NATIONAL QUALITY FORUM

National Voluntary Consensus Standards for Patient Outcomes
Summary of the GI/Biliary Technical Advisory Panel Conference Call
March 9, 2010

TAP members: David Johnson, MD (chair); John Allen, MD; Karen Hall, MD, PhD; Dick Johannes, MD, MS; Brian Jacobsen, MD, MPH; Rocco Ricciardi, MD, MPH

NQF staff: Reva Winkler, Heidi Bossley, Sarah Fanta, Hawa Camara, Suzanne Theberge

Measure Steward Representatives: Kay Schwebke (Ingenix); John Bott, Patrick Romano, Jeffrey Geppert (AHRQ); Bruce Hall (ACS)

Audience: Lee Fleisher, MD (Steering Committee co-chair); Joe Brill (AGA)

Dr. Johnson began the call with welcome and introductions by the Technical Advisory Panel (TAP) members. TAP members were asked to disclose any conflict with the measures being discussed.

Dr. Reva Winkler, NQF project consultant, provided an introductory slide presentation that described

- NQF and its activities;
- The HHS funded Patient outcomes project;
- The role of the TAP;
- NQF's standard measure evaluation criteria; and
- Identifying gaps in outcomes measures.

Dr. Johnson led TAP members through discussion of the sub-criteria for the five submitted measures. Measure developers were present and responded to questions from TAP members. The rating and issues discussed are summarized in the tables that follow.

OT2-008-09 Bariatric surgery and complications during the hospitalization or within 180 days of discharge (Ingenix)

IMPORTANCE TO MEASURE AND REPORT		
1a Impact	Partial	This is an important patient safety outcome for an elective surgery. A few studies identify significant complications after bariatric surgery, but there is not a lot of data on current performance, particularly regarding regional variation and late
1b Gap; Opportunity for Improvement	Partial	

1c Relation to Outcomes	Partial	occurring complications. Data presented on commercial population—no Medicare data. The strength is the large net cast to capture many complications such as strictures and fistulae may not be captured in a shorter timeframe. The major weakness is the timeframe—using 180 days to capture each complication in a population with so many co-morbidities that predispose patients to many of these events, will be difficult to link outcome with the initial bariatric procedure. What is the background risk for these events in this high risk population? Concerns noted regarding selection bias for volume and sophistication of reports. Evidence indicates that complications can be reduced and downward trend is being seen. No USPTF grading for 180 day complications since this list was based on a single retrospective study.
SCIENTIFIC ACCEPTABILTY OF THE MEASURE PROPERTIES		
2a Specs	Minimal	Included complications are identified by claims codes. Are the definitions standardized? Very broad inclusions. How to assess ER visits or urgent care that do not require hospitalization? Developer advised the included complications were based on the study by Echinosa and filtered by prevalence and availability of claims codes. Why not include GI bleed or hernia? The final reliability testing among a 12 million member database is still pending. Standardized follow-up important, patients lost to follow-up affect reliability. The validity testing appears to be more generic for use with several measures from a single data set. No exclusions, appropriate as this a relative pure population deemed suitable for surgery. Exclusion of unrelated second surgery or hospitalization suggested to avoid confounding risk exposures. Risk adjustment would be useful and meaningful. Key issue is a large number of disease states (e.g. bacterial pneumonia, DVTs) that may have a defined incidence in the obese patients without surgery. Need to have data on the incidence for this in a comparable population without surgery to allow for true risk adjustment Standardized full use and access to care would be a key issue to insure that the reported complications were not in part biased by differences in patient/physician exposure frequency. Disparities not addressed though important disparities are known to exist, particularly access, availability of procedure, age, co-morbidities.
2b Reliability	Partial	
2c Validity	Minimal	
2d Exclusions	Minimal	
2e Risk Adjustment	Not at All	
2f Meaningful Differences	Not at All	
2g Comparability	Not applicable	
2h Disparities	Not at All	
USEABILITY		
3a Distinctive	Partial	Not currently in use. No demonstration of usability. The overall value (numerator/denominator) would be easy to compare among institutions. But the specificity of the measure is of
3b Harmonization	Not Applicable	

