



TO: Consensus Standards Approval Committee (CSAC)

FR: Reva Winkler, Suzanne Theberge

RE: Gastrointestinal/Genitourinary Stage 2 Member Voting

DA: July 2, 2013

The CSAC will review recommendations from the GI/GU Stage 2 project at its July 10-11 in-person meeting. This memo includes a summary of the project, recommended measures, and themes identified from and responses to the public and member comments.

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](#).

## STAGE 2 REVIEW

Developers whose concepts were approved in Stage 1 have up to 18 months to fully develop, specify and test the measure before submitting it for Stage 2 review against the remaining NQF criteria. Eight measures were submitted for this Stage 2 review by January 11, 2013. One measure was withdrawn by the developer before Committee review began.

Stage 2 review includes:

- **Checklist Review:** A checklist was provided to each developer following the Committee's review and recommendations for concepts in Stage 1. Each checklist summarizes the Committee's recommendations for improving the concept(s) and considerations for specifying and testing the measure that must be addressed prior to submission to Stage 2.
- **Evaluation of approved concepts:** Fully specified measures for which the concepts were approved in Stage 1 are evaluated against the remaining criteria of scientific acceptability, usability, and feasibility.

Member voting on these recommended measures will end on July 9, 2013. The voting results will be provided to the CSAC at the July 10 meeting.

Accompanying this memo are the following documents:

1. [GI/GU Stage 2 Draft Report](#). The draft report has been updated to reflect the changes made following Steering Committee discussion of public and member comments. [The complete draft report](#) and supplemental materials are available on the project page.
2. [Comment table](#). Staff has identified themes within the comments received. This table lists 45 comments received and the NQF/Steering Committee responses.

### CSAC ACTION REQUIRED

Pursuant to the CDP, the CSAC may consider approval of 5 candidate consensus standards.

#### GI/GU Measures Recommended for Endorsement:

- [0098 Urinary Incontinence: Assessment, Characterization, and Plan of Care for Urinary Incontinence in Women Aged 65 Years and Older](#)
- [0635 Chronic Liver Disease - Hepatitis A Vaccination](#)
- [0658 Endoscopy/Polyp Surveillance: Appropriate follow-up interval for normal colonoscopy in average risk patients](#)
- [0659 Endoscopy/Polyp Surveillance: Colonoscopy Interval for Patients with a History of Adenomatous Polyps- Avoidance of Inappropriate Use](#)
- [2065 Gastrointestinal Hemorrhage Mortality Rate \(IQI #18\)](#)

#### GI/GU Measures Not Recommended

- [0622 GERD - Upper Gastrointestinal Study in Adults with Alarm Symptoms](#)

#### Approved Concepts That Did Not Pass Checklist Review

- [C 2056 Colonoscopy Quality Index](#)

### DRAFT REPORT

The GI/GU Draft Report presents the results of the evaluation of six measure concepts that were approved in Stage 1 and met the checklist requirements that have been considered under the CDP. Five are recommended for endorsement as voluntary consensus standards suitable for accountability and quality improvement; one was not recommended. The measures were evaluated against the 2011 version of the [measure evaluation criteria](#).

### GI/GU ENDORSEMENT MAINTENANCE, 2013 SUMMARY

	MAINTENANCE	NEW	TOTAL
Measures considered	5	2	7
Withdrawn from consideration	1		1
Recommended	4	1	5
Not Recommended	1		1
Reasons not Recommended	Scientific Acceptability- 1		
Approved Concepts that did not pass Checklist Review		1	1

## COMMENTS AND THEIR DISPOSITION

NQF received 45 comments from 13 organizations (including 12 member organizations) and 2 individuals pertaining to the general draft report and to the measures under consideration.

A [table of comments](#) submitted during the comment period, with the responses to each comment and the actions taken by the Steering Committee and measure developers, is posted to the [GI/GU project page](#) under the Public and Member Comment section.

### **Comment Themes and Committee Responses**

Comments were forwarded to the developers, who were invited to respond. At its review of all comments, the Steering Committee had the benefit of developer responses. Committee members focused their discussion on measures or topic areas with the most significant and recurring issues.

### ***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. These were added to the list of recommendations for future measure development.

### ***Measure Specific Comments***

#### **Comments on Recommended Measures**

#### ***2065: GASTROINTESTINAL HEMORRHAGE MORTALITY RATE (IQI #18)***

*Description:* 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. "

*Steering Committee Response:* 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.

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

*Description:* 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.”

*Steering Committee Response:* 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**

*Description:* 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 claims 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.

*Steering Committee Response:* 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. When asked, the developer was not able to provide any data on the frequency of positive immunity. The Committee acknowledged 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**

*Description:* 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.

*Steering Committee Response:* 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 wider use of the measure will identify a significant quality gap in the provision of "standard" care. The Committee did not change their recommendation on this measure.

**C 2056: COLONOSCOPY QUALITY INDEX COMPOSITE MEASURE CONCEPT**

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

*Steering Committee Response* 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 the in-person meeting on July 10-11, 2013, following the project discussion.

**REMOVE ENDORSEMENT OF MEASURES**

Two measures previously endorsed by NQF have not been re-submitted, withdrawn from maintenance of endorsement, or not recommended for continued endorsement:

Measure	Description	Reason for removal of endorsement
0030: Urinary Incontinence Management in Older Adults - a. Discussing urinary incontinence, b. Receiving	This is a patient-reported measure collected through the Health Outcomes Survey with two rates that address management of	Withdrawn by developer – will be replaced by updated measure in the future





Measure	Description	Reason for removal of endorsement
urinary incontinence treatment – A patient reported measure	urinary incontinence in older adults. Discussing urinary incontinence: Percentage of patients 65 years of age and older who self-report having a urine leakage problem in the last six months and who discussed their urinary leakage problem with their health care provider. Receiving urinary incontinence treatment: The percentage of patients 65 years of age and older who self-report having a urine leakage problem in the last six months and who received treatment for their current urine leakage problem.	
0622: Upper Gastrointestinal Study in Adults with Alarm Symptoms	The percentage of patients with in the overall and high risk population with gastroesophageal reflux disease (GERD) with alarm symptoms who have had an upper gastrointestinal study. (2 separate Denominators)	Not recommended – did not pass scientific acceptability

