

# THE NATIONAL QUALITY FORUM

## 'Hospital Care Outcomes & Efficiency' Steering Committee Meeting March 3-4, 2009

A meeting of the 'Hospital Care Outcomes & Efficiency' Steering Committee was held on March 3-4, 2009 in Washington, DC.

*Steering Committee members present:* Bruce Boissonnault, MBA (Co-Chair); Frank Opelka, MD, FACS (Co-Chair); Tanya Alteras; JudyAnn Bigby, MD (day 1); Caroline Blaum, MD, MS; Robert Bonow, MD, FACC; John Bramhall, MD, PhD; Niall Brennan, MPP (day 2); Donald Casey, MBA, MD, MPH, FACP; James Coates, MBA, MD; Amy Deutschendorf, MS, RN; Jill Fainter (by phone), RHIT; Lindsey Gilbert (phone day 1, in-person day 2); Charles Homer, MD, MPH; Clifford Ko, MD, MS, MSHS; Eliot Lazar, MBA, MD (by phone); Doris Peter, PhD; Steve Phillips, MPA; Ileana Piña, MD (in-person day 1, phone-am day 2).

*TAP members present:* Patricia Stone, PhD, FAAN (Co-Chair); Gabriel Escobar, MD (Co-Chair) (by phone day 2 pm)

*NQF Staff Present:* Helen Burstin, Karen Pace, Eric Colchamiro

*Measure Stewards Represented:* AHRQ (by phone), CMS, Health Benchmarks (by phone), Premier, 3M (by phone), The Leapfrog Group

### WELCOME, INTRODUCTIONS, AND DISCLOSURE OF INTERESTS

Mr. Boissonnault welcomed the Steering Committee members who then introduced themselves and stated any conflicts of interest<sup>1</sup>.

The purpose of the meeting was to:

- complete the evaluation of the candidate measures using the NQF standard evaluation criteria, including review of the measure evaluations and recommendations made by the Technical Advisory Panel (TAP);
- make recommendations on which measures are suitable as voluntary consensus standards (based on the results of the evaluation);
- identify additional recommendations to accompany the measures as indicated; and
- review plans for measures due for maintenance and some additional hospital measures.

---

<sup>1</sup> Frank **Opelka** – Physician Advisory Council for Amedisys and United Health Care Group; Bruce **Boissonnault** – publish measures in NY; Robert **Bonow** – executive committee of AMA PCPI; Donald **Casey** – technical advisory panel for Booz-Allen advising CMS on feasibility of use of patient safety indicators and Premier CACI, working American Board of Medical Specialites on AMI efficiency measures, ACC heart failure measures, ACP Performance Measures Subcommittee; Clifford **Ko** – works for American College of Surgeons, which runs NSQIP and investigator on contract to develop HOE-008, scientists that developed HOE-013 work with him at ACS, but he was not involved in the development those measures; Steve **Phillips** – works for Johnson & Johnson, the steward for HOE-011; Ileana **Piña** – participated in development of ACC heart failure performance measures, but not candidate measures.

## **INTRODUCTION**

After the audience and those on the conference call line introduced themselves, NQF staff provided background information.

### **Strategic Issues for NQF**

Dr. Burstin spoke of four strategic issues for NQF: driving toward high performance, shifting toward composite measures, moving toward outcomes measurement, and measuring disparities in all we do. She remarked that the measurement framework currently out for vote promotes shared accountability and measurement across patient-focused episodes of care rather than trying to isolate single entity accountability. Dr. Burstin noted the work of the National Priorities Partnership has led to the development of six core priority areas which guide NQF's work: patient and family engagement; population health; care coordination; safety; palliative and end-of-life care; and eliminating waste.

### **Consensus Development Process**

Dr. Burstin reviewed the Consensus Development Process (CDP) used by NQF to endorse measures, and noted the opportunities for NQF member and public input at both formal comment and appeal periods and the ability of members to request an adhoc review by the Consensus Standards Approval Committee (CSAC). She clarified the role of the Technical Advisory Panel (TAP) as advisory to the Steering Committee. The TAP's role is primarily to focus on scientific acceptability of the measure properties; its evaluation has been provided to you. This is an input for the Steering Committee to consider; however the Steering Committee, which is comprised of all stakeholders, will also evaluate all measures and make determinations as to which measures are recommended for endorsement as voluntary consensus standards. She emphasized that all measures, including those not recommended for endorsement, are submitted for public comment.

### **Intellectual Property**

Dr. Burstin also reviewed the new NQF Intellectual Property policy that was adopted in 2008. Under this policy, proprietary measures may be considered for NQF endorsement. Throughout the process, transparency is of paramount importance. Detailed specifications must be made available to Steering Committees and the TAP during the Consensus Development Process (CDP). While charges are allowed for complex measures, there is still a very clear, strong preference on the part of NQF's Board for measures that are free of charge, whenever possible. Although no measurement is truly without cost, charges for using the measures should be considered as one factor under feasibility, which also includes ease of measurement.

### **Hospital Outcomes Project and Work Plan for the Meeting**

Dr. Pace briefly reviewed the Hospital Outcomes project. The project is funded by CMS and the purpose is to achieve voluntary consensus on performance measures of hospital care outcomes and efficiency that are suitable for both public reporting and quality improvement. The measures before this committee fall into three categories: mortality, complications, and readmissions. NQF did not receive any submission for direct measures of efficiency (resource use paired with quality), however readmissions and complications are related to efficiency. The project timeline calls for endorsement action by July 31, 2009.

Dr. Pace reviewed the agenda and work plan for the meeting. The Committee will first have a discussion regarding methodological issues that need to be considered when evaluating outcome measures, then move into a discussion of all the candidate measures. Dr. Pace noted that Committee members have been assigned as primary and secondary reviewers for each measure

and have completed an initial evaluation of the measures prior to the meeting. The committee members' initial evaluations and the TAP's evaluation were handed out to the committee. During this meeting, Dr. Stone will review the TAP's evaluation, the assigned reviewers will present their evaluations, and the full Committee will discuss each measure and vote on the evaluation ratings for each major criterion. The measure stewards/developers were invited to attend the meeting and were provided time to make brief introductory remarks prior to the Committee's discussion as well as respond to questions that arose during its deliberations.

Dr. Pace identified that three measures were not previously reviewed by the Committee on the criterion of importance - two due to technical submission problems (HOE-017, HOE-018) and one due to changes in the submission (HOE-007) resulting in a delay of the information until after the January Steering Committee call. The TAP was asked to review those measures, however, if the Steering Committee determines they do not meet the importance criterion, they will not be further considered.

## **Measure Evaluation**

Dr. Pace reviewed the NQF measure evaluation criteria.

Importance to measure and report refers to the extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. The sub-criteria indicate how importance is reflected: high impact, gap in performance, and evidence-based.

Scientific acceptability of the measure properties refers to the extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. The sub-criteria reflect how scientific acceptability is evaluated:

- Precisely specified
- Reliability testing
- Validity testing
- Exclusions - justified
- Risk adjustment – evidence-based, factors at start of care (not related to disparities)
- Identification of statistically significant and practically/clinically meaningful differences in performance.
- Multiple data sources – comparable results
- Disparities - stratification

Usability refers to the extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. The sub-criteria address:

- Useful for both public reporting and quality improvement
- Harmonized
- Distinctive or additive value to existing measures

Feasibility refers to the extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. The sub-criteria address:

- Clinical data generated during care process
- Electronic sources
- Exclusions – no additional data source

- Susceptibility to inaccuracies identified
- Data collection strategy can be implemented
- For proprietary measures, fees associated with use of the measure

The Committee discussed questions or issues related to applying the evaluation criteria. A member observed that the outcome measures under review require statistical expertise to adequately evaluate them. NQF staff commented that not all Committee and TAP members are expected to have expertise in all areas. The TAP was convened specifically for expertise in this area.

