NATIONAL QUALITY FORUM

Memo

June 17, 2013

TO: NQF Members

FR: NQF Staff

RE: Voting Draft Report: *National Voluntary Consensus Standards: Gastrointestinal/Genitourinary Endorsement Maintenance, Stage 2*

DA: June 17, 2013

Background

Gastrointestinal conditions such as cancer, acid reflux, and GERD impact a large number of Americans. These disorders not only cause symptoms and pose a heavy burden of illness, but also impact the quality and length of life, as well as work productivity. Similarly, genitourinary (GU) conditions, including urinary tract infections (UTI), cystitis, benign prostate hypertrophy (BPH), and urinary incontinence (UI) pose a heavy burden on quality of life and healthcare spending.

NQF has endorsed several consensus standards to evaluate the quality of care for topic areas related to gastrointestinal and genitourinary diseases over the last several years. Evaluation of the NQF-endorsed® gastrointestinal and genitourinary measures and consideration of new measures ensures the currency of NQF's portfolio of voluntary consensus standards.

This GI/GU measure endorsement project is a pilot of the proposed two-stage consensus development process. The Stage 1 Final Report details the evaluation of the submitted concepts against the Importance to Measure and Report criterion. The evaluation, comments and feedback received during this project specifically related to the two-stage CDP process have been addressed separately in the two-stage evaluation report.

A Committee of 15 experts reviewed seven measures; five were recommended for endorsement. The public and NQF member comment period for these measures took place from April 25-May 24, 2013.

Comments and Revised Voting Report

NQF received 45 comments from 13 member organizations and two members of the public:

Consumers – 2	Professional – 1
Purchasers – 4	Health Plans – 2

Providers – 1	QMRI – 1
Supplier and Industry – 1	Public & Community Health - 0

A table of complete comments submitted during the comment period, with the responses to each comment and the actions taken by the Steering Committee, is posted to the GI/GU project page on the NQF website, along with the measure submission forms.

The Steering Committee reviewed and responded to all comments received. Revisions to the draft report and the accompanying measure specifications are identified as red-lined changes. (Note: Typographical errors and grammatical changes have not been red-lined, to assist in reading.)

Comments and their Disposition

Comments on the General Draft Report

NQF received a number of comments on the general draft report, many of which focused on additional areas for measure development:

- Takeda Pharmaceuticals America, Inc. strongly encourage the continued evaluation of the American Gastroenterological Association (AGA) Measure *2059: Inflammatory Bowel Disease Preventive Care: Corticosteroid Sparing Therapy*. "We believe it is important that this measure be considered and endorsed at the earliest possible time to expand its use in clinical practice and enhance care for CD and UC patients."

  Action Taken:

  AGA's measure *Inflammatory Bowel Disease Preventive Care: Corticosteroid Sparing Therapy* was not submitted to Stage 2 of this project since it has not yet been tested, but the Committee agrees it is an extremely important area. NQF looks forward to reviewing it in a future project, once the measure has been tested. In addition, we will add "measures of care for inflammatory bowel disease" to the list of recommendations for future measure development.

- AHIP recommends clinical registries be used as a data source for this measure set to facilitate better data collection and longitudinal use across care settings for a larger population of patients.

  Action Taken:

  The list of future recommendations was updated to add "measures that use clinical registries as data sources".

- AmeriHealth Caritas believes that individual measurements of immunization, colonoscopies, GI bleeding or GU pathology do not capture the overall quality of care rendered nor the success of health literacy. Certainly the issues of cultural competency and patient preference are not reflected in such measures. Rather, we believe that NQF should move to a next generation of measures of quality based upon systems approaches and population engagement.

  Action Taken:

  "Measures of population engagement and systems approaches, particularly those that address issues of cultural competency and patient preferences" has been added to the recommendations list.

- Ms. Lauren Agoratus acknowledges that "Measures for kidney transplants and GI complications" are being considered as a future measure. However, as 30% or renal transplant patients experience these complications which can be life-threatening, this should be put in place immediately. Most transplant patients had no prior GI diagnosis. Renal transplant is costly and unnecessary complications not only increase costs, including emergency room use and hospitalizations unrelated to the transplant itself, but affect the recipient's quality of life. Addressing this measure will improve morbidity and mortality for kidney transplant patients and result in better health outcomes."

  Action Taken:

  The Committee agrees that measures for kidney transplants and GI complications are an important topic for measurement and the quality of care. Unfortunately, no measures were brought forward on this topic at this time. The Committee strongly recommends the development of a measure in this topic area and looks forward to reviewing it at the next opportunity.