**Measure Evaluation Summary Tables**

**LEGEND: Y = Yes; N = No; H = High; M = Moderate; L = Low; I = Insufficient**

<p><b>0098 Urinary Incontinence: Assessment, Characterization, and Plan of Care for Urinary Incontinence in Women Aged 65 Years and Older</b></p> <p><a href="#">Measure Submission Form</a>   <a href="#">Specifications</a></p> <p><b>Status:</b> Maintenance, Original Endorsement: May 01, 2007</p> <p><b>Description:</b> This is a clinical performance measure which assesses whether women age 65+ were provided appropriate treatment for urinary incontinence (UI). This measure has three rates:          (A) Assessment for UI: Percentage of female patients aged 65 years and older who were assessed for the presence or absence of urinary incontinence within 12 months.          (B) Characterization of UI: Percentage of female patients aged 65 years and older with a diagnosis of urinary incontinence whose urinary incontinence was characterized at least once within 12 months          (C) Plan of Care for UI: Percentage of female patients aged 65 years and older with a diagnosis of urinary incontinence with a documented plan of care for urinary incontinence at least once within 12 months</p> <p><b>Numerator Statement:</b> This measure has three rate. The numerator for each of the rates is as follows:          (A)Assessment for UI: Patients who were assessed for the presence or absence of urinary incontinence within 12 months          (B)Characterization of UI: Patients whose urinary incontinence was characterized at least once within 12 months          (C)Plan of Care for UI: Patients with a documented plan of care for urinary incontinence at least once within 12 months</p> <p><b>Denominator Statement:</b> There are two denominators for the rates in this measure.          (A) Assessment of UI: All female patients aged 65 years and older who visited and eligible provider in the measurement year          (B&amp;C) Characterization and Plan of Care for UI: All female patients aged 65 years and older with a diagnosis of urinary incontinence who visited an eligible provider in the measurement year.</p> <p><b>Exclusions:</b> Documentation of medical reason(s) for not assessing the presence or absence of urinary incontinence within 12 months</p> <p><b>Adjustment/Stratification:</b> No risk adjustment or risk stratification N/A N/A</p> <p><b>Level of Analysis:</b> Clinician : Group/Practice, Clinician : Individual, Clinician : Team</p> <p><b>Type of Measure:</b> Process</p> <p><b>Data Source:</b> Paper Medical Records</p> <p><b>Measure Steward:</b> National Committee for Quality Assurance</p>
<p><b>STAGE 1</b></p>
<p><b>STAGE 1 PRE-REVIEW MEMBER COMMENTS (August 7-21, 2012)</b></p> <ul style="list-style-type: none"> <li>• None</li> </ul>
<p><b>STAGE 1 STEERING COMMITTEE MEETING (August 27-28, 2012)</b></p> <p><b>1. Importance to Measure and Report:</b></p> <p><b>1a. High Impact: H-15; M-0; L-0; I-0</b></p> <p><b>Discussion:</b> Similar impact as discussed in #0030. General agreement that incontinence is a high</p>



**0098 Urinary Incontinence: Assessment, Characterization, and Plan of Care for Urinary Incontinence in Women Aged 65 Years and Older**

impact area.

**1c. Evidence**

**13:** Yes, body of evidence meets guidance for quantity, quality, consistency

**0:** No, body of evidence does not meet guidance for quantity, quality, consistency

**2:** No, inadequate information to rate quantity, quality, consistency of body of evidence

**Discussion:** The Committee was concerned that the evidence presented by the developer indicated that incontinence should be treated but did not provide evidence that documentation in the medical record improved incontinence. Some expressed concern about the link between this process measure and patient outcomes. However, the Committee ultimately agreed this measure meets the evidence criteria since existing literature does link discussion with the provider about urinary incontinence to improved outcomes.

**1b. Performance Gap: H-7; M-8; L-0; I-0**

**Discussion:** While PQRS data does not show a performance gap, the Committee agreed that there is overall low performance and low reporting based on the data submitted.

**Recommendations to Developer for Stage 2:**

- eMeasure specifications are strongly recommended.
- Consider the addition of an option for patient choice of no treatment.
- Expand age group to include commercial and menopausal population.

**Stage 1 Steering Committee Recommendation for Approval of Concept: Y-14; N- 1**

**STAGE 1 MEMBER & PUBLIC COMMENT (September 26 – October 25, 2012)**

**Member & Public Comments:**

- Commenters were concerned that this measure was not closely linked to an outcome and would not be meaningful to all stakeholders.
- Suggestions were also made to pair the measure with an outcome measure or create a composite measure that accounts for urinary continence assessment, characterization, and plan of care.
- Commenters also desired a more inclusive age range.

**Committee response:**

- The Committee agrees that measures closely linked to an outcome will provide the most meaningful information to stakeholders. Additionally, the Committee recommended that the measure age range be expanded to include the commercial and menopausal population.

**STAGE 1 CSAC REVIEW (November 7-8, 2012)**

- **Decision:** Concept Approved, with the requirement that the Steering Committee recommendations must be addressed in the stage 2 submission.

**STAGE 1 BOD REVIEW (November 29 – December 11, 2012)**

- **Decision:** Concept Ratified

**STAGE 2**

**STAGE 2 PRE-REVIEW MEMBER COMMENTS (March 4-18, 2013)**

- Commenters stated that this measure is a standard of care, and a “check-the-box measure”, and were not supportive.
- Other concerns were raised around the threshold of “any urinary incontinence,” noting that it was too strict, especially in the population of women 65 years and older.



