trauma data bank: the more we look, the more we find. J Trauma 2008;64(4):932-6.

[28] Haut ER, Chang DC, Pierce CA, et al. Predictors of posttraumatic deep vein thrombosis (DVT): hospital practice versus patient factors-an analysis of the National Trauma Data Bank (NTDB). J Trauma 2009;66(4):994-9.

[29] http://journal.publications.chestnet.org/data/Journals/CHEST/23443/112299.pdf

[30] http://journal.publications.chestnet.org/article.aspx?articleID=1159520

[31] Bernardi E, Camporese G, Büller HR, et al. Serial 2-point ultrasonography plus D-dimer vs whole-leg color-coded Doppler ultrasonography for diagnosing suspected symptomatic deep vein thrombosis: a randomized controlled trial. JAMA 2008;300(14):1653-9.

[32] https://acsearch.acr.org/docs/69416/Narrative

[33] Suh TM, Cronan JJ, Healey TT. Dots are not clots: the over-diagnosis and over-treatment of PE. Emerg Radiol (2010) 17:347– 52.

[34] Righini M, Le Gal G, Bounameaux H. Venous thromboembolism diagnosis: unresolved issues. Thromb Haemost 2015;113(6):1184-92. Wiener RS, Schwartz LM, Woloshin S. Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. Arch Intern Med 2011;171(9):831-7.

[35] D'Apuzzo MR, Keller TC, Novicoff WM, et al. CT pulmonary angiography after total joint arthroplasty: overdiagnosis and iatrogenic harm? Clin Orthop Relat Res 2013;471(9):2737-42.

36 Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. JAMA 2001;286:415–20.

37 Dubois RW, Brook RH. Preventable deaths: who, how often, and why? Ann Intern Med 1988;109:582–9.

38 van Walraven C, Jennings A, Forster AJ. A meta-analysis of hospital 30-day avoidable readmission rates. J Eval Clin Pract. 2012 Dec;18(6):1211-8.

39 van Walraven C, Bennett C, Jennings A, Austin PC, Forster AJ. Proportion of hospital readmissions deemed avoidable: a systematic review. CMAJ. 2011 Apr 19;183(7):E391-402.

40 Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. Infect Control Hosp Epidemiol. 2011

Feb;32(2):101-14.

41 Pronovost PJ, Watson SR, Goeschel CA, Hyzy RC, Berenholtz SM. Sustaining Reductions in Central Line-Associated Bloodstream Infections in Michigan Intensive Care Units: A 10-Year Analysis. Am J Med Qual. 2015 Jan 21.

42 Palomar M, Álvarez-Lerma F, Riera A, Díaz MT, Torres F, Agra Y, et al. Bacteremia Zero Working Group. Impact of a national multimodal intervention to prevent catheter-related bloodstream infection in the ICU: the Spanish experience. Crit Care Med.

2013 Oct;41(10):2364-72.

43 Marsteller JA, Sexton JB, Hsu YJ, Hsiao CJ, Holzmueller CG, Pronovost PJ, et al. A multicenter, phased, cluster-randomized

controlled trial to reduce central line-associated bloodstream infections in intensive care units. Crit Care Med. 2012 Nov;40(11):2933-

Appendix C: 0352-Failure to Rescue In-Hospital Mortality (risk adjusted) memo

REF: 0352, "Failure to Rescue In-Hospital Mortality (risk adjusted)"

The Children's Hospital of Philadelphia
Center for Outcomes Research

Founded 1855

Jeffrey H. Silber, MD, PhD, *Director*
The Nancy Abramson Wolfson Professor of Health Services Research
Director, Center for Outcomes Research
The Children's Hospital of Philadelphia
Professor of Pediatrics, Anesthesiology and Critical Care, and
silber@email.chop.edu Health Care Management, The University of Pennsylvania

3535 Market Street,
Suite 1029
Philadelphia, PA 19104
Tel: (215) 590-2540
Fax: (215) 590-2378
Email:

August 14, 2015

REF: 0352, "Failure to Rescue In-Hospital Mortality (risk adjusted)"

To Whom It May Concern,

Thank you for giving us the opportunity to respond to the helpful comments of the National Quality Forum Patient Safety committee. We have responded to comments summarized by staff and abstracted from the meeting transcript and believe that the resulting application is stronger and clearer than our original. Below is a response to comments (committee request/question in **bold**, our response in normal font).

Committee's comments as summarized by NQF staff:

1. **Are there more current data available on the reliability of the measure? (the developer mentioned a 2014 paper)**

Yes. We have repeated an analysis of reliability using a more recent dataset of Medicare claims spanning

6 states (CA, GA, MD, NY, OH and PA), and the years 2005-2007. A data description can be found in Section 1.2 of the Measure Testing Form. Results were similar to the previous analysis and can be found in Section 2b2.2 of the Measure Testing form.

2. **Provide the failure to rescue rates that appeared more recently (ex. Table 2: Access to Hospital Distribution of Failure to Rescue Rates in Orthopedics and General Surgery that are Standardized)**

In Table 1 (page 3) of the Evidence form, an overview of studies reporting FTR (including the above) is provided. The reported associations with various hospital characteristics are provided for both unadjusted and adjusted associations when available. In the new analysis (noted in Question 1 above) we provide the FTR rates in Table 2 (page 5) of the Testing Measurement form, which was 3.68%.