Measure Specific Comments

Comments on Recommended Measures

2065: GASTROINTESTINAL HEMORRHAGE MORTALITY RATE (IQI #18)

This measure received comments from six organizations or individuals. Four comments from consumers and purchasers were supportive, noting that this is an outcome measure that focuses on a topic important to consumers, and that it appears to be both usable and feasible. One commenter noted the "small numbers problem" which may affect reliability, and suggested adding the secondary diagnosis to capture relevant data and increase reliability.

The American Hospital Association does not support the measure, raising concerns with the importance and the reliability:

Importance:

- "NQF-endorsed measures should focus on the most meaningful, highest impact areas with an opportunity for improvement. ... we do not believe IQI 18 is up to the task of helping to guide these efforts and provide reliable information to patients and providers. We agree that GI bleeds are common and important health problems that often warrant hospitalization, but are not confident that the data suggest it is important to measure mortality. Rather, the argument for including a GI bleed mortality measure among those that are endorsed by the NQF should convince us that GI Hemorrhage Mortality is more common than it could be if the right care was provided at the right time and in the right manner, and that by illuminating performance through measurement and reporting, we have the opportunity to spur efforts to produce better care and better outcomes."
- "The developer notes that among community hospitals in the Healthcare Cost and Utilization Project (HCUP), the risk-adjusted GI hemorrhage mortality rate was 1.94% in 2008. They also note that this rate "has steadily declined over the past 14 years, from 5.78% in 1994….to 3.02% in 2005." This decline shows significant progress in adopting new diagnostic and treatment modalities that have saved lives. Unfortunately, because the HCUP data have such a significant lag between the provision of the care and the production of the data, we have no idea if progress has hit a plateau or if it continues."
- "Further, for the Committee to know if it is important to recommend endorsement of this measure as a national standard, it needs information suggesting that the mortality rate could be lower than it is through the implementation of new strategies or more rigorous attention to implementing existing strategies. The studies and guidelines showing strategies for better diagnosis and management of bleeding that the developer has included are largely written just before or at the time that mortality began to decline precipitously, according to the submission. The developer does not offer compelling evidence that further improvement is likely to ensue from the collection and reporting of these IQI data. The measure developer fails to even address the question of how the use of a measure whose data are so out of date by the time numbers are produced is capable of informing the public or guiding improvement efforts. This critical question of how can we steer a clear path forward toward improved care and outcomes for patients by looking in the rearview mirror is the most essential question to be answered about this and all of the HCUP IQI / PSI measures, and the one that is simply unaddressed in this application."
- We believe hospitals should continue to take steps to minimize the risk of harm to patients with GI bleeding. An ongoing focus on interventions that more quickly identify and provide appropriate care to hospitalized patients with GI hemorrhages should result in a continued decline in mortality rates. Thus, an NQF-endorsed measure in this area does not appear to be warranted at this time.

Measure reliability

- While the developer presents reliability testing data in their submission, they do not provide final risk-adjusted performance scores. They also do not provide a recommendation on a minimum number of cases needed to reliably report the measure and compare results across multiple hospitals. We believe such information should inform the committee's determination of whether a measure is suitable for NQF endorsement, especially given that NQF-endorsed measures often become publicly reported.

- Available evidence suggests that IQI 18's reliability in a public reporting application is poor. In 2012, Mathematica conducted a study on behalf of CMS assessing the reliability of claims-based measures used in several CMS programs, including IQI 18. (Reference A) The CMS-commissioned study defines reliability of outcomes measures as "…the extent to which variation in the measure is due to variation in quality of care rather than random variation due to the sample of cases observed."
  To determine the reliability of IQI 18, the CMS-commissioned study uses a "reliability weight" in the AHRQ measure calculation software. This weight is equal to the ratio of the variance in scores between hospitals to the total variance divided by the number of observations. In this case, "total variance" is the sum of the variance in scores between different hospitals and the variance within a hospital's score.
- The study defines the "lower limit of moderate reliability" as R=0.4. The study shows that IQI 18 has a median reliability of R=0.12 using the same amount of data (12 months) as the measure submission. Even with 24 months of data, reliability improves only to R=0.22. With 24 months data, only 25% of hospitals would have a case size large enough to meet R=0.4. A measure that fails to meet even the lower limit of moderate reliability when applied in a public reporting program should not receive NQF endorsement.