**0098 Urinary Incontinence: Assessment, Characterization, and Plan of Care for Urinary Incontinence in Women Aged 65 Years and Older**

- While commenters noted the clinical importance of this topic, they raised several concerns about the feasibility, stating that it would require significant manual chart review or supplemental codes and suggested an eMeasure may be more feasible.
- Other commenters were concerned about the one measure having three rates, noting that this could be confusing, and that it is not clear how the three rates would be used together to assess provider performance.

**STAGE 2 STEERING COMMITTEE CHECKLIST REVIEW (March 2013):**

Checklist recommendations satisfactorily addressed: **Y-12; N-1; A-1**

Move forward to full Stage 2 review: **Y-13; N-0; A-1**

- The Committee had requested the developer add an option for patient choice of no treatment as a response. The developer explained they were unable to make that change in the limited time between the two stages, and also that they did not want to include patient refusal of treatment as a separate option because that can lead to “gaming” of measure reporting; they feel that the option of reassessing at a follow up visit would cover this issue. A Committee member suggested that not doing so meant the measure is not patient-centered and would not change outcomes, and that a patient choosing to opt-out is different than a reassessment and follow up; the developer explained this is consistent with other measures where patient refusal can be abused.
- The developer did not make additional checklist changes (adding eMeasure specifications and expanding the age group) due to funding and resource constraints. Committee members agreed this was a reasonable response and not a “fatal flaw” to keep the measure from moving forward.

**STAGE 2 STEERING COMMITTEE REVIEW (April 3, 2013)**

**2. Scientific Acceptability: The measure meets the Scientific Acceptability criteria**

*(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)*

2a. Reliability: **H-0; M-12; L-1; I-0** 2b. Validity: **H-0; M-11; L-2; I-0**

- The measure has three different indicators, but it is not a composite measure. Three separate ratings are reported.
- The measure is based on chart review.
- The Committee questioned the definition of assessment, i.e., what are the abstractors looking for? The developer responded that the assessment is just documentation of “yes or no” documented in response to “Have you had any urine leakage?” The characterization rate looks for more detail of the type of incontinence and more of the formal workup of the urinary incontinence
- The measure was tested for reliability at the data element level (numerator). All three rates had high Kappa scores.
- A systematic assessment of face validity was presented. The developer noted that they also assessed critical data elements of validity during the assessment of reliability.
- The analysis of exclusions indicated a very small number of patients are excluded.
- The measure is not stratified for disparities.

**3. Feasibility: H-3; M-10; L-0; I-0**

- The required data elements for this measure are routinely generated and used during care delivery and the measure is currently being used in PQRS.



**0098 Urinary Incontinence: Assessment, Characterization, and Plan of Care for Urinary Incontinence in Women Aged 65 Years and Older**

- However, the Committee noted some concern with capture of the data element(s) needed for the numerator.

**4. Use & Usability: H-2; M-8; L-3; I-0**

- This measure is used in at least one accountability application – in PQRS.
- Committee members noted that the performance data presented indicates little discrimination above the 10-25<sup>th</sup> percentile.
- The Committee agreed that there is not a lot of room for improvement among the providers currently reporting on this measure; the Committee agreed with the developers that there is likely to be more room for improvement once the measure is in wider use in 2015.

**5. Related and Competing Measures**

- No related or competing measures noted.

**Stage 2 Steering Committee Recommendation for Endorsement: Y-13; N-0**

**Public & Member Comment**

**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.

**Developer response:**

- We agree with the commenter that this measure currently shows "topped out" performance. However, we believe these rates are biased and do 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.
- We agree with the commenter that this measure is not usable for the commercial population. This measure was designed as part of a measure set focused solely on the geriatric population. Although Urinary Incontinence is prevalent among a younger population, it is more prevalent among older women. We will examine how the age range on this measure can be expanded to include a younger population of women, however this change was not feasible in the timeframe of measure review.

**Committee response:**

- The Committee agreed that this measure does have the potential to be a “check-the-box” measure, but thought that especially for the assessment portion, further testing and evaluation of the 3 components in a broader population will be important.
- While a urinary incontinence measure does apply to a broader population, older women do have higher rates of incontinence and expanded measures could be developed in the future.



**0098 Urinary Incontinence: Assessment, Characterization, and Plan of Care for Urinary Incontinence in Women Aged 65 Years and Older**

- The Committee agrees with the comment by the developer that the current reported rates from users of the PQRS system may represent a self-selected group and show high rates of meeting the current standard of care. When implemented in a broader setting, this measure may indicate a gap in quality care.
- Committee members reviewed the comments and the developer response and did not change their recommendation.

**0635 Chronic Liver Disease - Hepatitis A Vaccination**

[Measure Submission Form](#) | [Specifications](#)

**Status:** Maintenance, Original Endorsement: Dec 04, 2009

**Description:** The percentage of adult patients with chronic liver disease who have received a hepatitis A vaccine

**Numerator Statement:** Patients with chronic liver disease who have received a hepatitis A vaccine.

**Denominator Statement:** All patients, ages 18 and older, diagnosed with chronic liver disease

**Exclusions:** Specific Exclusions: 1. Patients with a previous history of viral hepatitis 2. Patients who report an allergy to Hepatitis A vaccine A. General exclusions: 1. Evidence of metastatic disease or active treatment of malignancy (chemotherapy or radiation therapy) in the last 6 months; 2. Patients who have been in a skilled nursing facility in the last 3 months (this exclusion is included to avoid holding physicians who care for patients during a transitional period, e.g. temporary SNF placement, for their ongoing care; hence, the time limitation of 3 months).

**Adjustment/Stratification:** No risk adjustment or risk stratification No risk adjustment necessary None

**Level of Analysis:** Population : National, Health Plan

**Type of Measure:** Process

**Data Source:** Other

**Measure Steward:** ActiveHealth Management

**STAGE 1**

**STAGE 1 MEMBER COMMENTS (August 7-21, 2012)**

- *America's Health Insurance Plans* - While this measure can be calculated using administrative data, there may be challenges with assessing the numerator at the health plan level in instances where patients have received the vaccination but who have also changed health plans.