3. **Address mistake made in table on co-morbidities. How is the data presented in the measure worksheet being presented?**
4. In our previous submission we mistakenly provided a validation model that included both patient and hospital characteristics. This was a mistake. In the revised submission we developed a new risk adjustment model with more recent data from 2005-7. The risk adjustment model now only includes patient characteristics. The complete list of 34 comorbidities is shown in the updated risk model in the Testing form, along with all DRGs and interaction terms included in the risk model (see page 8 of the Testing form, section 2b4.3, and the model starting on page 9, Table 4). **Correct the denominator to not only include complications, but also the number of patients who died without complications.**

Throughout all documents when defining the population included in the measure, we have tried to clarify that the numerator includes all patients who died in the hospital and that the denominator includes patients who either had an in-hospital complication or died within the hospital without a documented in-hospital complication.

5. **What is the rationale for excluded patients who are over 90 years old? (is this based on evidence?)**

We exclude patients over age 90, due to the increased likelihood that these patients will have DNR orders. This could introduce bias by yielding elevated failure-to-rescue rates at hospitals who treated more patients with DNR orders, potentially disproportionately penalizing these hospitals for deaths that were out of their control. If DNR status were available in the dataset, it could have been used as a more accurate exclusion criteria variable than age alone, however it is not available in CMS claims.

6. **Concern that board certified anesthesiologists are the only sole providers included when there are other sole providers**

Beginning on page 212, line 11 and continuing to page 213, line 6 of the meeting transcript, Ms. Ardizzone states,

"Hi, Laura Ardizzone. I just wanted to comment on some of their evidence. Number 5, which is their Silber study, anesthesiologist direction, is a highly controversial and actually, in my mind, fatally flawed study in how they compared outcome rates as compared to nurse anesthetists.

"So I mean, that's not up for discussion here. But just to kind of clarify that I think some of the evidence that they're using may be off point. I know that one is.

"On top of the fact that they talk about the percentage of board certified anesthesiologists and in 18 states, anesthesiologists are not the only sole providers in 18 states. Nurse anesthetists are sole providers of anesthesia. So I'm just questioning some of the evidence."

In response, we would state: (1) we disagree that this study was "fatally flawed." We have never seen a study that directly refuted its conclusions in the **same** population of Medicare patients undergoing the **same** procedures. (2) The **same** "flawed" model in that **same** paper showed that the nurse-to-bed ratio was the most important factor associated with mortality and failure-to- rescue. (3) That **same** model was used in the Aiken (and Silber) 2002 *JAMA* paper (winner of the AcademyHealth Article of the Year) that also showed the importance of nurse-to-bed ratio and failure-to-rescue. Finally, (4) we cited 35 papers. To discredit us because the reviewer did not like one finding out of many findings, seems a bit off-point regarding the merits

of the FTR measure itself. Many other studies have looked at nursing characteristics and their association with failure-to-rescue, such as nurse educational level and nurse-to-bed ratio (Aiken *JAMA* 2003; Aiken *JAMA* 2002). In Table 1 (page 3) in the Evidence form we now provide a table describing the research that demonstrates clear association between better failure-to-rescue rates and hospital characteristics associated with better quality of care.

7. **All characteristics that appear in the regression are not included (developer listed them vocally)**

The hospital characteristics, such as technology level, were included in previously submitted Risk-adjustment models in error (it was a validity model). The updated risk-adjustment model and results can be found in Section 2b4 of the Measure Testing Form.

8. **How is the risk adjustment model calculated?**

The revised risk-adjustment model includes only patient characteristics such as age, sex, comorbidities, emergency admission status, transfer-in status, DRG and principal procedure, with interactions between variables included when significant. Individual covariates were retained if they were significantly associated with in-hospital mortality, in-hospital complications, or in-hospital failure-to-rescue in univariate models at the $P < 0.15$ level. Some clinically important patient covariates were included even if they did not reach the P-Value criteria (e.g. AIDS). All pairwise interactions were also included in the final model if they were significant at the two-sided 5% level with the significance level based on the Bonferroni correction.

9. **There is variation in skill mix for day to day, how is this accounted for?**

Skill mix and other hospital characteristics are not included in the risk-adjustment model. This type of variation is not typically tracked in datasets to which we have access. For example, the Healthcare Cost Report Information System (HCRIS) and Medicare Provider of Service files offer annual information on these variables, but no files provide day-by-day measures of skill mix or any other hospital information that changes daily.

Committee's comments from the meeting transcript (Transcript location provided):

1. **Page 183, Lines 13-15: A question about the time frame for the data being used for the reliability. Is it still 1999 and 2000, or are there more current data?**

We have repeated the prior reliability work using Medicare claims from 6 states (CA, GA, NY, OH, and PA) spanning the years 2005-2007. A data description can be found in Section 1.2 of the Measure Testing Form.

2. **Page 185, Lines 15-21: I didn't see as many numbers in terms of different rates and stuff. It was more descriptive in articles, where other measures actually start talking about rates and ratios and different risk factors. Are they there somewhere I just missed it, or in the publications?**

We have added two tables describing the risk adjustments (Table 5 (page 26) in the Testing Measurement form) and study results (Table 1 (page 3) in the Evidence form) for various studies assessing associations with failure-to-rescue.

3. **Page 188, Lines 3-10: I'm curious how they are identifying comorbidities. They have a number that they have talked about age, sex, transfer status, whether it's a high-tech hospital, teaching hospital, bed size, bed-to-nurse ratio, staff mix. That's not claims data information, and so I'm wondering how they're gathering that information?**

We use a set of 34 comorbidities that are defined by the ICD-9 codes in Appendix C (an ICD-10 version is also provided in Appendix E). The comorbidities are defined using data from the patient's claims in at least the 3 months prior to their index admission (codes from the index admission are also used for most comorbidities; please see Appendices C and E). For Medicare data, all patient claims (inpatient, outpatient, and carrier/Part B) are scanned for the existence of these qualifying codes. Patient age is defined by taking the difference of the patient's admission date and their date of birth.