A member questioned if a candidate measure was similar to an endorsed measure, should the endorsed measure be considered the gold standard. He went on to ask how to handle a measure that is based on an endorsed measure for which there are concerns. Dr. Pace responded that there is no presumption that the endorsed measure is better and that is an open question for the committee to decide. If an endorsed measure has been found to be problematic, the issue can be brought to NQF for an adhoc review. Another member pointed out a 2008 study that demonstrated how different measures of the same topic can provide divergent rankings and that measure harmonization and selecting one measure is less confusing.<sup>2</sup>

One member commented that measures that are not based on publicly available data (e.g., AHRQ QIs from billing and discharge data) sources allow providers to opt out by not providing data and that should be a consideration under usability and perhaps feasibility regarding whether the measure can be implemented. In addition to a framework for evaluation, there needs to be more work on the data highway including the data in electronic records for quality measurement and how that data will be available.

## **CROSS-CUTTING OUTCOME MEASURE EVALUATION ISSUES**

Dr. Stone presented an overview of some general observations and cross-cutting issues discussed at the TAP meeting on February 4-5. Some of the problems with the measure submissions that the TAP identified included issues with precision of specifications (e.g., allow use of any risk model, use of adjusted or unadjusted mortality rates in a measure, time window not specified); risk model specifications (e.g., inclusion of race and socioeconomic status and factors that occur after start of care); and other issues (e.g., little or no information on model performance, limited information on justification for risk variables, lack of risk adjustment).

### **Cross-cutting Issues**

Dr. Stone outlined the cross-cutting issues discussed by the TAP.

- Modeling
  - Hierarchical models are appropriate, but not the only way to address clustered data and small case volumes; may give a false sense of making a precise estimate for small volume providers
  - Important to provide the confidence intervals and sample size information for outcome measures
- Use of administrative claims data to identify the outcome

---

<sup>2</sup> Rothberg MB, et. al. "Choosing the Best Hospital: The Limitations of Public Quality Reporting." *Health Affairs*. 2008;27;1680-1687.

- Dr. Stone clarified that the TAP did not have an overall objection to using administrative claims data for outcome measures. It recommended one based on claims and did not recommend some others. The issue was whether there was any attempt to validate the identification of the outcome.
- The TAP recognized that the limitations of administrative claims data are accepted and used in many measures.
- When used to identify the outcome of interest, inaccuracies are more problematic and the validity of the score results must be assessed to determine if the variability in scores represents variability in the outcome vs. variability in coding practices
  - Present on admission (POA) coding is promising, but new and needs reliability testing
  - Differences in number of codes present in different data sets affect results
- Rare occurrence
  - Need information on the rates expected and whether they are large enough to be able to detect significant differences
  - Rates per patients may not be able to discriminate
- Time window
  - Must be specified
  - Minimize confounding with differences in discharge/transfer practices
- Actionability of outcome measures
  - Variability in risk-adjusted outcomes suggests opportunity for lower quality hospitals to improve
  - Outcome measures indicate area for improvement; require provider to investigate specific actions
  - Composite endpoints (e.g., global complications/readmissions) are more challenging for usability
    - Subsume many different complications/readmissions
    - Require ability to “drill-down” to identify area for improvement
    - Consumers also may find specific information more useful for decision-making

Dr. Stone reviewed the TAP’s initial recommendations that will be further developed when the TAP is reconvened in the future to provide more guidance on evaluating outcome measures. Some things to consider for the future include the following.

- Provide more specific guidance and examples.
- Ask for a context or framework for the submitted measure, the purpose, and how the measure was developed.
- Provide a checklist of risk model performance metrics to be reported.
- Provide a reference database that could be used to run the proposed measure and risk models for claims-based measures.
- Ask for comparison to other models for the outcomes cited in the literature.
- Ask for justification for the risk variables included/excluded from the model.
- Clarify the different types of reliability: reliability of the data elements and reliability (precision) of the score produced by the measure, which pertains to true vs. random variation.
- Be more specific of expectations for reliability and validity testing.
- Ask for information on the rates expected and whether they are large enough to be able to detect significant differences.
- Consider pre-screening measure submission by statisticians.

- Measure and model testing should use the most current data available.
- Ask for information/data on improvement in a proposed outcome.

The Steering Committee participated in a discussion of the issues presented by the TAP. A committee member commented on the important statistical issues of risk modeling and uncertainty around a point estimate of the outcome, especially in light of the fact that many of the data are not normally distributed. He expressed concern that the uncertainty about the scores needs to be transparent when reported. A committee member asked if the TAP recommended a particular risk adjustment method. Dr. Stone responded that the TAP did not rule out any of the adjustment methods that were presented in the measures reviewed.

The committee discussed the pros and cons of Present on Admission (POA) coding. Most agreed it should be helpful in distinguishing complications that arise after the start of care; however, some pointed out it's new in most states and may not be a reliable variable. Others noted that coding improves when it is known that it will be used such as in quality measures. Some expressed concern that too much emphasis on POA coding can take away from focusing on what is most important for that admission. A question was asked about whether the AHRQ QI measures, some of which are endorsed by NQF, incorporate POA coding. A response was received that the AHRQ measures allow for computation with or without POA coding.

The Steering Committee discussed the benefits and limitations of administrative data. One committee member questioned whether the paper record should be considered the gold standard and that he's heard of hospitals training how to document in the paper record similar to training regarding coding. The committee had a discussion about coding practices and agreed that improving coding is appropriate and there's a distinction between unethical practices. Another member commented that there are many safeguards and quality checks as well as criminal penalties for fraudulent coding and that should not be a consideration in using a particular data source. One member noted a benefit of using administrative data is that all providers are included and cannot opt out by not submitting data. He opined that the limitations of administrative data are overrated and the impact of not having all providers represented is underrated. Dr. Stone reiterated that the TAP was not against using administrative data and did not even consider intentional inaccuracies – it was evaluating whether there was any evidence for the NQF criteria of reliability and validity.

A committee member questioned whether measures could be identified for what audience and purpose they are intended. Dr. Burstin noted that measures endorsed by NQF are considered suitable for both public reporting and quality improvement; NQF does not endorse measures considered only appropriate for internal QI purposes.

A committee member commented that he thought the technical quality expected by NQF is higher than in his previous experience several years ago. Dr. Burstin agreed and noted that the updated criteria were intended to improve measurement, however, the criteria are guidelines and the committee will need to weight the various aspects and importance of the various criteria. Another committee member also expressed that another difference is that outcome measures have more technical specifications and are more challenging and complex than process measures. The committee also discussed that measure recommendations should not require perfection.

The Committee discussed the actionability of outcome measures. Some members expressed the opinion that many outcome measures are not actionable because they don't indicate exactly what should be done to improve and thought that evidence of improvement should be required. Others agreed that variability in risk-adjusted rates indicate that those with poorer performance can improve. Some were hesitant about measures with time windows beyond the immediate hospital admission that are influenced by other factors and would require collaborative efforts with other provider settings to improve.

## COMMENTS

NQF members and public audience members were given the opportunity to make comments. The following points were addressed.

- Comments on the PCI 30-day mortality measures (HOE-009/HOE-010) included: probabilistic matching should not be used when the measures are implemented; cardiogenic shock needs validation; and outpatients should not be excluded. A letter from the Society for Cardiovascular Angiography and Interventions was distributed to the Steering Committee.
- The suggestion of "gaming" should not be discussed in relation to providers and performance measures.
- Administrative data won't get better until we start using it. The TAP suggestions of a standard test data set to facilitate comparisons and to focus on the most important elements of a measure or for global measures the components that contribute most to the score of measures are worthwhile.
- NQF should consider implementation issues when evaluating and endorsing measures.

## EVALUATION OF INDIVIDUAL MEASURES

The evaluation of each measure began with an introduction by the measure steward assessment, followed by a presentation of the evaluation by the primary and secondary reviewers, followed by discussion and voting by the entire Committee. Questions that arose were referred to the measure stewards/developers. The following table provides the Committee ratings and recommendations and a summary of the major points related to its evaluation. Six measures were recommended by the Committee to advance in the consensus process.