**STAGE 1 STEERING COMMITTEE MEETING (August 27-28, 2012)**

**Recommendations to Developer for Stage 2:**

- The numerator is inconsistent with title of measure; consider changing the title of the measure to more closely align with the measure focus.
- There could be a potential validity issue in stage 2 with the assumption this concept makes that if a person was tested, they were positive and received the vaccination. Consider how to address this issue.
- Understanding there are differences in data sources, harmonize with #0399 under review in the NQF Infectious Disease project
  - **Developer Response:** Developers acknowledged and agreed to the suggested changes.



<b>0635 Chronic Liver Disease - Hepatitis A Vaccination</b>
<b>Stage 1 Steering Committee Recommendation for Approval of Concept: Y-15; N- 0</b>
<b>STAGE 1 MEMBER &amp; PUBLIC COMMENT (September 26 – October 25, 2012)</b> <b>Member &amp; Public Comments:</b>
<ul style="list-style-type: none"> <li>No comments received</li> </ul>
<b>STAGE 1 CSAC REVIEW (November 7-8, 2012)</b>
<ul style="list-style-type: none"> <li><b>Decision:</b> Concept Approved with the requirement that the Steering Committee recommendations must be addressed in the stage 2 submission.</li> </ul>
<b>BOD REVIEW (November 29 – December 11, 2012)</b>
<ul style="list-style-type: none"> <li><b>Decision:</b> Concept approval ratified</li> </ul>
<b>STAGE 2</b>
<b>STAGE 2 PRE-REVIEW MEMBER COMMENTS (March 4-18, 2013)</b>
<ul style="list-style-type: none"> <li>Two commenters stated that this is a standard of care process measure, and did not think it was useful to improve health care and outcomes.</li> </ul>
<b>STAGE 2 CHECKLIST REVIEW (March, 2013):</b>
Checklist recommendations satisfactorily addressed: <b>Y-11; N-1; A-2</b>
Move forward to full Stage 2 review: <b>Y-12; N-0; A-2</b>
<ul style="list-style-type: none"> <li>The Committee requested that the developer work with AMA PCPI on harmonization for a related measure. The two developers have indicated they are working towards this.</li> <li>The Committee requested the numerator be changed to no longer allow testing for antibodies to be sufficient. The developer made this change. However, one Committee member suggested they would also need to be removed from the denominator if the results were not available, in order to prevent doctors from being inappropriately penalized.</li> <li>The revised measure specifications are not yet tested.</li> </ul>
<b>STAGE 2 STEERING COMMITTEE REVIEW (April 8, 2013)</b>
<b>2. Scientific Acceptability: The measure meets the Scientific Acceptability criteria</b> <i>(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)</i>
2a. Reliability: <b>H-0; M-11; L-0 I-0</b> 2b. Validity: <b>H-0; M-8; L-3; I-0</b>
<ul style="list-style-type: none"> <li>The developer clarified that that chronic liver disease is defined by the diagnosis codes from claims or health information exchange (HIE) in the past year, and that non-alcoholic fatty liver disease is included as a chronic liver disease. The included codes are provided in a spreadsheet that accompanied the measure submission.</li> <li>While the Committee agreed that there are multiple guidelines recommending hepatitis A vaccine in people with chronic liver disease, they were concerned that if the measure is also recommending that people with fatty liver disease receive the vaccination, this would widen the population. Committee members were also concerned that this measure would lead to overuse of the hepatitis A vaccination because while there are no serious adverse effects from the vaccine, the vaccine is less likely to be effective in advanced liver disease. Committee members agreed that despite these concerns the measure was still appropriate.</li> </ul>
<b>3. Feasibility: H-2; M-6; L-3; I-0</b>
<ul style="list-style-type: none"> <li>The measure is already in use and the submission form included data from this use. The Committee noted the results of this measure are consistent with similar data reported by the Veterans' Administration.</li> <li>The measure uses a proprietary analytic engine, so the Committee questioned whether other organizations would be able to implement it. The developers confirmed that the measure is</li> </ul>



**0635 Chronic Liver Disease - Hepatitis A Vaccination**

reproducible by others, and that the specifications and code sets are publically available. Other users may alter the measure, but ActiveHealth reviews all modifications, generally based on data availability, to make sure the measure is aligned with the original intent, and the modified measure would not be considered an NQF-endorsed measure.

**4. Use & Usability: H-1; M-8; L-2; I-0**

- ActiveHealth reports that they are planning to report yearly, national results on their website.
- The Committee was unsure how useful this measure is as a “national population” measure.
- The Committee noted a potential for over-vaccination of non-susceptible individuals.

**5. Related and Competing Measures**

- There is one related measure, 0399: Hepatitis C: Hepatitis A Vaccination (AMA-PCPI); the developers are working on harmonization. (See Appendix D)

**Stage 2 Steering Committee Recommendation for Endorsement: Y-10; N-1**

**Public & Member Comment**

**Comments received:**

- 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 claims 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.

**Developer response:**

- As we move towards a more electronic-data-based healthcare world, the aforementioned gap of data collection should start to diminish. Our algorithms use data from health information exchanges, provider feedback, as well as patient-reported data to help fill this gap.
- The previous version of this measure included Hepatitis A immunity testing in the numerator. This was removed at the request of the NQF due to coding limitations. Because codified HAV antibody results are reported as reactive or nonreactive, as opposed to a numeric value, use of HAV Ab LOINC codes are not possible at this time in the measure algorithm. It would be possible to modify the measure to use feedback or self-reported data indicating immunity to hepatitis A in either the denominator exclusions or as a part of the numerator. However, because such modifications would require retesting of the measure and data elements, further modifications are not possible at this time. AHM remains open to reasonable change requests, given adequate time and notice, during future endorsement maintenance or update cycles.

**Committee response:**

- 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.





**0635 Chronic Liver Disease - Hepatitis A Vaccination**

- 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.