Hospital teaching status, technology status, nurse-to-bed ratio, and nurse mix are not patient comorbidities, but we define them to study the relationship between failure-to-rescue and hospital characteristics, in order to establish validity. These variables can be defined using the Medicare Provider of Services (POS) file

or the Healthcare Cost Report Information System (HCRIS).

4.
5. **Page 189, Lines 8-10: Just a question for the developer on the exclusion of patients over the age 90. Can you talk us through the rationale on that?**

Patients over age 90 are excluded on the assumption that a larger proportion of these patients will have DNR orders. This could introduce a bias towards increased failure-to-rescue due to DNR status census, potentially disproportionately penalizing hospitals for deaths that were out of their control. If DNR status was included in the dataset, this could be used as a more accurate exclusion criteria variable.

6. **Page 195, Lines 7-10: I thought the other criteria was that it showed performance improvement over time, and I thought that those were missing.**

A recent study was completed by Kutney-Lee at al. showing changes over time in failure-to- rescue in a sample of hospitals that attained Magnet recognition between 1999 and 2007 with hospitals that remained non-Magnet. A description of the study and its results can be found in Section 1b.2 of the Main Form.

7. **Page 211, Lines 8-14: I'm on Page 1 and it says, "In summary, failure rate was a function of anesthesia board certification and the presence of surgical staff, but not a function of admission severity or illness score." It does not mention nurse-to-bed ratio, nurse mix.**

We apologize for the confusion. We only intend that "in summary" statement to refer to that study specifically (the first FTR study from 1992), not to all evidence we cited linking failure-to- rescue rates to hospital characteristics that occurred over the past 25 years. This statement has been deleted to avoid confusion. Table 1 (page 3) in the Evidence form provides the results of fifteen studies (both ours and many performed by other researchers) that assessed the associations of various hospital/provider characteristics with failure-to-rescue.

8. **Page 214, Lines 1-4: They had a testing sample that was the 65 to 90 year olds for general surgery, but the measure is for 18 to 90 year olds for general, vascular, and orthopedic surgery.**

The data on which this risk-adjustment model is based was Medicare claims for general, vascular, and orthopedic surgery patients age 65 to 90. We do not have validated models for adults below 65. Others published studies that have applied the FTR measure to populations including less than 65 years old including: Aiken *JAMA* 2002, Aiken *JAMA* 2003, Friese *Surgery* 2010, Glance *Ann Surg* 2011, Kendall-Gallagher *J Nurs Scholarsh* 2011, Kutney-Lee *Med Care* 2015, Sheetz *J Am Coll Surg* 2014, and Wright *J Clin Oncol* 2012.

9. **Page 214, Lines 6-11: There were patients who died with a complication and they also included, as in the previous measure, patients who died without documented complications. So I just have a little concern about including them in the numerator statement.**

The designation of having a complication is determined by the codes described in the Appendix. If a patient dies after surgery without a complication, we presume there was an undocumented complication. This represents a small number of patients (Silber *Med Care* 2007).

10. **Page 221-222, Lines 15-22;1-4: After the developer's last comment about the sort of finite list of things that they identified as really being impactful to the measure, like the availability of anesthesia, high tech, all of those things, if they could clarify then the pretty sophisticated risk-adjusted model that has 160 different characteristics, so the rationale for having a risk-adjustment model that has 160 characteristics given that it's things like the availability of a CAT scan and an anesthesiologist.**

We apologize again for erroneously including the wrong model. We mistakenly reported a validity model instead of the risk adjustment model. In our presentation of the findings and definitions, we did not intend to conflate the patient characteristics for which we risk-adjust and the hospital characteristics that have been found to be associated with performance on the risk-adjusted measure. The correct risk-adjustment model includes only patient characteristics such as age, sex, comorbidities, emergency admission status, transfer-in status, DRG, and principal procedure, with interactions between variables included when statistically significant using the Bonferroni correction. Each patient characteristic was tested in a univariate model in the development sample, and retained if statistically significant at the $P < 0.15$ level Interactions between pairs of variables were also included in the final model if they were significant at the two-sided 5% level with

the significance level based on the Bonferroni correction..

The updated risk-adjustment model and results can be found in Section 2b4 of the Measure Testing Form. Additionally, as this metric can be used for various populations that are very diverse, users should construct a model with coefficients that would be applicable to their study sample. Examples of coefficients that have been included in published studies are included in Table 5 (page 26) within the Measure Testing Form. There are some coefficients that are commonly used, and others that are specific to the study sample. For example, Glance *Ann Surg* 2011 adjusted for mechanism of trauma, as they were comparing low- and high-mortality hospitals caring for trauma patients.

11. **Page 222, Lines 5-15: Then also if they could please provide some more detail on how this is calculated? They're saying that it uses administrative data, but it also does seem to have some patient level characteristics in the risk adjustment. So I'm trying to understand, is there software that's been developed? How is the measure actually calculated? So, like, what is the math behind that? Particularly, the risk adjustment piece would be extremely helpful.**

For unadjusted FTR, we compute a ratio. In the numerator is the number of patients who died, and in the denominator is the number of patients who developed a complication or died with no complication (something we say is an undocumented complication—as the patient was initially alive and died after an elective surgery, so the event must have occurred and was not documented.)
Examining adjusted FTR requires some sort of O/E or (O-E)/N computation (where O=observed, E = expected, and N = the number of patients analyzed), reflecting indirect standardization, or a direct standardization computation. This can be accomplished with regression models or matching analyses, and in our citations we provide examples for both approaches
In the simple regression approach for FTR we create a data set which includes only patients who developed a complication or who died without a complication. This population (N) includes all the patients who died and all the patients who developed a complication. We simply then run a logistic regression model with the "y" dependent variable being death (0=alive, 1 = dead) and the "x" or independent variables being all the patient characteristics. Once the regression model is fit, we can then estimate a predicted probability of dying (in this population of patients who had a complication or who died without a complication) for every patient in the data set. Then we examine each patient at each hospital who had a complication or who died without a complication and estimate a predicted probability of FTR. We sum each patient risk of FTR to obtain the E, and count the deaths (failures) to get the O (observed) so we can then compute a hospital specific O/E or (O-E)/N. Of course there are many ways to create the regressions, but this description represents the simplest indirect standardization method. Note: The risk- adjustment model we provide was produced through the more complex process described elsewhere.