Measure Developer Response:

"The reliability of a measure is actually an attribute of the measure when applied to a particular population. What is meant by "the reliability of a measure" is the average reliability across a set of hospitals for a particular population. The particular population in the CMS-commissioned study was Medicare fee-for-service patients only, which is only a portion (indeed, a minority at many hospitals) of the total denominator eligible population. It is not surprising therefore that applying IQI 18 to a smaller population of patients results in risk-adjusted rates with less reliability. The results in our submission reported an average reliability across 4,000 community hospitals and an all-payer (including uninsured) population of R=0.47, which exceeds the threshold proposed by the commenter. In addition, the average reliability alone does not determine the value of a measure for purposes of public reporting. Rather, we use the hospital level reliability as a "shrinkage weight" to calculate each organization's performance score. The usefulness of this reliability-adjusted performance score for purposes of public reporting is discussed below. The shrinkage approach adopted by both AHRQ and CMS in its risk-adjusted outcome measures obviates the need for a minimum volume threshold by "shrinking" performance scores for small hospitals toward the overall mean value.

Potential opportunities for improvement related to GI hemorrhage mortality were extensively addressed in AHRQ's Stage 1 submission. Specific opportunities noted at that time included:

1. Prompt recognition of gastrointestinal hemorrhage as the cause of a patient's symptoms, necessitating inpatient admission for further evaluation and treatment.

2. Prompt assessment of the severity of the patient's hemorrhage and the associated risk of mortality, to guide initial decisions about where to admit the patient and how much nursing care to provide.

3. Appropriate stabilization of acutely ill patients with prompt but safe administration of fluids, blood products, vaspressors, and other resuscitative maneuvers.

4. Appropriate diagnostic and evaluation processes to identify the source of bleeding and to characterize the risk of rebleeding.

5. Appropriate monitoring by nurses, physicians, and other health professionals to identify early warning signs of clinical deterioration and to implement "rapid response" as appropriate.

6. Appropriate treatment of high-risk bleeding sources with pharmacologic and procedural interventions that have been demonstrated to reduce the risk of re-bleeding and transfusion requirements.

7. Appropriate timing of transfer from the intensive care setting to the regular unit setting, with appropriate handoffs to ensure that all important information is transmitted and that the care plan is continued and modified as needed.

In the Stage 1 submission, 16 references were provided to clinical practice guidelines, observational studies, and randomized controlled trials on the topic of GI hemorrhage management and mortality.

Hospitals may download the AHRQ Quality Indicator software and calculate the IQI 18 rate on the hospital's patient population in real-time (or as soon as an abstract of discharge data are available). The capacity to calculate baseline rates and to evaluate the impact of current interventions is an important component of usability of IQI 18 for purposes of quality improvement. The data suggest that hospitals will find opportunity for improvement (see attached Table 8). Using the reliability adjusted performance scores, our estimate is that 24.7% of IQI 18 events are potentially preventable, if all patients selected hospitals that performed at the benchmark level of performance (defined as the 20th percentile in the probability score distribution).

The recent trend data suggest that performance on IQI 18 may, in fact, have reached a plateau. The 2008 reference population used to estimate measure prevalence in Version 4.4 of the AHRQ QI software had an observed rate of 2.46%. The 2010 reference population used in the recently released Version 4.5 (May, 2013) has an observed rate of 2.41%. However, disparities across hospitals persist, and hospital performance scores are persistent over time (see data above), meaning that past performance is predictive of current performance (and that past performance is in fact more predictive of current performance than other hospital attributes such as case volume or overall transfer-out rate). Thus, the data suggest that the performance scores provide useful information to consumers and other stakeholders. "

Action Taken:

The Committee discussed both the comments and the developers' responses. They agreed that while the measure will miss some cases that have GI hemorrhage as the secondary code, restricting the measure to the primary diagnosis code allows for a

greater degree of confidence in those being counted. Committee members acknowledge that miscoding is possible. Committee members discussed the small numbers issue and noted that the shrinkage methodology is intended to account for this issue. The Committee did not change their recommendation on the measure.

0658: ENDOSCOPY/POLYP SURVEILLANCE: APPROPRIATE FOLLOW-UP INTERVAL FOR NORMAL COLONOSCOPY IN AVERAGE RISK PATIENTS

This measure received six comments; all were supportive. Commenters noted the measure's usability and feasibility, and applauded the focus on reducing unnecessary care and decreasing costs. Two comments suggested it be linked with 0659 and reported as a paired measure.

Action Taken:

After review of the comments, the Committee did not change their recommendation for this measure.

0659: ENDOSCOPY/POLYP SURVEILLANCE: COLONOSCOPY INTERVAL FOR PATIENTS WITH A HISTORY OF ADENOMATOUS POLYPS- AVOIDANCE OF INAPPROPRIATE USE

This measure received six comments; only one was supportive. The supportive comment noted the concerns with the measure, but that "we should not let the perfect be the enemy of the good", as there is wide variability with the follow up recommendations, and that this is a decent interim measure.