The measure stewards/developers were invited to present introductory remarks on their measures to the Committee. The following representatives did so, and also were available during the discussion of the individual measures to respond to questions that arose.

- Dr. Bruce Hall, from the Washington University and Barnes Hospital in St. Louis; HOE-015, the LEB bypass mortality/complications measure developed in conjunction with CMS
- Dr. Jephtha Curtis (and Dr. Harlan Krumholz via telephone), from the Yale-New Haven Health System; HOE-009 and HOE-010 PCI 30-day mortality measures developed in conjunction with CMS
- Dr. Judy Chen, Director of Clinical Development, Health Benchmarks; HOE-004, the Risk Adjusted 30-day Readmission Rate for Heart Failure
- Barbara Rudolph, Director of Leaps and Measures for the Leapfrog Group; HOE-013, Survival Predictors (6 individual mortality measures – CABG, AVR, PCI, AAA, Esophagectomy, Pancreatectomy)
- Eugene Kroch, Chief Scientist with Premier; HOE-006 and HOE-018, Inpatient Comorbidity Adjusted Morbidity Index and Inpatient Comorbidity Adjusted Complication Index

- Larry Westfall, Johnson & Johnson, HOE-011, Measure of the Occurrence of Deep-vein thrombosis/ Pulmonary embolism Following Hip or Knee Replacement Surgery

### Competing Measures

Several of the candidate measures are similar to NQF-endorsed measures. The committee grappled with how to compare so as to make its recommendation. Some of the full specifications had been obtained and were handed out at the meeting, however not all of them. None of the competing candidate measure information included an analysis of how the measure results were different. The committee discussed potential conditional recommendations based on further analysis or response to a series of questions; however, the specific analyses and questions were not identified. Ultimately, the committee voted on recommendations with the information it had at hand.

After the meeting, an internal NQF review was conducted and it was determined that more information was needed to put forth a recommendation for competing measures. NQF recognized that even though we do not have further analysis from the measure submitters, the Steering Committee was not provided all the specifications on the endorsed measures. NQF notified the stewards of both the candidate and endorsed measures that the Steering Committee would need to further evaluate the measures at a minimum comparing the specifications to fully understand if the measures were competing or complementary. The information will be provided to the Steering Committee and it will be discussed on a conference call on April 13.

Candidate Measures	Endorsed Measures	Competing Candidate Measures
HOE-013 Survival Predictor (6 individual mortality measures - CABG, AVR, PCI, AAA, Esophagectomy, Pancreatectomy) (Leapfrog Group)	<p>NQF# 0359, Abdominal Aortic Artery (AAA) Repair Mortality Rate (IQI 11) (risk adjusted) (AHRQ)</p> <p>NQF# 0360, Esophageal Resection Mortality Rate (IQI 8) (risk adjusted) (AHRQ)</p> <p>NQF# 0365, Pancreatic Resection Mortality Rate (IQI 9) (risk adjusted) (AHRQ)</p> <p>NQF# 0120, Risk-Adjusted Operative Mortality for Aortic Valve Replacement (STS)</p> <p>NQF# 0133, PCI mortality risk-adjusted (ACC)</p> <p>NQF# 0119, Risk-Adjusted Operative Mortality for CABG (STS)</p>	<p>HOE-009-08 30-day all-cause risk-standardized percutaneous coronary intervention (PCI) mortality rate for patients without ST segment elevation myocardial infarction (STEMI) and without cardiogenic shock (Centers for Medicare and Medicaid Services)</p> <p>HOE-010-08 30-day all-cause risk-standardized Percutaneous Coronary Intervention (PCI) mortality rate for patients with ST segment elevation myocardial infarction (STEMI) or cardiogenic shock (Centers for Medicare and Medicaid Services)</p>
HOE-004-08 Risk-Adjusted 30-Day Readmission Rate For Heart Failure (Health Benchmarks, Inc)	NQF# 0330, 30-Day All-Cause Risk Standardized Readmission Rate Following Heart Failure Hospitalization risk adjusted (CMS)	

During the course of voting on the complications measures, it was clarified that the vote was on recommending measures for endorsement, not just to advance for comment. All measures whether recommended or not will be presented for public comment.

# THE NATIONAL QUALITY FORUM

Hospital Outcomes & Efficiency  
 Technical Advisory Panel - February 2009  
 Steering Committee - March 2009

## Summary of Review of Measures

**NQF Evaluation Criteria:** I=Importance to measure and report; S=Scientific acceptability of measure properties; U=Usability; F=Feasibility  
**Importance to measure and report:** this is a threshold criterion and the Committee votes: Y=yes, N=no, or A=abstain. Measures that do not pass the importance criterion are not further evaluated and not recommended for consensus standards.

**Remaining Criteria:** Extent to which the NQF evaluation criteria are met: H=high; M=moderate; L=low. The Committee votes or reaches consensus on ratings.