An alternative approach is to make comparisons through matching, and we also cite these papers. With matching, we compare similar patients (based on patient characteristics, types of comorbidities, etc.) and examine FTR rates in these similar patients. Matching is another form of risk adjustment.

Administrative data can be used to identify patient characteristics such as age, sex, principal procedure, DRG, comorbidities, and admission status. Patient age is defined by taking the difference of the patient's admission date and their date of birth. We use a set of 34 comorbidities that are defined by the ICD-9 codes in Appendix C (an ICD-10 version is also provided in Appendix E). The comorbidities are defined using data from the patient's claims in at least the 3 months prior to their index admission (codes from the index admission are also used for most comorbidities; please see Appendices C and E). When using Medicare data, all patient claims (inpatient, outpatient, and carrier/Part B) from the admission and lookback period are scanned for the existence of these qualifying codes. The comorbidity and complication definitions can be applied to any dataset that includes patient hospitalizations linked to longitudinal health services utilization data.

Hospital teaching status, technology status, nurse-to-bed ratio, and nurse mix are not patient comorbidities, but we define them to study the relationship between FTR and hospital characteristics, in order to establish validity. These variables can be defined using the Medicare Provider of Services (POS) file or the Healthcare Cost Report Information System (HCRIS).

The failure-to-rescue rate is simply the rate of deaths in the hospital among those who had an in- hospital complication or death in the hospital in the target case population. The calculation algorithm and measure length are described more detail in Section S18 of the Main form as such: "Patients admitted to an acute care facility with a stay characterized by a principal procedure and DRG of interest as outlined in the attached Appendix that can also be found on the website (http://www.research.chop.edu/programs/cor/node/26). Those patients both alive and without complications were excluded, as were any below 18 years of age or above 90 years old. Cases meeting the target process were therefore between the ages of 18-90 years old, admitted to an acute care facility for a DRG of interest, and had a complication or died without a documented complication in the hospital. The event of interest is death."

12. **Page 222-223, Lines 22;1-3: Often nurse staffing is variable during a patient's stay and you may not have exactly the same skill mix every time during the day, nor every day. And where is the count being taken?**

Skill mix and other hospital characteristics are not included in the proper risk-adjustment model. This type of variation is not typically tracked in datasets to which we have access. For example, the Healthcare Cost Report Information System (HCRIS) and Medicare Provider of Service files offer annual information on these variables, but no files provide day-by-day measures of skill mix or any other hospital information that changes daily.

For studies by our group that adjusted for hospital characteristics, covariates like nurse staffing are absolute, not variable, as they are a cross-sectional study assessing the FTR rate at a given time point. Kutney-Lee et al., 2015 compared hospitals at two different time points, adjusting for hospital characteristics which were reported for each of those time points. Showing that changes in hospital characteristics over time can be accounted for in FTR rate comparisons.

Preliminary analysis comments (section location included):

1. **Page 3, Section 2a1: ICD-10 codes are NOT included in the specifications.**

We have updated the Appendix to include proposed ICD-10 codes, these can be found in Appendices D and E. These will still require testing once claims using ICD-10 are available, but we have reviewed them for clinical comparability with the ICD-9 codes.

2. **Page 3, Section 2a1: The developer notes that "when Physician Part B is available, the definition of complications and comorbidities are augmented to include CPT codes." It is therefore unclear whether CPT codes should always be used when calculating the measure.** Regardless of the data source, any time Current Procedural Terminology (CPT) codes are available, the definition of complications and comorbidities should be augmented to include them. This has been clarified in the forms.

3. **Page 4, Section 2a2: The developer indicates that reliability testing was conducted at the data element level. However, no information regarding individual data elements was provided. The developer described the method of reliability testing as "split half sample correlation", which would suggest that testing at the <u>performance measure score level</u> was conducted. No information was provided to describe how the split-sample method was used or which correlation statistic was calculated. The reliability statistic reported was 0.32 (0.56 when a correction factor was applied), but no interpretation of these values was provided. Because information regarding the methodology was not provided, it is unclear whether the method used was an appropriate method of testing for reliability of the performance measure score and how to interpret the results.**

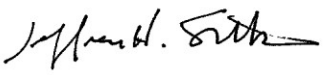
Our approach to reliability was based on the classic book by Lord and Novick (Lord & Novick 1968). Our approach, using the Lord and Novick methodology to present reliability was published in *Medical Care* in 2007 (Silber *Med Care* 2007). Using the same approach, we have updated the data set for this new NQF revision. We ask how reliable is the FTR measure for assessing a hospital rank. We do this by forming two random split groups inside each hospital ("wing A" and "wing B") and ranking hospitals on one wing and then the other. When we discuss reliability, we are interested in the correlation in FTR ranking between the two random hospital "wings". Further description of this methodology, as well as the results can be found in

Section 2a2 of the Measure Testing form.

A highly reliable FTR measure will provide similar rankings for the same hospital across its random "wings" or the random samples of its patients. In our previous work we compared our FTR measure with the FTR measure used by AHRQ and also certified by the NQF. The Silber FTR had higher reliability. In the new models we developed on a more recent data set, we found similar results to the 2007 Medical Care paper (Silber *Medical Care* 2007).