**0658 Endoscopy/Polyp Surveillance: Appropriate follow-up interval for normal colonoscopy in average risk patients**

[Measure Submission Form](#) | [Specifications](#)

**Status:** Maintenance, Original Endorsement: Jan 17, 2011 , Time-limited status not yet removed

**Description:** Percentage of patients aged 50 years and older receiving a screening colonoscopy without biopsy or polypectomy who had a recommended follow-up interval of at least 10 years for repeat colonoscopy documented in their colonoscopy report.

**Numerator Statement:** Patients who had a recommended follow-up interval of at least 10 years for repeat colonoscopy documented in their colonoscopy report

**Denominator Statement:** All patients aged 50 years and older receiving screening colonoscopy without biopsy or polypectomy

**Exclusions:** Documentation of medical reason(s) for not recommending at least a 10 year follow-up interval (eg, above average risk patient, inadequate prep)

**Adjustment/Stratification:** No risk adjustment or risk stratification N/A We encourage the results of this measure to be stratified by race, ethnicity, gender, and primary language, and have included these variables as recommended data elements to be collected.

**Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team

**Type of Measure:** Process

**Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Registry

**Measure Steward:** American Medical Association - Physician Consortium for Performance Improvement (AMA-PCPI)

**STAGE 1**

**STAGE 1 PRE-REVIEW MEMBER COMMENTS (August 7-21, 2012)**

- None

**STAGE 1 STEERING COMMITTEE MEETING (August 27-28, 2012)**

**1. Importance to Measure and Report:**

**1a. High Impact: H-15; M-0; L-0; I-0**

**Discussion:** There is general agreement this concept meets the high impact criterion.

**1c. Evidence**

**15:** Yes, body of evidence meets guidance for quantity, quality, consistency

**0:** No, body of evidence does not meet guidance for quantity, quality, consistency

**0:** No, inadequate information to rate quantity, quality, consistency of body of evidence



**0658 Endoscopy/Polyp Surveillance: Appropriate follow-up interval for normal colonoscopy in average risk patients**

**Discussion:**

- There is a significant amount of evidence to support this measure focus.
- There was discussion on whether the 10 year interval specified in this concept is based on evidence or consensus. Most polyps > 1 cm in diameter appear to grow for 5-10 years before becoming colorectal cancer. Usefulness of an interval beyond 10 years has not been studied. Committee members noted that prospective studies have demonstrated that very few patients (< 3%) have advanced adenomas when colonoscopy is repeated 5 years after a normal screening colonoscopy. Evidence in the submission form was not graded, but it is supported in the guidelines.

**1b. Performance Gap: H-15; M-0; L-0; I-0**

**Discussion:** Based on the data provided for this maintenance measure, the Committee agrees there is still a performance gap and an opportunity for improvement.

**Recommendations to Developer for Stage 2:**

- Rather than measuring whether the appropriate interval was recommended, consider specifying the measure, for example, patients aged 60 years or older receiving a screening colonoscopy who are documented to have had their last screening colonoscopy 10 or more years prior. Implementing these changes would make the measure closer to an outcome measure that would be more impactful. The Committee recognized that to implement a prospective outcome measure is difficult based on availability of data.
- Account for patients aged 50 years and older receiving a screening colonoscopy that had a recommendation to repeat colonoscopy in 1 year or less due to poor bowel cleansing.
- Consider adjusting the upper age limit for older patients, including inflammatory bowel disease, and better define "above average risk".
- Clarify in the specifications whether the exceptions are included in the denominator or should be calculated as a separate measure.
- Due to the differences in populations and the measure focus, harmonization between this concept and 0659 will not be needed.

**Stage 1 Steering Committee Recommendation for Approval of Concept: Y-15; N- 0**

**STAGE 1 MEMBER & PUBLIC COMMENT (September 26 – October 25, 2012)**

**Member & Public Comments:**

- Commenters were concerned that the measure exclusions provide loopholes for providers to manipulate the measure results.
- The lack of information about previous colonoscopies may hide evidence of poor care.
- Pairing this measure with an appropriate outcome measure would make it more meaningful to all stakeholders.

**Committee response:**

- The Committee agreed that manipulation of results through gaming is a concern; however, the specific medical reason for the exclusion must be documented through the use of CPT-II codes. The Committee recommended that the specifications for the exclusions include a specific list of the types of medical reasons that are acceptable for this exclusion when the measures are submitted for the Stage 2 measure evaluation.

**STAGE 1 CSAC REVIEW (November 7-8, 2012)**

- **Decision:** Concept Approved with the requirement that the Steering Committee



**0658 Endoscopy/Polyp Surveillance: Appropriate follow-up interval for normal colonoscopy in average risk patients**

recommendations must be addressed in the stage 2 submission.

**STAGE 1 BOD REVIEW (November 29 – December 11, 2012)**

- **Decision:** Ratification of concept approval

**STAGE 2**

**STAGE 2 PRE-REVIEW MEMBER COMMENT (March 4-18, 2013)**

- Several supportive comments were received for this measure, noting that it was both useable and feasible, and would improve the safety and quality of care.
- One commenter did not support the measure without the addition of an upper age limit of 75 years, as recommended in the March 2013 Guidance Statement of the American College of Physicians (ACP).

**STAGE 2 STEERING COMMITTEE CHECKLIST REVIEW (MARCH 2013)**

Checklist recommendations satisfactorily addressed: **Y-12; N-1; A-1**

Move forward to full Stage 2 review: **Y-13; N-0; A-1**

- The Committee's checklist recommendations included suggesting that the measure be re-specified to be a look-back measure, but the developer thought it would be an undue burden on the referring and performing physician 10 years later. The Committee also requested some clarifications and suggested an adjustment to the upper age limit and the exceptions. The developer will review these recommendations when the measure is due for their internal review, and also clarified that the exceptions are not included in the denominator if there is a valid medical reason. They encourage exception rates to be reported as well, but do not require it.
- During the checklist review, the Committee clarified that the measure looks at whether the colonoscopist knows that the patient's last normal colonoscopy was 10 years ago (which they agreed was important) but they had actually wanted the developer to add whether the provider recommends a repeat colonoscopy in ten years if the results are normal. The developer does not have the ability to do so at this time, but the Committee agreed the measure should move on to Stage 2.