Other comments noted that while the measure focus—reducing overuse—is strong, the measure has numerous problems. Commenters noted the concern that the broad exclusions allow for "gaming" and provider manipulations of results, and were very concerned about hiding poor care by not including information from prior colonoscopies. One commenter stated that "With the growth of electronic health records and the mandate for care coordination, it is not acceptable to approve to colonoscopy without including information about previous colonoscopies."

Action Taken:

The Committee acknowledged that the measure is not as strong as it could be and looks forward to a better measure in the future. After review of the comments, the Committee did not change their recommendation.

0635: CHRONIC LIVER DISEASE - HEPATITIS A VACCINATION

This measure received seven comments. Four did not support the measure, noting that it is a "strict process measure" that reflects a "standard of care". These commenters noted that while it may be feasible, it would not improve care or outcomes. Commenters were also concerned with the usability, noting that it is only submitted as a " national population

level" measure and noted the Committee's uncertainty as to what entity would be accountable with this measure.

Another comment supported the concept but noted that this measure would be subject to data issues due to incomplete health-plan clairms data records. Another negative comment also raised the issue of incomplete records for patients who change health plans, as well as noting this measure is subject to a small numbers problem, raising reliability issues. An additional comment supported the concept of hepatitis A vaccination for patients with chronic liver disease, but did not comment on the measure under consideration.

A commenter suggested potential overuse of vaccination since those with positive antibodies might receive the vaccination to keep the measure results high.

Action Taken:

The Committee discussed the level of analysis and suggested that the measure is used by health plans and the testing data was presented for health plans. The developer agreed to add health plan as a level of analysis. The Committee also clarified that this measure is not an eMeasure in HQMF format. When asked, the developer was not able to provide any data on the frequency of positive immunity. The Committee acknowledges the issues raised in the comments but did not change their recommendation of the measure.

0098: URINARY INCONTINENCE: ASSESSMENT, CHARACTERIZATION, AND PLAN OF CARE FOR URINARY INCONTINENCE IN WOMEN AGED 65 YEARS AND OLDER

Comments Received

Six comments were received on this measure. Of those, five were not in support of the measure, noting that it is a "check the box" measure that is a "standard of care process." Commenters noted the limited usability (only in women over 65) and the complexity of the multiple numerators and denominators as additional reasons not to support the measure. The measure received one supportive comment from the American Urological Association that the measure was rigorously developed, focuses on an area of importance in an elderly population, and is reliable and valid.

Action Taken:

The Committee generally agreed with the commenter that this measure currently shows "topped out" performance but noted that the data does not reflect the majority of providers. The PQRS program is currently designed to allow providers to choose which measures to report on and rewards for satisfactory reporting. Fewer than 1% of providers currently choose to report on this measure and it is likely this self-selecting sample does not reflect the broader provider population. In fact, results from other quality measures such as 0030, suggest rates of screening and plan of care for urinary incontinence are much lower (59% and 35%

respectively). These rates suggest that when the PQRS program is more widely spread this measure will identify a significant quality gap in the provision of "standard" care. The Committee did not change their recommendation on this measure.

Comments on Measures Not Recommended

0622: GERD - UPPER GASTROINTESTINAL STUDY IN ADULTS WITH ALARM SYMPTOMS

One commenter stated that ""patients with GERD and with alarm symptoms are a very small population" and that "The sensitivity of the practice to identify cancers in patients with alarm symptom is about 67%." I agree with "While the Committee agreed the evidence submitted was insufficient, there was agreement that they would exercise the evidence exception to continue to review the concept, since the quality, quantity, and consistency of the evidence would support this measure focus if provided." Therefore, I think that this should be further investigated and addressed as soon as possible."

Action Taken:

The Committee agrees this measure covers an important aspect of quality of care, and the measure did pass the importance criteria (using the evidence exception.) However, the measure submitted to the Committee for review was not tested, and the specifications were not precisely specified. The measure as submitted did not pass either the reliability or validity criteria, so it did not meet the must-pass criterion of Scientific Acceptability. The Committee encouraged the developers to revise and test the measure and submit it for review in the future.

C 2056: COLONOSCOPY QUALITY INDEX COMPOSITE MEASURE CONCEPT

NQF received multiple comments from consumers and purchasers expressing concern that the measure was not recommended.

Action Taken:

The measure concept was evaluated in Stage 1 of this 2-stage pilot project. The measure was not evaluated in Stage 2 because it did not meet the checklist requirements to enter Stage 2 (see **Stage 1 report** for details). The CSAC will discuss the concerns raised by these comments during their in-person meeting on July 10-11, 2013.

NQF Member Voting

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

Please note that voting concludes on July 9, 2013 at 6:00 pm ET – no exceptions.