To get a feel for the reliability we provide, it should be viewed in context to other measures. That is why in our report we compare this reliability to using the 30-day mortality rate and the AHRQ-based FTR measures (that we compute by utilizing only those patients experiencing the AHRQ-based complications). The comparisons are now found in Table 2 (page 5) of the Measure Testing form.

Sincerely,



The Nancy Abramson Wolfson Professor of Health Services Research Director,
Center for Outcomes Research
The Children's Hospital of Philadelphia
Professor of Pediatrics and Anesthesiology & Critical Care The
Perelman School of Medicine
Professor of Health Care Management, The Wharton School
The University of Pennsylvania

Appendix D: 0353-Failure to Rescue 30-Day Mortality (risk adjusted) Memo

The Children's Hospital of Philadelphia

Center for Outcomes Research

Founded 1855

Jeffrey H. Silber, MD, PhD, *Director*
Suite 1029
The Nancy Abramson Wolfson Professor of Health Services Research
Director, Center for Outcomes Research
The Children's Hospital of Philadelphia
Professor of Pediatrics, Anesthesiology and Critical Care, and
silber@email.chop.edu Health Care Management, The University of Pennsylvania

3535 Market Street,

Philadelphia, PA 19104
Tel: (215) 590-2540
Fax: (215) 590-2378
Email:

August 14, 2015

REF: 0353, "Failure to Rescue 30-Day Mortality (risk adjusted)"

To Whom It May Concern,

Thank you for giving us the opportunity to respond to the helpful comments of the National Quality Forum Patient Safety committee. We have responded to comments summarized by staff and abstracted from the meeting transcript and believe that the resulting application is stronger and clearer than our original. Below is a response to comments (committee request/question in **bold**, our response in normal font).

Committee's comments as summarized by NQF staff:

1. **Are there more current data available on the reliability of the measure? (the developer mentioned a 2014 paper)**

Yes. We have repeated an analysis of reliability using a more recent dataset of Medicare claims spanning 6 states (CA, GA, MD, NY, OH and PA), and the years 2005-2007. A data description can be found in Section 1.2 of the Measure Testing Form. Results were similar to the previous analysis and can be found in Section 2b2.2 of the Measure Testing form.

2. **Provide the failure to rescue rates that appeared more recently (ex. Table 2: Access to Hospital Distribution of Failure to Rescue Rates in Orthopedics and General Surgery that are Standardized)**

In Table 1 (page 3) of the Evidence form, an overview of studies reporting FTR (including the above) is provided. The reported associations with various hospital characteristics are provided for both unadjusted and adjusted associations when available. In the new analysis (noted in Question 1 above) we provide the FTR rates in Table 2 (page 5) of the Testing Measurement form, which was 4.48%.

3. **Address mistake made in table on co-morbidities. How is the data presented in the measure worksheet being presented?**

In our previous submission we mistakenly provided a validation model that included both patient and hospital characteristics. This was a mistake. In the revised submission we developed a new risk adjustment model with more recent data from 2005-7. The risk adjustment model now only includes patient characteristics. The complete list of 34 comorbidities is shown in the updated risk model in the Testing form, along with all DRGs and interaction terms included in the risk model (see page 8 of the Testing form, section 2b4.3, and the model starting on page 9, Table 4).

4. **Correct the denominator to not only include complications, but also the number of patients who died without complications.**

Throughout all documents when defining the population included in the measure, we have tried to clarify that the numerator includes all patients who died within 30 days of admission and that the denominator includes patients who either had an in-hospital complication or died within 30 days of admission without a documented in-hospital complication.

5. **What is the rationale for excluded patients who are over 90 years old? (is this based on evidence?)**

We exclude patients over age 90, due to the increased likelihood that these patients will have DNR orders. This could introduce bias by yielding elevated failure-to-rescue rates at hospitals who treated more patients with DNR orders, potentially disproportionately penalizing these hospitals for deaths that were out of their control. If DNR status were available in the dataset, it could have been used as a more accurate exclusion criteria variable than age alone, however it is not available in CMS claims.

6. **Concern that board certified anesthesiologists are the only sole providers included when there are other sole providers.**

Beginning on page 212, line 11 and continuing to page 213, line 6 of the meeting transcript, Ms. Ardizzone states,

"Hi, Laura Ardizzone. I just wanted to comment on some of their evidence. Number 5, which is their Silber study, anesthesiologist direction, is a highly controversial and actually, in my mind, fatally flawed study in how they compared outcome rates as compared to nurse anesthetists.

"So I mean, that's not up for discussion here. But just to kind of clarify that I think some of the evidence that they're using may be off point. I know that one is.

"On top of the fact that they talk about the percentage of board certified anesthesiologists and in 18 states, anesthesiologists are not the only sole providers in 18 states. Nurse anesthetists are sole providers of anesthesia. So I'm just questioning some of the evidence."

In response, we would state: (1) we disagree that this study was "fatally flawed." We have never seen a study that directly refuted its conclusions in the **same** population of Medicare patients undergoing the **same** procedures. (2) The **same** "flawed" model in that **same** paper showed that the nurse-to-bed ratio was the most important factor associated with mortality and failure-to- rescue. (3) That **same** model was used in the

Aiken (and Silber) 2002 *JAMA* paper (winner of the AcademyHealth Article of the Year) that also showed the importance of nurse-to-bed ratio and failure-to-rescue. Finally, (4) we cited 35 papers. To discredit us because the reviewer did not like one finding out of many findings, seems a bit off-point regarding the merits of the FTR measure itself. Many other studies have looked at nursing characteristics and their association with failure-to-rescue, such as nurse educational level and nurse-to-bed ratio (Aiken *JAMA* 2003; Aiken *JAMA* 2002). In Table 1 (page 3) in the Evidence form we now provide a table describing the research that demonstrates clear association between better failure-to-rescue rates and hospital characteristics associated with better quality of care.