3c Added Value	Partial	concern given the types of events captured in the numerator over a 180 day time frame. For example, an obese person gets a DVT three months after gastric bypass while flying across the country. Is that a complication of their surgery? Will have more data as Ingenix uses the measure. Public reports without appropriate risk adjustment or unrelated to surgery may cause patients to avoid sugery.
FEASIBILITY		
4a Data a by Product of Care	Complete	Claims based measure. One worries that with so many codes to enter the numerator, would each one be accurate? A revision would need to have granular data to support the ability to capture needed data over the 180 day timeframe. How to track between healthcare systems? How to assess ER visits or urgent care or outpatient office visits that do not require hospitalization. Minimal testing in the 180 day population.
4b Electronic	Complete	
4c Exclusions	Complete	
4d Inaccuracies/Errors	Minimal	
4e Implementation	Partial	

OT2-012-09 Bariatric surgery and complications during the hospitalization or within 30 days of discharge (Ingenix)

IMPORTANCE TO MEASURE AND REPORT		
1a Impact	Partial	Much stronger than the 180 day measure since proximity to surgery makes outcome more likely to be associated with the surgery.
1b Gap	Partial	
1c Relation to Outcomes	Complete	
SCIENTIFIC ACCEPTABILTY		
2a Specs	Partial	Full reliability, testing still pending. Need to standardize the intents for assessment, What level of complication and how defined? Why not include GI bleed, hernia, or endoscopy code for control of hemorrhage? Standardized reporting variances for intervals. How to assess ER visits or urgent care that do not require hospitalization. Key issue is a large number of disease states (e.g. bacterial pneumonia, DVTs) that may have a defined incidence in the obese patients without surgery. Need to have data on the incidence for this in a comparable population without surgery to allow for true risk adjustment. This is more important for the 180 days but would provide better granularity of true impact for surgery effect even for the 30 day full use.
2b Reliability	Partial	
2c Validity	Minimal	
2d Exclusions	Minimal	
2e Risk Adjustment	Not at all	
2f Meaningful Differences	Not at all	
2g Comparability	Not applicable	
2h Disparities	Not at all	
USEABILITY		
3a Distinctive	Partial	There are no disparity stratifications and this may be relevant for interpretation of outcome. However, 30-day morbidity is a well-recognized measure.
3b Harmonization	Not Applicable	
3c Added Value	Partial	
FEASIBILITY		

4a Data a by Product of Care	Complete	More appropriate time frame. How to track between healthcare systems? Entirely from claims data. Limited experience with the measure.
4b Electronic	Complete	
4c Exclusions	Complete	
4d Inaccuracies/ Errors	Partial	
4e Implementation	Complete	

OT2-009-09 Gastrointestinal hemorrhage mortality rate (AHRQ)

IMPORTANCE TO MEASURE AND REPORT		
1a Impact	Partial	Death from GI hemorrhage is an important outcome; not clear that gains will be made with this measure. Developers stated that users of the GI hemorrhage measure desired a specific measure related to esophageal bleeding (essentially varices). This reduces the number of patients per institution substantially and lessens the impact. This is a very "resource intensive" episode of care (increasing its impact) but relatively few patients and even in these patients survival from an acute bleeding episode does not necessarily translate into longer term survival (hence lessening its impact). To substantiate the impact of this specific measures (and thus have it be considered for the final measure set), an estimate of number of patients per institution would be needed.
1b Gap	Partial	
1c Relation to Outcomes	Partial	
SCIENTIFIC ACCEPTABILTY		
2a Specs	Partial	Measure generically focuses on area of prioritized importance but is not focused. Suggested that "present on admission" codes might be valuable. Definition of GI bleed population is three groups using esophageal varicies in as a primary or secondary diagnosis. The measure developer provided an excellent overview of how the GI bleeding cases were stratified. There was no validation provided however to corroborate that the process in fact correctly identified the specific patient population of interest- need to define accuracy for extraction – no validation against charts, just against large datasets. Identification relies on a three-step process combining CPT with ICD-9 codes —this may be acceptable and accurate but prior to endorsing a measure such as this (i.e. complex) this hypothesis would need empiric proof. Risk adjustments—relatively few factors for a complex population—alcoholism very important as it affects the basic physiology of a patient, very different than other causes of liver disease. Suggest stratification on urban vs. suburban, transplant evaluation, etc.
2b Reliability	Not at All	
2c Validity	Not at All	
2d Exclusions	Minimal	
2e Risk Adjustment	Minimal	
2f Meaningful Differences	Partial	
2g Comparability	Minimal	
2h Disparities	Partial	