**STAGE 2 STEERING COMMITTEE REVIEW (April 3 & 8, 2013)**

**2. Scientific Acceptability: The measure meets the Scientific Acceptability criteria**

*(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)*

2a. Reliability: **H-4; M-7; L-0; I-0** 2b. Validity: **H-0; M-10; L-1; I-0**

- The Committee agreed the results of reliability testing of the measure score were straightforward with good results.
- Specifications for an eMeasure were submitted. The developer did not provide a crosswalk for the codes between the eMeasure and the original measure specification. Committee members pointed out several differences between the specifications, particularly which procedures are included as a "screening colonoscopy" and conditions listed as high risk. The Committee was concerned this would lead to different results from each version of the measure. The developer agreed that the measures should be consistent and will bring back a crosswalk of both specifications for review by the Committee at the post-comment call. Final recommendation of the eMeasure will wait until that review.
- Validity was tested by a systematic assessment of face validity.
- In response to the comment from ACP about the lack of an upper age limit, the Committee



**0658 Endoscopy/Polyp Surveillance: Appropriate follow-up interval for normal colonoscopy in average risk patients**

acknowledged that ACP and the US Preventive Services Task Force do not recommend screening colonoscopy beyond age 75 years (USPSTF Grade C recommendation for ages 76-85 years and D recommendation for >85 years.) However, Committee members noted that the guidelines from the three GI professional societies do not include an upper age limit. The measure is aligned with the guidelines of the professional societies. Some Committee members were concerned about potential harms to elderly patients.

**3. Feasibility: H-0; M-13; L-0; I-0**

- This measure is reported through data collection for the American Gastroenterological Association (AGA) outcome registry. The committee agreed it is feasible and is currently in use.

**4. Use & Usability: H-0; M-13; L-0; I-0**

- Committee members noted that older specifications have two different calculations, one for reporting on the measure and one for performance. Committee members were concerned that the two calculations would yield very different percentages, leading to confusion. The developer explained this modification was to allow the measure to be implemented in PQRS. NQF staff clarified that the measure being evaluated for endorsement is the percentage of patients aged 50 and older receiving screening colonoscopy with or without biopsy or polypectomy with a recommended follow up of at least ten years as specified above.

**5. Related and Competing Measures**

- No related or competing measures noted.

**Steering Committee Recommendation for Endorsement: Y-13; N-0**

- 
- After reviewing the crosswalk between the eSpecifications and the original measure coding, the Committee agreed the two versions were aligned. The Committee voted 11-0 to also recommend the eMeasure for endorsement.

**Public & Member Comment**

**Comments received:**

- 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. However, the developer responded this was not possible:
  - **Developer response:** Although 0658 and 0659 are both intended to reduce unnecessary colonoscopies, there remain differences in populations and the measure focus between the two measures and therefore linking them together would not be feasible.

**Committee response:** The Committee did not change their recommendation.



**0659 Endoscopy/Polyp Surveillance: Colonoscopy Interval for Patients with a History of Adenomatous Polyps- Avoidance of Inappropriate Use**

[Measure Submission Form](#) | [Specifications](#)

**Status:** Maintenance, Original Endorsement: Jan 17, 2011 , Time-limited status not yet removed

**Description:** Percentage of patients aged 18 years and older receiving a surveillance colonoscopy, with a history of a prior colonic polyp(s) in previous colonoscopy findings, who had an interval of 3 or more years since their last colonoscopy

**Numerator Statement:** Patients who had an interval of 3 or more years since their last colonoscopy

**Denominator Statement:** All patients aged 18 years and older receiving a surveillance colonoscopy, with a history of a prior colonic polyp(s) in previous colonoscopy findings

**Exclusions:** Documentation of medical reason(s) for an interval of less than 3 years since the last colonoscopy (eg, patients with high risk for colon cancer, last colonoscopy incomplete, last colonoscopy had inadequate prep, piecemeal removal of adenomas, or last colonoscopy found greater than 10 adenomas)

Documentation of system reason(s) for an interval of less than 3 years since the last colonoscopy (eg, unable to locate previous colonoscopy report)

**Adjustment/Stratification:** No risk adjustment or risk stratification N/A We encourage the results of this measure to be stratified by race, ethnicity, gender, and primary language, and have included these variables as recommended data elements to be collected.

**Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team

**Type of Measure:** Process

**Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Registry

**Measure Steward:** American Medical Association - Physician Consortium for Performance Improvement (AMA-PCPI)

**STAGE 1**

**STAGE 1 PRE-REVIEW MEMBER COMMENTS (August 7-21, 2012)**

- None

**STAGE 1 STEERING COMMITTEE MEETING (August 27-28, 2012)**

**1. Importance to Measure and Report:**

**1a. High Impact: H-15; M-0; L-0; I-0**

**Discussion:** There is general agreement this measure focus addresses a high impact area as it is one of the most overused procedures.

**1c. Evidence**

**14:**Yes, body of evidence meets guidance for quantity, quality, consistency

**0:** No, body of evidence does not meet guidance for quantity, quality, consistency

**1:** No, inadequate information to rate quantity, quality, consistency of body of evidence

**Discussion:**

- The Committee discussed the length of screening intervals and the yield of identifying adenomas.
- The Committee reviewed evidence cited in the guidelines that was not specifically provided by the measure developer. Based on this review, the Committee determined that there is high quality of evidence demonstrating that these are appropriate intervals, and that the expected benefits are consistent.
- The interval specified in the measure does not match the recommendations in the evidence 3+ years versus 5 years



**0659 Endoscopy/Polyp Surveillance: Colonoscopy Interval for Patients with a History of Adenomatous Polyps- Avoidance of Inappropriate Use**

**1b. Performance Gap: H-4; M-10; L-0; I-1**

**Discussion:**

- While the PQRS data does not suggest a performance gap, few physicians reported on this measure. However, the Committee did not believe that the submitted data is representative of the likely performance gap. The use of EHRs for this measure could demonstrate a larger performance gap. PQRS also only takes patients 65 years and older, so it is not capturing patients in the commercial population.