7. **All characteristics that appear in the regression are not included (developer listed them vocally)**

The hospital characteristics, such as technology level, were included in previously submitted risk-adjustment models in error (it was a validity model). The updated risk-adjustment model and results can be found in Section 2b4 of the Measure Testing Form.

8. **How is the risk adjustment model calculated?**

The revised risk-adjustment model includes only patient characteristics such as age, sex, comorbidities, emergency admission status, transfer-in status, DRG and principal procedure, with interactions between variables included when significant. Individual covariates were retained if they were significantly associated with 30-day mortality, in-hospital complications, or 30-day failure-to-rescue in univariate models at the $P < 0.15$ level. Some clinically important patient covariates were included even if they did not reach the P-Value criteria (e.g. AIDS). All pairwise interactions were also included in the final model if they were significant at the two- sided 5% level with the significance level based on the Bonferroni correction.

9. **There is variation in skill mix for day to day, how is this accounted for?**
10. Skill mix and other hospital characteristics are not included in the risk-adjustment model. This type of variation is not typically tracked in datasets to which we have access. For example, the Healthcare Cost Report Information System (HCRIS) and Medicare Provider of Service files offer annual information on these variables, but no files provide day-by-day measures of skill mix or any other hospital information that changes daily **Specific to 0353, how is 30 day mortality data retrieved after a patient has been discharged?**

Our typical experience has been with Medicare data which includes vital status information for all Medicare beneficiaries. It is also our experience that hospital discharge records for all adult hospitalizations can be linked with vital status information available from many states, including California, Missouri, New York, and Pennsylvania. We agree that it is always better to use data that accounts for deaths outside the hospital. When such linked vital status data is not available, investigators can apply the in-hospital measure of failure-to-rescue (0352).

Committee's comments from the meeting transcript (Transcript location provided):

1. **Page 183, Lines 13-15: A question about the time frame for the data being used for the reliability. Is it still 1999 and 2000, or are there more current data?**

We have repeated the prior reliability work using Medicare claims from 6 states (CA, GA, NY, OH, and PA) spanning the years 2005-2007. A data description can be found in Section 1.2 of the Measure Testing Form.

2. **Page 185, Lines 15-21: I didn't see as many numbers in terms of different rates and stuff. It was more descriptive in articles, where other measures actually start talking about rates and ratios and different risk factors. Are they there somewhere I just missed it, or in the publications?**

We have added two tables describing the risk adjustments (Table 5 (page 28) in the Testing Measurement form) and study results (Table 1 (page 3) in the Evidence form) for various studies assessing associations with failure-to-rescue.

3. **Page 188, Lines 3-10: I'm curious how they are identifying comorbidities. They have a number that they have talked about age, sex, transfer status, whether it's a high-tech hospital, teaching hospital, bed size, bed-to-nurse ratio, staff mix. That's not claims data information, and so I'm**

wondering how they're gathering that information?

We use a set of 34 comorbidities that are defined by the ICD-9 codes in Appendix C (an ICD-10 version is also provided in Appendix E). The comorbidities are defined using data from the patient's claims in at least the 3 months prior to their index admission (codes from the index admission are also used for most comorbidities; please see Appendices C and E). For Medicare data, all patient claims (inpatient, outpatient, and carrier/Part B) are scanned for the existence of

these qualifying codes. Patient age is defined by taking the difference of the patient's admission date and their date of birth.

Hospital teaching status, technology status, nurse-to-bed ratio, and nurse mix are not patient comorbidities, but we define them to study the relationship between failure-to-rescue and hospital characteristics, in order to establish validity. These variables can be defined using the Medicare Provider of Services (POS) file or the Healthcare Cost Report Information System (HCRIS).

4. **Page 189, Lines 8-10: Just a question for the developer on the exclusion of patients over the age 90. Can you talk us through the rationale on that?**

Patients over age 90 are excluded on the assumption that a larger proportion of these patients will have DNR orders. This could introduce a bias towards increased failure-to-rescue due to DNR status census, potentially disproportionately penalizing hospitals for deaths that were out of their control. If DNR status was included in the dataset, this could be used as a more accurate exclusion criteria variable.

5. **Page 195, Lines 7-10: I thought the other criteria was that it showed performance improvement over time, and I thought that those were missing.**

A recent study was completed by Kutney-Lee at al. showing changes over time in failure-to- rescue in a sample of hospitals that attained Magnet recognition between 1999 and 2007 with hospitals that remained non-Magnet. A description of the study and its results can be found in Section 1b.2 of the Main Form.

6. **Page 211, Lines 8-14: I'm on Page 1 and it says, "In summary, failure rate was a function of anesthesia board certification and the presence of surgical staff, but not a function of admission severity or illness score." It does not mention nurse-to-bed ratio, nurse mix.**

We apologize for the confusion. We only intend that "in summary" statement to refer to that study specifically (the first FTR study from 1992), not to all evidence we cited linking failure-to- rescue rates to hospital characteristics that occurred over the past 25 years. This statement has been deleted to avoid confusion. Table 1 (page 3) in the Evidence form provides the results of fifteen studies (both ours and many performed by other researchers) that assessed the associations of various hospital/provider characteristics with failure-to-rescue.