**Recommendation:** The Committee/TAP votes on the overall recommendation for endorsement: Y=yes, N=no, or A=abstain.

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
HOE-011-08 Measure of the Occurrence of deep-vein thrombosis/pulmonary embolism (DVT/PE) Following Hip or Knee Replacement Surgery (Johnson & Johnson Health Care Systems, Inc.)	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Yes (SC)    <b>S:</b> H-0;M-0;L9-;A-    <b>U:</b> cannot determine    <b>F:</b> cannot determine  <b>Recommend for Time-Limited Endorsement:</b> Y-0;N-9;A-0  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> S: Measure is untested. The measure is intended to identify treatment for DVT/PE 30 days after discharge; however the specifications do not provide any detail for ambulatory coding and linking index hospitalization to post-discharge hospital and ambulatory claims. No risk adjustment strategy is planned because the steward states that DVT/PE is considered "preventable for all patient risk profiles"; however, in item #19 of the submission form, the steward identified factors associated with disparate outcomes including cancer, obesity, age, previous VTE, oral contraception, which indicates the need for some method of risk adjustment or risk stratification.  F: Only feasible if able to link claims across time and settings.</p> <p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-14;N-1;A-1    <b>S:</b> H-0;M-0;L-15;A-1    <b>U:</b> H-0;M-5;L-9;A-1    <b>F:</b> H-2; M-5;L-7;A-1  <b>Recommend for Time-Limited Endorsement:</b> Y-0;N-14;A-1  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> I: DVT is an important topic of measurement and relates to NPP goal. Information was provided on impact, but not variability in performance. Issues to address in further evaluation: it is untested and there are already many measures related to DVT; 30-day time frame  S: The SC agreed with the TAP that the measure specifications need to be more precise in order to implement and that a risk adjustment strategy is needed. A SC member discussed that the majority of events following hospitalization are clinically silent and this measure only looks at those that are identified, which is a smaller percentage. Others noted that practice guidelines are discordant. One member questioned whether this was an appropriate topic for the 30-day time window; another suggested that would encourage better surveillance and prophylaxis.  F: Health plans and CMS do have the ability to link claims across time and settings, so they would be able to implement such a measure.  The SC agreed this measure as specified was not ready for endorsement, but expressed they would like to see this measure brought</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	back to NQF in the future after testing.
<p><b>HOE-015-08</b> Postoperative Respiratory Failure (PSI #11) (Agency for Healthcare Research and Quality)</p>	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Yes (SC) <b>S:</b> H-3;M-6;L-;A- <b>U:</b> H-9;M-;L-;A- <b>F:</b> H-9;M-;L-;A-  <b>Recommend for Endorsement w/ Condition:</b> Y-7;N-2;A-0  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> S: Criterion validity is high, suggesting that the measure is identifying a high number of true positives or true events. The risk-adjustment methodology appears sound, has been used in numerous indicators and settings, and takes into account clustering within hospitals. The indicator is used specifically to examine the quality of care within a specific hospitalization, so that measurement is relatively precise. The measure has been used in several settings with comparable results and high positive predictive validity. Someone questioned whether the false negative rates had been evaluated; however others pointed out that has not been a requirement for testing and this measure has had other appropriate reliability and validity testing.  F: Use of administrative data makes feasible.  The TAP recommended this measure on the condition that the results of current validation testing are reported as soon as possible. A suggestion also was made that at the time of maintenance review, an assessment of the use the POA indicator be included.</p> <p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-15;N-1;A-0 <b>S:</b> H-10;M-6;L-0;A- <b>U:</b> H-11 ;M-5;L-0;A- <b>F:</b> H-12;M-4;L-0;A-  <b>Recommend for Endorsement:</b> Y-16;N-0;A-0  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> I: Although there was some question of the source of estimates for variability 2.3-29.2% and whether wide confidence intervals would negate much variability, because this measure is being used, the committee thought it warranted further evaluation.  The SC agreed that the measure met scientific acceptability, usability and feasibility criteria.  S: Because all measures are reviewed under maintenance on a 3-year cycle, it recommended for endorsement with any updated information provided at the time of maintenance review. In response to a question, the developer stated the validation study will address false negatives, but it is a sampling challenge to do efficiently because relatively infrequent. Also clarified that the variability data should have been 2.3-29.2 per 1000 (not percent).</p>
<p><b>HOE-017-08</b> Postoperative Stroke or Death in Asymptomatic Patients undergoing Carotid Endarterectomy (Society for Vascular Surgery)</p>	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Y-9;N-0;A- <b>S:</b> H-;M-5;L-4;A- <b>U:</b> H-0;M-2;L-7;A- <b>F:</b> H-0;M-0;L-9;A-  <b>Recommend for Time-Limited Endorsement:</b> Y-0;N-9;A-0  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> I: This measure submission had been inadvertently missed. The TAP agreed it met the importance criterion. These are important outcomes and measure also would encourage selection of appropriate patients for the procedure.  S: The measure is untested. The measure would require physician claims be linked to hospital claims in order to have the information in the G-code that indicates the patient was asymptomatic for a year prior to the procedure. Although the measure would not need risk adjustment if restricted to the asymptomatic patients, testing of the reliability and validity of the G-code, especially for under-reporting is necessary. The TAP also did not think that a cumulative lifetime rate for individual physicians was a sound approach for performance measurement and that other approaches to deal with small volume should be explored (e.g., rolling time periods).  F: G-code not yet established and G-codes not used in hospital claims.  These issues do not warrant granting time-limited endorsement – the measure should be tested and then brought back to NQF.</p> <p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-11;N-1;A-3 <b>S:</b> H-0;M-0;L-15;A- <b>U:</b> H-0;M-0;L-15;A- <b>F:</b> H-0;M-0;L-15;A-  <b>Recommend for Time-Limited Endorsement:</b> Y-0;N-15;A-0  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b></p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	<p>The SC agreed with the issues identified by the TAP.</p> <p>S: The SC discussed whether there should be some risk adjustment even among asymptomatic patients. The steward submission indicates that the measure is targeted for asymptomatic patients because practice guidelines recommend CEA only be performed in asymptomatic patients and it is incumbent on the surgeon to only select patients for this prophylactic and elective operation who will have a low stroke or death rate." Committee members identified that factors such as age and other comorbidities such as renal failure, diabetes, etc. affect risk.</p> <p>The SC expressed they would like to see this measure brought back to NQF in the future after refinement and testing.</p>
<p><b>HOE-018-08</b>            Inpatient Co-morbidity Adjusted Complication Index (Premier, Inc)</p>	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Y-8;N-0;A-1 <b>S:</b> H-0;M-8;L-1;A- <b>U:</b> H-0;M-0;L-9;A- <b>F:</b> H-1;M-5;L-3;A-</p> <p><b>Recommend for Endorsement:</b> Y-0;N-9;A-0</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> Please note that the measure steward was notified that for this project, specific measures need to be proposed and evaluated. Although the Premier classification system facilitates drilling down into the data for various levels of analyses, the measures being evaluated for potential endorsement are for total complications or total severe complications across all hospitalized patients. Measures HOE-018 and HOE-006 are both measures of complications using the same methodology. HOE-018 includes all complications; HOE-006 includes severe complications. The following comments pertain to both HOE-006 and HOE-018.</p> <p>I: Measure HOE-018 submission had been inadvertently missed and not previously reviewed by the Steering Committee. The TAP agreed it met the importance criterion.</p> <p>S: A primary concern was the replicability of the classification system. The definition of what constitutes a complication is dependent upon an evaluation of principal-secondary diagnosis pairs selected by volume and reviewed by physician panels using modified Delphi consensus techniques to determine the probability that the secondary dx is a complication rather than a comorbidity. Complications also are classified by severity on a 5-point Likert scale (A-E) by internal panels of clinicians and those rated D&amp;E are used to denote the severe complications for HOE-006. The risk adjustment model includes race and income variables, which the NQF evaluation criteria suggest should not be used in risk adjustment. The developer stated these are considered proxies for access to care. The risk model also includes valid procedure that occurs during the hospitalization ("certain procedures can serve as effective proxies for lab reports and treatment history that are not available in the current database, as well as for other unobservable critical factors.") and discharge status (which occurs after care is provided). Risk model performance metrics for a development and validation sample were not provided. The number of diagnoses included in a data source (e.g., CA-25, MEDPAR-9) can affect rates of complications unless hospitals only compared within the same data source. Only face validity is addressed, and there was no testing to determine if variability in scores reflects variability in complication rates or in coding practices. The TAP discussed that use of administrative ICD-9 codes to identify an outcome such as all complications vs. variables used in risk models, necessitates an understanding of the reliability of those data.</p> <p>U: The TAP discussed whether a global complications measure based on diagnosis codes can be used for public reporting because of the reliability and validity issues identified above, although using such methods as screening tools for quality improvement activities might be helpful. In addition, the overall scores subsume many different complications so that the type of complications could differ greatly from one hospital to another. There also was some discussion of whether a global complications measure can be used for quality improvement without data on the specific complications. However, risk-adjusted complication rates from coded data could identify situations that require further investigation. The classification system used to compute the measures also can be used in the QI investigation to identify patients or various groups of patients (e.g., by diagnosis) with the complications. The developer stated that hospitals would need to do that analysis on their own and it could be done in a simple spreadsheet.</p> <p>F: The measure is based entirely on administrative data. The measure steward plans to make the measure freely available.</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	<p>The TAP agreed that such a measure is useful for screening and the system a useful tool for QI investigations, but it is not ready for publicly benchmarking performance.</p> <p><b>Measure Steward Response:</b> The measure steward submitted an additional 42 pages of materials and a letter regarding concerns about process.