**Recommendations to Developer for Stage 2:**

- The developer should expand on the available evidence and on the details of the meta-analysis to better demonstrate the body of evidence available to support this measure focus.
- eMeasure specifications should be submitted in stage 2.
- The interval specified in the measure does not match the recommendations in the evidence 3+ years versus 5 years; consider how these can be aligned to ensure the measure is evidence-based.
- Due to the differences in populations and the measure focus, harmonization between this concept and 0658 will not be needed.

**Stage 1 Steering Committee Recommendation for Approval of Concept: Y-15; N- 0**

**STAGE 1 MEMBER & PUBLIC COMMENT (September 26 – October 25, 2012)**

**Member & Public Comments:**

- Commenters were concerned that the measure exclusions provide loopholes for providers to manipulate the measure results.
- The lack of information about previous colonoscopies may hide evidence of poor care.
- Pairing this measure with an appropriate outcome measure would make it more meaningful to all stakeholders.

**Committee response:**

- The Committee agreed that manipulation of results through gaming is a concern; however, the specific medical reason for the exclusion must be documented through the use of CPT-II codes. The Committee recommended that the specifications for the exclusions include a specific list of the types of medical reasons that are acceptable for this exclusion when the measures are submitted for the Stage 2 measure evaluation.

**STAGE 1 CSAC REVIEW (November 7-8, 2012)**

- **Decision:** Concept Approved with the requirement that the Steering Committee recommendations must be addressed in stage 2 submission.

**BOD REVIEW (November 29 – December 11, 2012)**

- **Decision:** Ratification of concept approval.

**STAGE 2**

**STAGE 2 PRE-REVIEW MEMBER COMMENT (March 4-18, 2013)**

- This measure received one supportive comment and several critical comments.
- Critical comments raised the issue that the exclusions provide a large loophole for providers to manipulate the results, and a lack of information about previous colonoscopies may hide evidence of poor care. The commenters suggested that the exclusions be better defined, and stated that it is unacceptable that a patient would be asked to undergo an unnecessary procedure because the provider is unable to track down their prior medical records.



<p><b>0659 Endoscopy/Polyp Surveillance: Colonoscopy Interval for Patients with a History of Adenomatous Polyps- Avoidance of Inappropriate Use</b></p> <ul style="list-style-type: none"> <li>One commenter was concerned that the measure could lead to overuse, as not all patients need another screen at a three-year interval. Another commenter suggested that the measure include minimum and maximum interval ranges.</li> </ul>
<p><b>STAGE 2 STEERING COMMITTEE CHECKLIST REVIEW (March 2013)</b>  Checklist recommendations satisfactorily addressed: <b>Y-9; N-4; A-1</b>  Move forward to full Stage 2 review: <b>Y-11; N-2; A-1</b></p> <ul style="list-style-type: none"> <li>The Committee requested additional evidence to support the measure, eMeasure specifications, and requested that the Committee align the interval specified in the measure with the guideline recommendations. The developer submitted an eMeasure for review in stage 2 but did not expand on the evidence as requested. The developer explained that the interval for the measure (at least three years) was consistent with evidence-based guidelines recommending an interval of 3-5 years.</li> <li>The Committee recognized that the availability of data limits the ability of the measure capture as many inappropriate colonoscopies as they would wish, but agreed that this measure should catch the intermediate-risk patients by excluding those with normal results who would not need follow up for 10 years, and those with cancer or other high-risk conditions that would need earlier screening.</li> </ul>
<p><b>STAGE 2 STEERING COMMITTEE REVIEW (April 8, 2013)</b>  <b>2. Scientific Acceptability: The measure meets the Scientific Acceptability criteria</b>  <i>(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)</i>  2a. Reliability: <b>H-1; M-12; L-0; I-0</b> 2b. Validity: <b>H-0; M-11; L-2; I-0</b></p> <ul style="list-style-type: none"> <li>Similar to measure 0658, the Committee identified issues with alignment of the original specifications and the eMeasure specifications. The developer and the Committee agreed the code sets should be harmonized between both versions of the measure.</li> <li>Committee members discussed the 3-5 year target and acknowledged that the measure still allows for inappropriate repeat colonoscopy in the 3-5 year interval but will capture inappropriate use prior to 3 years.</li> <li>The Committee agreed this measure had been tested in the same manner as the previous measure, 0658.</li> <li>Some Committee members questioned whether some of the conditions in the high-risk value set, such as eosinophilic gastroenteritis, are really high-risk for malignancy.</li> </ul>
<p><b>3. Feasibility: H-0; M-12; L-0; I-0</b></p> <ul style="list-style-type: none"> <li>The Committee again noted the similarity to measure 0658, and had no additional concerns about the feasibility of the original measure.</li> <li>Concerns about feasibility of the eMeasure will be discussed when the Committee reviews the crosswalk of the specifications.</li> </ul>
<p><b>4. Use &amp; Usability: H-0; M-12; L-0; I-0</b></p> <ul style="list-style-type: none"> <li>The measure is currently in use in PQRS.</li> </ul>
<p><b>5. Related and Competing Measures</b></p> <ul style="list-style-type: none"> <li>No related or competing measures were noted.</li> </ul>
<p><b>Steering Committee Recommendation for Endorsement: Y-11; N-1</b></p> <ul style="list-style-type: none"> <li>The Committee noted the measure could be improved but they agreed the measure does help prevent some overuse, particularly for repeat colonoscopies performed in less than three years.</li> </ul>



**0659 Endoscopy/Polyp Surveillance: Colonoscopy Interval for Patients with a History of Adenomatous Polyps- Avoidance of Inappropriate Use**

The Committee strongly recommends that the developer bring forward a better measure in the future.