7. **Page 214, Lines 1-4: They had a testing sample that was the 65 to 90 year olds for general surgery, but the measure is for 18 to 90 year olds for general, vascular, and orthopedic surgery.**

The data on which this risk-adjustment model is based was Medicare claims for general, vascular, and orthopedic surgery patients age 65 to 90. We do not have validated models for adults below 65. Others published studies that have applied the FTR measure to populations including less than 65 years old including: Aiken *JAMA* 2002, Aiken *JAMA* 2003, Friese *Surgery* 2010, Glance *Ann Surg* 2011, Kendall-Gallagher *J Nurs Scholarsh* 2011, Kutney-Lee *Med Care* 2015, Sheetz *J Am Coll Surg* 2014, and Wright *J Clin Oncol* 2012.

8. **Page 214, Lines 6-11: There were patients who died with a complication and they also included, as in the previous measure, patients who died without documented complications. So I just have a little concern about including them in the numerator statement.**

The designation of having a complication is determined by the codes described in the Appendix. If a patient dies after surgery without a complication, we presume there was an undocumented complication. This represents a small number of patients (Silber *Med Care* 2007).

9. **Page 221-222, Lines 15-22;1-4: After the developer's last comment about the sort of finite list of things that they identified as really being impactful to the measure, like the availability of anesthesia, high tech, all of those things, if they could clarify then the pretty sophisticated risk-**

adjusted model that has 160 different characteristics, so the rationale for having a risk-adjustment model that has 160 characteristics given that it's things like the availability of a CAT scan and an anesthesiologist.

We apologize again for erroneously including the wrong model. We mistakenly reported a validity model instead of the risk adjustment model. In our presentation of the findings and definitions, we did not intend to conflate the patient characteristics for which we risk-adjust and the hospital characteristics that have been found to be associated with performance on the risk- adjusted measure. The correct risk-adjustment model includes only patient characteristics such as age, sex, comorbidities, emergency admission status, transfer-in status, DRG, and principal procedure, with interactions between variables included when statistically significant using the Bonferroni correction. Each patient characteristic was tested in a univariate model in the development sample, and retained if statistically significant at the $P < 0.15$ level Interactions between pairs of variables were also included in the final model if they were significant at the two-sided 5% level with the significance level based on the Bonferroni correction..

The updated risk-adjustment model and results can be found in Section 2b4 of the Measure Testing Form. Additionally, as this metric can be used for various populations that are very diverse, users should construct a model with coefficients that would be applicable to their study sample. Examples of coefficients that have been included in published studies are included in Table 5 (page 28) within the Measure Testing Form. There are some coefficients that are commonly used, and others that are specific to the study sample. For example, Glance *Ann Surg* 2011 adjusted for mechanism of trauma, as they were comparing low- and high-mortality hospitals caring for trauma patients.

10. **Page 222, Lines 5-15: Then also if they could please provide some more detail on how this is calculated? They're saying that it uses administrative data, but it also does seem to have some patient level characteristics in the risk adjustment. So I'm trying to understand, is there software that's been developed? How is the measure actually calculated? So, like, what is the math behind that? Particularly, the risk adjustment piece would be extremely helpful.**

For unadjusted FTR, we compute a ratio. In the numerator is the number of patients who died, and in the denominator is the number of patients who developed a complication or died with no complication (something we say is an undocumented complication—as the patient was initially alive and died after an elective surgery, so the event must have occurred and was not documented.)

Examining adjusted FTR requires some sort of O/E or (O-E)/N computation (where O=observed, E = expected, and N = the number of patients analyzed), reflecting indirect standardization, or a direct standardization computation. This can be accomplished with regression models or matching analyses, and in our citations we provide examples for both approaches.

In the simple regression approach for FTR we create a data set which includes only patients who developed a complication or who died without a complication. This population (N) includes all the patients who died and all the patients who developed a complication. We simply then run a logistic regression model with the "y" dependent variable being death (0=alive, 1 = dead) and the "x" or independent variables being all the patient characteristics. Once the regression model is fit, we can then estimate a predicted probability of dying (in this population of patients who had a complication or who died without a complication) for every patient in the data set. Then we examine each patient at each hospital who had a complication or who died without a complication and estimate a predicted probability of FTR. We sum each patient risk of FTR to obtain the E, and count the deaths (failures) to get the O (observed) so we can then compute a hospital specific O/E or (O-E)/N. Of course there are many ways to create the regressions, but this description represents the simplest indirect standardization method. Note: The risk- adjustment model we provide was produced through the more complex process described elsewhere.

An alternative approach is to make comparisons through matching, and we also cite these papers. With matching, we compare similar patients (based on patient characteristics, types of comorbidities, etc.) and examine FTR rates in these similar patients. Matching is another form of risk adjustment.

Administrative data can be used to identify patient characteristics such as age, sex, principal procedure,

DRG, comorbidities, and admission status. Patient age is defined by taking the difference of the patient's admission date and their date of birth. We use a set of 34 comorbidities that are defined by the ICD-9 codes in Appendix C (an ICD-10 version is also provided in Appendix E). The comorbidities are defined using data from the patient's claims in at least the 3 months prior to their index admission (codes from the index admission are also used for most comorbidities; please see Appendices C and E). When using Medicare data, all patient claims (inpatient, outpatient, and carrier/Part B) from the admission and lookback period are scanned for the existence of these qualifying codes. The comorbidity and complication definitions can be applied to any dataset that includes patient hospitalizations linked to longitudinal health services utilization data.

Hospital teaching status, technology status, nurse-to-bed ratio, and nurse mix are not patient comorbidities, but we define them to study the relationship between FTR and hospital characteristics, in order to establish validity. These variables can be defined using the Medicare Provider of Services (POS) file or the Healthcare Cost Report Information System (HCRIS).