</p> <p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-18;N-0;A-0 <b>S:</b> H-0;M-10;L-8;A- <b>U:</b> H-1;M-13;L-4;A- <b>E:</b> H-7;M-11;L-0;A-</p> <p><b>Recommend for Endorsement:</b> Y-5;N-11;A-2</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> The following comments pertain to both HOE-006 and HOE-018.</p> <p>S: The Delphi process may be an appropriate tool for reaching consensus, however, SC members expressed concern about 1) the lack of reliability and validity testing to verify the probabilities that a secondary diagnosis is a complication, 2) the use of a subjective process in areas where evidence is available, and 3) not using the POA indicator. The steward indicated it was planning to use POA in the future. The committee discussed that the POA is important to distinguish co-morbidities, but is not yet ready for implementation. The committee also noted the use of race and SES as risk factors and the NQF criteria suggesting stratification rather than risk adjustment. The steward indicated it found differences in complication rates by race and SES, but felt that was related to access to care rather than differences in care provided. Note: an analysis by the measure developer submitted in support of the measures states, "the process of care differences that we and others have observed suggest that some of the difference in complication rates can be attributable to systematic racial differences in the way patients are treated after admission to a hospital." (Kroch, Eugene A., et. al. "Racial Disparities in Inpatient Complication and Mortality Rates." Academy Health Presentation. June 2003)</p> <p>U: Most committee members agreed that the aggregate rate of complications is less useful than more specific measures for both consumers and providers; however, a member noted that such an aggregate score would be of interest to purchasers. The SC agreed that the system is useful for screening and detailed analysis to identify specific problems for quality improvement, but not public reporting.</p> <p>F: The measure is based on administrative claims data and does not require additional data collection. Premier has indicated that if endorsed, it would make use of the measure free of charge and will provide a mechanism for entering data online. If hospitals want detailed reports to disaggregate the data to assist with targeting quality improvement it would need to subscribe to Premier's quality improvement system.</p>
<p><b>HOE-006-08</b> Inpatient Co-morbidity Adjusted Morbidity Index (Premier, Inc)</p>	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Yes (SC) <b>S:</b> H-0;M-3;L-6;A- <b>U:</b> H-0;M-0;L-9;A- <b>E:</b> H-1;M-5;L-3;A-</p> <p><b>Recommend for Endorsement:</b> Y-0;N-9;A-0</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> See comments for HOE-018.</p>
<p><b>HOE-007-08</b> 3M™ Potentially Preventable Complications (PPCs) (3M Health Information Systems)</p>	<p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-12;N-4;A-0 <b>S:</b> H-0;M-13;L-5;A- <b>U:</b> H-1;M-13;L-4;A- <b>E:</b> H-6;M-10;L-1;A-</p> <p><b>Recommend for Endorsement:</b> Y-7;N-10;A-1</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> See comments for HOE-018.</p> <p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Y-9;N-0 <b>S:</b> H-0;M-4;L-5;A- <b>U:</b> H-0;M-6;L-3;A- <b>E:</b> H-0;M-2;L-7;A-</p> <p><b>Recommend for Endorsement:</b> Y-0;N-9;A-0</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> Please note that the measure steward was notified that for this project, specific measures need to be proposed and evaluated. Although the 3M classification system facilitates drilling down into the data for various levels of analyses, the measure being evaluated for potential endorsement is for total complications across all hospitalized patients.</p> <p>I: This measure was not previously reviewed by the SC. The TAP agreed it met the importance criterion.</p> <p>S: This measure builds on the AHRQ Patient Safety Indicators (PSI) and the Complications Screening Program (CSP). The measure will be sensitive to present-on-admission (POA) coding practices, and the developers point out that hospitals have 2 incentives to increase POAs: 1) to decrease complication rate and 2) increase severity of illness. It was developed using CA data where POA has been</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	<p>implemented. The number of diagnoses included in a data source (e.g., CA-25, MEDPAR-9) can affect rates of complications unless hospitals are only compared within the same data source. A TAP member noted that CA data tends to be quite different and that validation with other data sets from other states or a national set would be desirable. Only face validity is addressed, and there was no testing to determine if variability in scores reflects variability in complication rates or in coding practices. Risk adjustment is accomplished by indirect standardization using APR DRGs further subdivided by 4 severity of illness subclasses and 4 risk of mortality subclasses (developed by and iterative process of formulating clinical hypotheses and then testing the hypotheses with historical data). The TAP discussed that use of administrative ICD-9 codes to identify an outcome such as all complications vs. variables used in risk models, necessitates an understanding of the reliability of those data. It agreed that POA coding will assist with distinguishing complications from co-morbidity; however POA coding is relatively recent in many states and measure scores are subject to variability in coding practices.</p> <p>U: The TAP discussed whether a global complications measure based on diagnosis codes can be used for public reporting because of the issues identified above, although using such methods as screening tools for quality improvement activities might be helpful. In addition, the overall scores subsume many different complications so that the type of complications could differ greatly from one hospital to another. There also was some discussion of whether a global complications measure can be used for quality improvement without data on the specific complications. However, risk-adjusted complication rates from coded data could identify situations that require further investigation. The classification system used to compute the measures also can be used in the QI investigation to identify patients or various groups of patients (e.g., by diagnosis) with the complications.</p> <p>F: The measure is based entirely on administrative data. The measure steward intends to charge for use of the measure, which would require both the PPC system and APR DRGs (stated PPC is roughly half the cost of APR DRG).</p> <p>The TAP agreed that such a measure is useful for screening and the system useful tool for QI investigations, but it is not ready for publicly benchmarking performance.</p> <p><b>Measure Steward Response:</b> In response to the issue regarding the effect of the number of secondary diagnoses in a data set, the steward indicated that you would need to compute the reference norm for the risk adjustment method and make comparisons by data set. The steward also noted that the historical problems related to POA coding were related to inadequate guidelines and now that CMS is requiring POA coding, there has been a lot of work on laying out precise coding guidelines.</p> <p>3M submitted a letter with concerns about process.</p>
	<p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-18;N-0;A-0    <b>S:</b> H-1;M-15;L-2;A-    <b>U:</b> H-4;M-13;L-1;A-    <b>F:</b> H-2;M-13;L-3;A-</p> <p><b>Recommend for Endorsement:</b> Y-8;N-10;A-0</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b></p> <p>S: Although this measure uses the POA indicator, which helps distinguish a complication from a pre-existing co-morbidity, the SC did not think POA is ready for use in a measure suitable for public reporting. It also discussed that it expects coding to improve since it is now required on the UB 04 claim form and the coding guidelines have been improved. It was noted that although this classification system did not use a Delphi process as in the Premier measures, it is based on clinical panels deciding what was a potentially preventable complication. SC members expressed concern about the lack of reliability and validity testing to verify the that complications identified are in fact complications.</p> <p>U: Most committee members agreed that the aggregate rate of complications is less useful than more specific measures for both consumers and providers; however, a member noted that such an aggregate score would be of interest to purchasers. The SC agreed that the system is useful for screening and detailed analysis to identify specific problems for quality improvement, but did not reach consensus on its usefulness for public reporting.</p> <p>F: The measure is based on administrative claims data and does not require additional data collection. 3M will charge for the use of its</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	measure, but that also includes the entire system, information and support.
<p><b>HOE-008-08</b> Hospital specific risk-adjusted measure of mortality or one or more major complications within 30 days of a lower extremity bypass (LEB). (Centers for Medicare and Medicaid Services)</p>	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Yes (SC) <b>S:</b> H-4;M-5;L-;A- <b>U:</b> H-0;M-6;L-3;A- <b>F:</b> H-0;M-4;L-5;A-  <b>Recommend for Time-Limited Endorsement:</b> Y-8;N-1;A-0  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> S: Submission indicates not fully developed and tested and will be completed within 24 months, however development testing reported is quite extensive. Data fields are well defined, but the developer indicated reliability testing would be completed prior to implementation. The TAP questioned whether reliability would hold up when implemented outside of NSQIP's training and auditing and also noted that the risk model would need to be recalibrated. The measure steward noted that NSQIP currently captures about 90% of cases, so would expect relatively few changes. It was clarified that although the measure was developed using NSQIP database, participation in NSQIP is not a requirement for implementation. The measure has a multiple endpoints because of the low occurrence of each event individually, but is not submitted as a composite measure. The reliability of the functional status risk variable was questioned, as well as the validity of RVU as a risk factor. Others commented that accuracy of risk factors overall are less a problem than accuracy of the outcome data. Creatinine&gt;1.2 is a risk factor - should also consider code for dialysis. Developed using 3 years of data, but anticipate computing yearly rates when implemented. The presentation of interval estimates is a strength of the proposed methodology.  U: In response to a question whether rates could be improved, the developer stated that they have seen improvement in NSQIP.  F: Uses clinical data that until electronic records are available must be collected and reported (now to NSQIP registry, possibly some other mechanism). Participation in NSQIP is not a requirement. Feasibility cannot be entirely evaluated because a national data collection strategy has not yet been proposed.  The TAP recommended time-limited endorsement due to development using registry data vs. implementation intended nationally, no report about reliability testing, and need to recalibrate the risk model when implemented nationally.  <b>Measure Steward Response:</b> Response regarding RVU: Years ago in the NSQIP the program attempted to control for "procedural complexity" by creating an in-house scale of complexity developed by a panel of experts, but it became apparent that this same information was largely captured already by the CMS designation of work RVUs, with the added advantage that this was an independent body doing the assessments, and that the assessments were updated periodically. It was demonstrated within NSQIP then that the correlation between work RVUs and the in-house "complexity score" was high, and so the complexity score was dropped and the work RVUs were adopted. Again, the aim was to provide some control for procedural complexity, within or across procedure types. Thus, there is now many years of experience using work RVU as a risk adjuster within the NSQIP, and that experience was carried forward into this project. In this vascular project, wRVU continued to demonstrate explanatory value as a risk adjuster (as reflected in the submitted materials). Keep in mind, however, that this LEB measure only deals with a well-defined subset of vascular procedures; controlling for complexity of procedures is always less important within a small procedure subset than it would be for comparisons across large sets of disparate procedures. Nonetheless, the inclusion of wRVU in this vascular measure did contribute to explanatory power.</p> <p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-14;N-1;A-0 <b>S:</b> H-15; M-1;L-0;A-1 <b>U:</b> H-10;M-6;L-1;A-1 <b>F:</b> H-5;M-9;L-3;A-1  <b>Recommend for Time-Limited Endorsement:</b> Y-15;N-1;A-1  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> I: All sub-criteria were met. One committee member questioned whether it was high enough volume to be considered high impact or best for internal QI only. Another committee member thought it is also an indicator of appropriate pre-operative patient selection. In response to a question regarding variability Bruce Hall, NSQIP stated that 17% experience an event and variability of risk-adjusted predicted:expected ratio is 0.75 to 1.25.  The SC thought the measure was well developed, and concurred with much of the TAP evaluation.</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	<p>S: In response to a question about trauma patients, the steward noted trauma patients are not intended to be included and are not in the NSQIP database. A SC member questioned the reliability of functional status data and another questioned whether other risk factors might enter the model (MI past 6 months) if possibly correlated factors (SGOT, albumin) were removed. The steward responded that functional status has been used in NSQIP and is rigorously defined.</p> <p>U: In response to a comment about understandability of hierarchical modeling, several committee members agreed that understanding a methodology is not essential, as long as the data is presented in a understandable and useful way.</p> <p>F: The steward noted that participation in NSQIP is not a requirement of using this measure. The SC thought that the TAP suggestion for time-limited endorsement probably relates to feasibility of national implementation. The measure requires data after hospitalization that must be obtained from ambulatory records or patient follow-up.</p>
<p><b>HOE-009-08</b> 30-day all-cause risk-standardized percutaneous coronary intervention (PCI) mortality rate for patients without ST segment elevation myocardial infarction (STEMI) and without cardiogenic shock (Centers for Medicare and Medicaid Services)</p>	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Yes (SC) <b>S:</b> H-9;M-L;-A- <b>U:</b> H-6;M-3;L;-A- <b>F:</b> H-3;M-6;L;-A-</p> <p><b>Recommend for Time-Limited Endorsement:</b> Y-9;N-0;A-0</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> Measures HOE-009 and HOE-010 are basically the same except for the denominator populations (with or without STEMI/cardiogenic shock), which are clearly distinct, both from a clinical standpoint as well as from a data collection standpoint. The measures were discussed and voted on together. The following comments pertain to both HOE-009 and HOE-010.</p> <p>S: Submission indicates not fully developed and tested and will be completed within 24 months, however development testing was reported. Data fields are well defined, but the developer indicated reliability testing would be completed prior to implementation. It was clarified that the measure was developed using NCDR CathPCI registry database, but participation in the registry is not a requirement for implementation. The measure submitted requires matching registry data to Medicare claims and enrollment data. The developer indicated that availability of patient identifier would improve the measure through ability to link with actual outcome (rather than probabilistic matching to outcomes). The TAP agreed that probabilistic matching to endpoint would not be acceptable for a publicly reported measure. The TAP also agreed that 30-day mortality was preferable to in-hospital mortality and that the use of clinical data as in the registry is preferred to administrative data alone. The definition of cardiogenic shock needs reliability testing and may need refinement.</p> <p>U: The endpoint for this measure is easily comprehensible to both the general public as well as to clinicians; it is useful for both consumers and providers.</p> <p>F: Measure is based on an existing registry in which the majority of hospitals that perform PCI already participate and for whom feasibility is high. Those not already participating will need to allocate staff for data collection. Although all data elements are in the electronic registry, they are not currently extracted from an electronic medical record. Feasibility cannot be entirely evaluated because a national data collection strategy has not yet been proposed.</p> <p>The TAP recommended time-limited endorsement due to development using registry data vs. implementation intended nationally, probabilistic matching for testing vs. unique identifiers for implementation, need to recalibrate the risk model when implemented nationally, and no report about reliability testing for key cohort identification and risk adjustment variables.</p> <p><b>Measure Steward Response:</b> Response regarding additional testing: In our NQF applications for the PCI mortality models, we indicated we would conduct additional testing within 24 months (Page 1, Question D) because CMS plans to refit the models once a national dataset with patient identifiers is assembled for public reporting (this was probably a conservative interpretation of the question). As you know, because we were not able to use direct patient identifiers during the process of measure development, we used a probabilistic match to merge CathPCI registry data with administrative data available on the subset of Medicare patients. The characteristics of patients who matched are virtually identical to those of patients excluded from measure development because they were not matched (Table 5 of the technical report). Accordingly, we are confident that the patients in our analysis are representative of</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	<p>the larger cohort of Medicare patients. In the course of measure implementation, we will have to refit the models using direct identifiers in all PCI patients. However, we would consider these steps part of measure maintenance as opposed to measure development. We have reviewed the criteria for “adequate field testing” set forth in NQF’s guidance on time-limited endorsement and believe that the PCI measures meet these criteria. The measures were developed in a large, representative cohort of PCI patients. Specifically, we analyzed data from more than 125,000 patients undergoing PCI at more than 600 hospitals that submit data to the American College.</p> <p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-15;N-0;A-0 <b>S:</b> H-15;M-2;L-0;A- <b>U:</b> H-16;M-0;L-1;A- <b>F:</b> H-0;M-17;L-0;A-  <b>Recommend for Endorsement:</b> Y-16;N-0;A-1  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> The following comments pertain to both HOE-009 and HOE-010.  <b>I:</b> All sub-criteria met. Issues to address in further evaluation: be cognizant of other reporting initiatives and harmonization; whether shock should be separate.  <b>S:</b> The SC concurred with much of the TAP's evaluation. The two main points of discussion were about 1) including non-STEMI with the non-MI patients or with the STEMI and 2) detecting true cardiogenic shock. The stewards responded that the non-STEMI mortality rate was more similar to non-MI than STEMI and the risk model will adjust for different levels of severity; another consideration was that adding a third category would introduce problems with case volume size. It was discussed that the National Cardiovascular Registry has defined cardiogenic shock consistent with the literature, but in MA they have reviews of designation of cardiogenic shock. The steward suggested that the data could be monitored by identifying hospitals that seem to have an unusual distribution of patients with cardiogenic shock. In regards to the issue raised by SCAI regarding outpatient procedures, the steward noted that inclusion in the measures was not dependent on a hospital admission, just if the PCI was done.  The steward clarified that on the submission the future reliability testing was in reference to statistical reliability, not data reliability which has been established. Therefore, the SC recommended for endorsement rather than time-limited endorsement.  <b>Note:</b> Because there are competing endorsed and candidate measures on the same topic (NQF# 0133, PCI mortality risk-adjusted; HOE-013, Leapfrog survival predictor for PCI), the committee’s recommendations are conditional on further evaluation that HOE-009/010 are superior or provide distinctive or additive value to the existing endorsed and candidate measures.</p>
<b>HOE-010-08</b> 30-day all-cause risk-standardized Percutaneous Coronary Intervention (PCI) mortality rate for patients with ST segment elevation myocardial infarction (STEMI) or cardiogenic shock (Centers for Medicare and Medicaid Services)	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Yes (SC) <b>S:</b> H-9;M-;L-;A- <b>U:</b> H-6;M-3;L-;A- <b>F:</b> H-3;M-6;L-;A-  <b>Recommend for Time-Limited Endorsement:</b> Y-9;N-0;A-0  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> See comments for HOE-009.  <b>Measure Steward Response:</b> See comments for HOE-009.</p> <p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-15;N-0;A-0 <b>S:</b> H-15;M-2;L-;A- <b>U:</b> H-16;M-0;L-1;A- <b>F:</b> H-0;M-17;L-0;A-  <b>Recommend for Endorsement:</b> Y-16;N-0;A-2  <b>Rationale for ratings (I, SA, U, F)/recommendation:</b> See comments for HOE-009.</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
<p><b>HOE-013-08</b> Survival Predictor (6 individual mortality measures – CABG, AVR, PCI, AAA, Esophagectomy, Pancreatectomy) (Leapfrog Group)</p>	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Yes (SC) <b>S:</b> H-0;M-3;L-6;A- <b>U:</b> H-1;M-5;L-3;A- <b>F:</b> H-0;M-0;L-8;CA-1;A-</p> <p><b>Recommend for Endorsement:</b> Y-0;N-8;A-1</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> Although only one measure submission form was submitted that referred to a survival predictor and listed 12 component measures, the measure steward clarified that there are 6 separate mortality measures (CABG, AVR, PCI, AAA, Esophagectomy, Pancreatectomy). S: Although the TAP agreed that the Bayesian methodology and modeling is elegant and cutting edge, it agreed it was not ready for endorsement. The proposed composite measures are a weighted average of a facility's mortality and the expected mortality given its volume (<math>E[m]=a+b*\log[\text{volume}]</math>). Facilities with a small number of cases then get weighted more heavily towards the expected mortality given their volume. This expected mortality is likely to be higher than the mean across all facilities since lower volumes are generally associated with higher mortality. The TAP noted the controversy surrounding the volume-outcome relationship and questioned the premise of using a volume-predicted mortality rate as a component of the composite. Although the methodology employed by this measure was recently published in a prestigious journal (Medical Care), the panel noted that publication of a single article often marks the beginning, not the end, of a discussion of a controversial subject. The panel expects that this paper will trigger much discussion as well as the publication of counter-examples and critiques, and that this process will take some time before consensus is reached on the volume-outcome relationship. Another issue identified was the lack of standardization regarding risk adjustment - the specifications allow for either risk-adjusted or raw mortality rates. The developer stated risk adjustment makes no difference in predicting future risk-adjusted rate. The competing NQF-endorsed mortality measures are all risk-adjusted. U: Because there are already NQF-endorsed mortality measures for the six procedures, the question is whether these represent additive value or superior methodology. The measure steward noted that the current NQF-endorsed cardiovascular measures from STS and ACC/AHA are not currently publicly reported. The TAP did not think these measures were ready to replace the existing endorsed measures. <b>Measure Steward Response:</b> The steward stated it will use risk-adjusted rates for hospitals that have them available and the survival predictor with unadjusted rates for those that don't. The steward noted that the survival predictor did predict future risk-adjusted mortality, possibly because only including elective procedures.</p>