USEABILITY		
3a Distinctive	Partial	The measure is currently used and public reported at the 95 percent confidence interval though AHRQ doesn't follow the reporting – no information on trends. AHRQ also has an "all" GI hemorrhage measure, but due to feedback has focused this measure more narrowly due to new treatment for varicies. Consider whether this would be a good stand alone measure, or, if placed side by side with another measure that captures all ICU-level mortality, would it survive the "best in class" test. Since this is geared at in-hospital mortality, and designed to capture multiple specialties and coordination of care, a broad ICU-survivorship may be more appropriate. The other issue is that since the overall prognosis of those with variceal hemorrhage is poor, decisions about withdrawal of support (i.e., DNR) may be clinically relevant and important to capture when looking at mortality. A hospital with a good palliative care service might appear to have unacceptable mortality rates but might indeed be providing the most appropriate care.
3b Harmonization	Partial	
3c Added Value	Partial	
FEASIBILITY		
4a Data a by Product of Care	Partial	Claims based measure. The outcome is death within a specific hospitalization. The complexity of these patients make that relationship less accurate. Specific centers would be biased towards alcohol-induced variceal bleeding (large city, urban centers with substantial Medicaid or pro bono care) – these patients do much worse than variceal bleeding from viral-induced cirrhosis or cirrhosis that is compensated (with high portal pressures)—thus a clear bias towards adverse outcomes would be likely based on cause of the liver disease.
4b Electronic	Partial	
4c Exclusions	Partial	
4d Inaccuracies/ Errors	Partial	
4e Implementation	Partial	

OT2-014-09 Opioid-related symptom distress scale (Pfizer)

IMPORTANCE TO MEASURE AND REPORT		
1a Impact		**Measure documentation is insufficient for assessment.**
1b Gap		
1c Relation to Outcomes		
SCIENTIFIC ACCEPTABILTY		
2a Specs		There is no denominator defined. Not really clear on outcome—less use of opioids? More use of a competing medication? Validation was done in patients having surgery —what about
2b Reliability		
2c Validity		

2d Exclusions		extrapolation to other populations?
2e Risk Adjustment		
2f Meaningful Differences		
2g Comparability		
2h Disparities		
USEABILITY		
3a Distinctive		
3b Harmonization		
3c Added Value		
FEASIBILITY		
4a Data a by Product of Care		
4b Electronic		
4c Exclusions		
4d Inaccuracies/ Errors		
4e Implementation		

OT2-002-09 Risk adjusted colorectal surgery outcome measure (ACS)

IMPORTANCE TO MEASURE AND REPORT		
1a Impact	Complete	Range of outcome observed; measurement would help highlight outliers. Prioritized area of high impact.
1b Gap	Complete	
1c Relation to Outcomes	Complete	
SCIENTIFIC ACCEPTABILTY		
2a Specs	Complete	Excellent definitions. NSQIP based - most hospitals not reporting to NSQIP, but does not require participation in NSQIP. This is "an aggregate clinical measure" not a composite. O/E ratio method is well published. Inclusion criteria – estimated 85 percent colectomies would be captured. Low volume hospitals wouldn't qualify for the measure. Tested on NSQIP hospitals—primarily academic and VA hospitals—is it appropriate to extrapolate to all hospitals? Exclusions are clinically meaningful. Risk adjustment
2b Reliability	Partial	
2c Validity	Partial	
2d Exclusions	Complete	
2e Risk Adjustment	Partial	
2f Meaningful Differences	Complete	
2g Comparability	Not Applicable	

2h Disparities	Partial	well described—should include pre-op functional status, prior procedures, and surgical complexity. Not clear how specific risk adjustment made. What about tertiary referral bias and transfers or ischemic colitis in vascular patient or post op/transplant patients? How do you account for management variations—individual or regional—outpatient, urgent care/ED vs inpatient. Reliability for model is only fair (0.4) Note disparities issue not stratified; Applications are hospital specific—how do you account for migration to other care centers? <u>None of the outcomes have been tied to outcome impact such as length of stay, reoperation, further treatment required or graded with a severity index. What about grading of complications?</u>
USEABILITY		
3a Distinctive	Complete	Application to only 42 percent of hospitals meeting volume characteristics still referral bias adjustments? Measure developer reports that non-participants for NSQIP would still be able to participate but would have to submit separately.
3b Harmonization	Complete	
3c Added Value	Complete	
FEASIBILITY		
4a Data a by Product of Care	Partial	Weakness is reliance on clinical measurements such as fever, respiratory rate or heart rate. This information would require data abstraction, albeit this is apparently done for other measures currently in use Data extraction issues raised as a concern for impact on FTE time—although measure developer suggests that this would be 0.125-0.333 FTE
4b Electronic	Partial	
4c Exclusions	Complete	
4d Inaccuracies/ Errors	Complete	
4e Implementation	Partial	