The failure-to-rescue rate is simply the rate of deaths within 30 days of admission among those who had an in-hospital complication or 30-day death in the target case population. The calculation algorithm and measure length are described more detail in Section S18 of the Main form as such: "Patients admitted to an acute care facility with a stay characterized by a principal procedure and DRG of interest as outlined in the attached Appendix that can also be found on the website (http://www.research.chop.edu/programs/cor/node/26). Those patients both alive and without complications were excluded, as were any below 18 years of age or above 90 years old. Cases meeting the target process were therefore between the ages of 18-90 years old, admitted to an acute care facility for a DRG of interest, and had a complication or died without a documented complication within 30 days of admission. The event of interest is death."

11. **Page 222-223, Lines 22;1-3: Often nurse staffing is variable during a patient's stay and you may not have exactly the same skill mix every time during the day, nor every day. And where is the count being taken?**

Skill mix and other hospital characteristics are not included in the proper risk-adjustment model. This type of variation is not typically tracked in datasets to which we have access. For example, the Healthcare Cost Report Information System (HCRIS) and Medicare Provider of Service files offer annual information on these variables, but no files provide day-by-day measures of skill mix or any other hospital information that changes daily.

For studies by our group that adjusted for hospital characteristics, covariates like nurse staffing are absolute, not variable, as they are a cross-sectional study assessing the FTR rate at a given time point. Kutney-Lee et al., 2015 compared hospitals at two different time points, adjusting for hospital characteristics which were reported for each of those time points. Showing that changes in hospital characteristics over time can be accounted for in FTR rate comparisons.

12. **Page 223, Lines 12-16: One is the collection of 30 day mortality. That's not something that hospitals normally have. So is the expectation like NSQIP that you actually call all these patients and see where they are 30 days out?**

Our typical experience has been with Medicare data that includes vital status information for all Medicare beneficiaries nationwide. It is also our experience that hospital discharge records for all adult hospitalizations can be linked with vital status information available from many states, including California, Missouri, New York, and Pennsylvania. We agree that it is always better to use data that accounts for deaths outside the hospital. When such linked vital status data is not available, investigators can apply the in-hospital measure of failure-to-rescue (0352).

13. **Page 223, Lines 17-19: And then how you actually know that the complication that may have occurred in the hospital related to their death 30 days out?**

We believe there is a misunderstanding of how we use complications in the measure definition. We are not evaluating hospitals on their complication rates; we are simply using the in-hospital complications to define

a more homogenous patient group with elevated risk. As all deaths with or without a documented complication are included, any selection bias arising from differences in documentation of a complication that lead to death is negated.

Preliminary analysis comments (section location included):

1. **Page 3, Section 2a1: ICD-10 codes are NOT included in the specifications.**

We have updated the Appendix to include proposed ICD-10 codes, these can be found in Appendices D and E. These will still require testing once claims using ICD-10 are available, but we have reviewed them for clinical comparability with the ICD-9 codes.

2. **Page 3, Section 2a1: The developer notes that "when Physician Part B is available, the definition of complications and comorbidities are augmented to include CPT codes." It is therefore unclear whether CPT codes should always be used when calculating the measure.** Regardless of the data source, any time Current Procedural Terminology (CPT) codes are available, the definition of complications and comorbidities should be augmented to include them. This has been clarified in the forms.

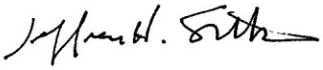
3. **Page 4, Section 2a2: The developer indicates that reliability testing was conducted at the data element level. However, no information regarding individual data elements was provided. The developer described the method of reliability testing as "split half sample correlation", which would suggest that testing at the performance measure score level was conducted. No information was provided to describe how the split-sample method was used or which correlation statistic was calculated. The reliability statistic reported was 0.32 (0.56 when a correction factor was applied), but no interpretation of these values was provided. Because information regarding the methodology was not provided, it is unclear whether the method used was an appropriate method of testing for reliability of the performance measure score and how to interpret the results.**

Our approach to reliability was based on the classic book by Lord and Novick (Lord & Novick 1968). Our approach, using the Lord and Novick methodology to present reliability was published in *Medical Care* in 2007 (Silber *Med Care* 2007). Using the same approach, we have updated the data set for this new NQF revision. We ask how reliable is the FTR measure for assessing a hospital rank. We do this by forming two random split groups inside each hospital ("wing A" and "wing B") and ranking hospitals on one wing and then the other. When we discuss reliability, we are interested in the correlation in FTR ranking between the two random hospital "wings". Further description of this methodology, as well as the results can be found in Section 2a2 of the Measure Testing form.

A highly reliable FTR measure will provide similar rankings for the same hospital across its random "wings" or the random samples of its patients. In our previous work we compared our FTR measure with the FTR measure used by AHRQ and also certified by the NQF. The Silber FTR had higher reliability. In the new models we developed on a more recent data set, we found similar results to the 2007 Medical Care paper (Silber *Medical Care* 2007).

To get a feel for the reliability we provide, it should be viewed in context to other measures. That is why in our report we compare this reliability to using the 30-day mortality rate and the AHRQ-based FTR measures (that we compute by utilizing only those patients experiencing the AHRQ-based complications). The comparisons are now found in Table 2 (page 5) of the Measure Testing form.

Sincerely,



The Nancy Abramson Wolfson Professor of Health Services Research Director,
Center for Outcomes Research
The Children's Hospital of Philadelphia

Professor of Pediatrics and Anesthesiology & Critical Care The Perelman School of Medicine
Professor of Health Care Management, The Wharton School
The University of Pennsylvania

NATIONAL QUALITY FORUM

Memo

NQF MEMBER votes are due [Month DD, YYYY] by 6:00 PM ET