	<p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-14;N-0;A-1 <b>S:</b> H-1;M-12;L-2;A- <b>U:</b> H-1;M-8;L-6;A-1 <b>F:</b> H-8;M-7;L-0;A-</p> <p><b>Recommend for Endorsement:</b> Y-9;N-6;A-0</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> I: All components are NQF-endorsed so already determined to be important. Issues to address in further evaluation: weighting of components, how to sort out elective from emergent procedures (e.g., AAA). One committee member questioned inclusion of volume component in an outcome measure; and another noted the volume-outcome relationship for some conditions/procedures. It was noted that NQF has a Composite Steering Committee that has been developing evaluation criteria that will be provided to the committee and TAP The SC was divided on recommending these measures. The primary reasons for recommending these measures despite the TAP's reservations were 1) even though there is controversy in general regarding volume-outcome relationships, the steward noted that the relationship is established for these procedures and 2) they allow reporting on hospitals with small case volumes (the steward indicated that 1 case can be reported). The minority position against these measures was based on 1) agreement with the TAP on the controversy regarding volume-outcome relationships including the role of surgeon vs. hospital; and therefore should not be the premise for the rate for small volume providers, which will be characterized primarily by the volume-predicted rate and 2) the measures only provide a slight marginal benefit over the existing measures by being able to report on small volume providers and whether reporting a rate for a provider with as few as 1 case, which would be based primarily on a volume-predicted rate, is useful information. One member also commented that by removing the emergent cases, you are perhaps missing an assessment of the skill of</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	<p>those providers.</p> <p>U: The specifications for the 3 endorsed AHRQ QI measures was handed out at the meeting, but the detailed specifications for the other 3 endorsed measures had not been obtained as of the meeting time, and the SC was not able to fully compare the measures. Committee members thought that being able to measure small providers was an advantage.</p> <p><b>Note:</b> Because there are competing endorsed measures on the same topic, the committee’s recommendation is conditional on further evaluation that the candidate measures are superior to the existing endorsed measures or provide distinctive or additive value.</p> <p>Existing measures: NQF# 0359, Abdominal Aortic Artery (AAA) Repair Mortality Rate (IQI 11) (risk adjusted); NQF# 0360, Esophageal Resection Mortality Rate (IQI 8) (risk adjusted); NQF# 0365, Pancreatic Resection Mortality Rate (IQI 9) (risk adjusted); NQF# 0133, PCI mortality risk-adjusted; NQF# 0119, Risk-Adjusted Operative Mortality for CABG; NQF# 0120, Risk-Adjusted Operative Mortality for Aortic Valve Replacement. Candidate measures: HOE-009/010, 30-day All-cause Risk-standardized PCI Mortality.</p>
<p><b>HOE-004-08 RISK-ADJUSTED 30-DAY READMISSION RATE FOR HEART FAILURE (Health Benchmarks, Inc)</b></p>	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Yes (SC) <b>S:</b> H-0;M-8;L1-;A- <b>U:</b> H-0;M-7;L-2;A- <b>F:</b> H-0;M-8;L-1;A-</p> <p><b>Recommend for Endorsement:</b> Y-0;N-9;A-0</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> S: A number of issues were identified. It appears the risk models are fit to each plan rather than one that applies to all hospitals. An exclusion (hospice before &amp; up to 30 days after) and risk factor (discharge to nursing home 1-30 days after) occur after discharge and may inappropriately exclude or adjust for outcomes that are the result of care. There is conflicting information on how age is used (dichotomous, categorical). The comorbidity index includes some of the other individual risk factors (e.g., COPD, renal failure).</p> <p>U: Although this measure would apply to potentially all patients vs. the competing endorsed measure that applies only to Medicare patients, it would be limited to health plans because of the need to link claims over time.</p> <p>The TAP agreed this measure was not strong scientifically.</p> <p><b>Measure Steward Response:</b> Issue 1: The risk model should be fitted to all plans rather than to each plan.</p> <p>Response: We fit the risk model to all plans and present the results and statistics regarding the model in table 1 and 2.</p> <p>Issue 2: Discharged to nursing home should not be included as a variable for risk-adjustment.</p> <p>Response: This measure is designed to use administrative claims data only to maximize ease of use and widest adoption. Administrative claims data do not capture direct information regarding severity of heart failure; thus we used discharged to nursing home as proxy measure for more severe heart failure.</p> <p>Issue 3: Members with heart failure and on hospice before and after discharge may be inappropriately excluded.</p> <p>Response: We conceptually excluded patients who receive hospice on discharge because members on hospice most likely have end stage heart failure and would be admitted only for palliative treatment. Less than 1% of our commercially insured sample received hospice during the time period specified and including and excluding these patients made no difference in the results.</p> <p>Issue 4: There is conflicting information on how age is used.</p> <p>Response: Thank you for this feedback. We revisited the age variable in our model and determined that the best way to use age is as a continuous variable with a quadratic term to capture the nonlinear relationship (Table 1).</p> <p>Issue 5: Comorbidity index includes some of the other individual factors (e.g., COPD and renal failure).</p> <p>Response: We apologize for this misunderstanding. We used the modified Elixhauser Comorbidity Index in which we excluded CHF, renal failure, and COPD calculation of the index.</p>
	<p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Y-15;N-0;A-0 <b>S:</b> H-8;M-8;L-0;A- <b>U:</b> H-5;M-7;L-3;A-1 <b>F:</b> H-7;M-5;L-3;A-1</p> <p><b>Recommend for Endorsement:</b> Y-12;N-2;A-2</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> I: SC already agreed on importance of readmission in Phase I. Issues to address in</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	<p>further evaluation: attention to left ventricular assistive devices and transplant; comparison to other endorsed measures</p> <p>In response to a question regarding whether there was adequate case volumes of CHF patients &lt;65 who are hospitalized, a SC member stated the average age for CHF in a 3,000-patient clinical trial is 59. The steward also stated it had run its model on a dataset with Medicare Advantage patients. The committee thought a measure for all payers and all ages was appropriate.</p> <p>The primary reason for approving this measure was that it includes all payers and ages and the issues raised by the TAP were addressed. It was noted that the steward should not refer to the Elixhauser index if it has been modified.</p> <p><b>Note:</b> Because there is a competing endorsed measure on the same topic (NQF# 0330, 30-Day All-Cause Risk Standardized Readmission Rate Following Heart Failure Hospitalization risk adjusted), the committee's recommendation is conditional on further evaluation that the candidate measure is superior or provides distinctive or additive value to the existing endorsed measure.</p>
<p><b>HOE-012-08</b> 3M™ Potentially Preventable Readmissions (PPRs) (3M Health Information Systems)</p>	<p><b>TAP Measure Evaluation criteria:</b> <b>I:</b> Yes (SC) <b>S:</b> H-0;M-2;L-7;A- <b>U:</b> H-0;M-0;L-9;A- <b>F:</b> H-0;M-1;L-8;A-</p> <p><b>Recommend for Time-Limited Endorsement:</b> Y-0;N-9;A-0</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> Please note that the measure steward was notified that for this project, specific measures need to be proposed and evaluated. Although the 3M classification system facilitates drilling down into the data for various levels of analyses, the measure being evaluated is for total preventable readmissions across all hospitalized patients.</p> <p><b>S:</b> Although there is some appeal to isolating preventable readmissions, the TAP questioned the reproducibility and validity of the designation of preventable readmissions by clinical panels (the developer indicated "each of the 98,596 cells contain a specification of whether the combination of the base APR DRG for the Initial Admission and for the readmission were clinically-related and therefore potentially preventable"). A question also was raised about the stability of the empiric estimates for the PPRs based on one state (FL). The submission form indicates any risk adjustment method could be used but APR DRGs is recommended; however, one method would need to be specified to result in a standard measure. A limitation of the risk adjustment method is reliance on ICD-9 codes without POA indicator and whether can adequately distinguish what was present at the start of care from conditions that developed during care.</p> <p><b>U:</b> There is an NQF-endorsed risk-adjusted measure for all readmissions. Comparison of results and rankings from this candidate measure of preventable readmissions with the endorsed risk-adjusted all readmission measure is needed to justify the complexity of this measure.</p> <p><b>F:</b> The measure is based entirely on administrative data. The measure steward intends to charge for use of the measure, which would require both the PPR system and APR DRGs (stated PPR is roughly half the cost of APR DRG).</p> <p><b>Measure Steward Response:</b> 3M submitted a letter with concerns about process.</p> <hr/> <p><b>SC Measure Evaluation criteria:</b> <b>I:</b> Yes-see 004, Phase I <b>S:</b> H-2;M-11;L-4;A- <b>U:</b> H-2;M-12;L-3;A- <b>F:</b> H-4;M-8;L-5;A-</p> <p><b>Recommend for Endorsement:</b> Y-7;N-11;A-0</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> <b>I:</b> SC already agreed on importance of readmission in Phase I.</p> <p><b>S:</b> The SC agreed that risk adjustment, time window, and readmission to same hospital or any hospital need to be standardized in the measure specifications and should not be up to any implementer - the steward seems willing to do that. In regards to the 98,000-cell matrix, the steward noted that 2/3 were assessed as not causally related. The SC also expressed concern with reliability of the judgment process used to determine if a readmission was related to the prior admission.</p> <p><b>U:</b> The steward stated that FL is now publicly reporting on a PPR-based set of reports where in the past it used an all-cause readmission that the providers objected to and did not improve. The SC noted the utility of the "3M system" for quality improvement, but expressed concern that the specific aggregate measure submitted for consideration would not be that useful for consumers or providers; however, one SC member noted it could be useful to purchasers.</p> <p><b>F:</b> Administrative claims data are feasible to use. However, not all states have a database like FL and even though each payer would</p>

Meas# / Title/ (Owner)	TAP/Steering Committee Discussion/Evaluation
	have the data for its patients, that would not provide a complete picture for any hospital. The fees for using the measure makes the feasibility score somewhat lower, but some members did not think it was a big issue.
<b>HOE-005-08</b> Postprocedural Stroke or Death in Asymptomatic Patients undergoing Carotid Angioplasty and Stenting (Northwestern University, Society for Vascular Surgery)	<p><b>TAP Measure Evaluation criteria:</b> <u>I</u>: No (SC) <u>S</u>: <u>U</u>: <u>F</u>:</p> <p><b>Recommend for :</b></p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b></p> <p><b>SC Measure Evaluation criteria:</b> <u>I</u>: Y-4;N-11;A-0 <u>S</u>: <u>U</u>: <u>F</u>:</p> <p><b>Recommend for :</b> Not evaluated further.</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> I: Although preventing stroke and death are worthwhile goals, there are no data on this relatively new procedure and the procedure itself is not yet supported outside of clinical trials.</p>
<b>HOE-014-08</b> Postoperative Hemorrhage and Hematoma (PSI #9) (Agency for Healthcare Research and Quality)	<p><b>TAP Measure Evaluation criteria:</b> <u>I</u>: No (SC) <u>S</u>: <u>U</u>: <u>F</u>:</p> <p><b>Recommend for :</b></p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b></p> <p><b>SC Measure Evaluation criteria:</b> <u>I</u>: Y-5;N-11;A-0 <u>S</u>: <u>U</u>: <u>F</u>:</p> <p><b>Recommend for :</b> Not evaluated further.</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> I: Although relates to NPP safety goal, it is infrequent (96 deaths); there is little variability (2.3 to 2.9) and large number with 0 rate. Although it is an outcome measure, it would have little improvement impact.</p>
<b>HOE-016-08 RISK-ADJUSTED COMPLICATION LIKELIHOOD FOR SURGERIES: APPENDECTOMY AND CHOLECYSTECTOMY (Health Benchmarks, Inc)</b>	<p><b>TAP Measure Evaluation criteria:</b> <u>I</u>: No (SC) <u>S</u>: <u>U</u>: <u>F</u>:</p> <p><b>Recommend for :</b></p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b></p> <p><b>SC Measure Evaluation criteria:</b> <u>I</u>: Y-5;N-9;A-1 <u>S</u>: <u>U</u>: <u>F</u>:</p> <p><b>Recommend for :</b> Not evaluated further.</p> <p><b>Rationale for ratings (I, SA, U, F)/recommendation:</b> I: This is a fairly common procedure and relates to NPP safety goal; however, the quality problem was not demonstrated. In response to question of why these procedures were chosen, the measure steward indicated it was because they are relatively common.</p>