- After reviewing the crosswalk between the eSpecifications and the original measure coding, the Committee agreed the two versions were aligned. The Committee voted 11-0 to also recommend the eMeasure for endorsement.

**Public & Member Comment**

**Comments received:**

- 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.
- The unsupportive 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 give a green light to colonoscopy without including information about previous colonoscopies.” Another comment suggested collecting the data for the measure would be burdensome.

**Developer response:**

- This measure was submitted for endorsement with electronic clinical/registry data as the intended data source. The measure is currently in use in 2 different GI-focused registries.
- Contrary to the commenter’s suggestion that this measure fails to include information about previous colonoscopies, eligibility for the measure is defined by the finding of prior colonic adenomatous polyp(s) in a previous colonoscopy. The identification of pre-cancerous colon polyps on a previous colonoscopy implies a certain level of quality as the primary goal of colon cancer screening is the timely removal of such lesions to prevent colon cancer. The published literature indicates that repeat colonoscopy is often overutilized and is not tied to clinical data on initial colonoscopy. The use of this measure is intended to increase physicians' adherence to the evidence based guideline and subsequently may reduce unnecessary tests, costs, and patient risk.
- The use of “medical reason” was not and is not intended to be purposefully vague, but rather stems from current PCPI exception methodology. This method uses three categories of reasons for which a patient may be removed from the denominator of an individual measure. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. A non-exhaustive list of examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For this measure, exceptions may include medical reason(s) (eg, patients with high risk for colon cancer, last colonoscopy incomplete, last colonoscopy had inadequate prep, piecemeal removal of adenomas, or last colonoscopy found greater than 10 adenomas) or system reason(s) for an interval of less than 3 years since the last colonoscopy (eg, unable to locate previous colonoscopy report). Where examples of exceptions are included in the measure language, value sets for these examples are developed and included in the eSpecifications.





**0659 Endoscopy/Polyp Surveillance: Colonoscopy Interval for Patients with a History of Adenomatous Polyps- Avoidance of Inappropriate Use**

- Since this measure's review during the 2nd stage of NQF's process and posting for public comment, we have made a few minor modifications to the language and corresponding value set for one of the examples noted above (ie, patients with high risk for colon cancer). In consultation with expert work group members, we have further clarified the conditions that would signify a patient being at high risk for colon cancer to include 4 categories: Crohn's disease, ulcerative colitis, rectal/lower GI bleeding, personal or family history of colon cancer. Accordingly, we have revised the description of the medical reason example to read "...[eg, patients with high risk for colon cancer (ie, Crohn's disease, ulcerative colitis, lower gastrointestinal bleeding, personal or family history of colon cancer)]..." and created new value sets that reflect those categories. This will hopefully correct the misperception that the list of exclusions included are so broad as to enable exception reporting.
- This issue of exception reporting --the potential for physicians to inappropriately exclude patients to enhance their performance statistics- has been raised by these comments. Research has indicated that levels of exception reporting occur infrequently and are generally valid. (Doran et al., 2008), (Kmetik et al., 2011) Furthermore, exception reporting has been found to have substantial benefits: "it is precise, it increases acceptance of [pay for performance] programs by physicians, and it ameliorates perverse incentives to refuse care to "difficult" patients." (Doran et al., 2008) A recent study conducted by the PCPI in 47,075 outpatients with coronary artery disease seen during 2006 and 2007 in 5 medical practices that used electronic health records, reported that the overall exception percentage for all 4 measures studied was 3.5%. The vast majority (92.6%) of those exceptions were confirmed during manual review.(Kmetik et al., 2011)
- Although this methodology does not require the external reporting of more detailed exception data, the PCPI recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The PCPI also advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

References:

Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of pay for performance targets by English Physicians. *New Engl J Med.* 2008; 359: 274-84.

Kmetik KS, Otoole MF, Bossley H et al. Exceptions to Outpatient Quality Measures for Coronary Artery Disease in Electronic Health Records. *Ann Intern Med.* 2011;154:227-234.

**Committee response:** The Committee acknowledged the comments, but agreed they had already discussed these issues and they did not wish to "let the perfect be the enemy of the good" as the commenter stated. They did not change their recommendation for the measure.



**2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)**

[Measure Submission Form](#) | [Specifications](#)

**Status:** New Submission

**Description:** Percent of discharges with an in-hospital death among cases with a principal diagnosis of gastrointestinal hemorrhage

**Numerator Statement:** Number of in-hospital deaths among cases meeting the inclusion and exclusion rules for the denominator

**Denominator Statement:** All discharges, age 18 years and older, with a principal diagnosis code for gastrointestinal hemorrhage OR a principal diagnosis of predisposing condition for esophageal varices and a secondary diagnosis of esophageal varices in condition classified elsewhere with bleeding (456.20)

**Exclusions:** Exclude cases:

- transferred to another short-term hospital
- with MDC 14 (pregnancy, childbirth, and puerperium)
- with missing discharge disposition, gender, age, quarter, year or principal diagnosis

**Adjustment/Stratification:** Statistical risk model The predicted value for each case is computed using a two-stage hierarchical model (the first stage is a logistic regression using Generalized Estimating Equations (GEE) to account for clustering of patients within hospitals; the second stage is a reliability weight). The covariates in the logistic regression include age (in 5-year age groups pooled), APR-DRG and APR-DRG Risk of Mortality subclass, MDC and transfer-in status. The reference population used in the regression is the universe of discharges for states that participate in the HCUP State Inpatient Data (SID) for the years 2008, a database consisting of 42 states and approximately 30 million adult discharges.

**Level of Analysis:** Facility

**Type of Measure:** Outcome

**Data Source:** Administrative claims

**Measure Steward:** Agency for Healthcare Research and Quality

**STAGE 1**

**STAGE 1 PRE-REVIEW MEMBER COMMENTS (August 7-21, 2012)**

- *America's Health Insurance Plans* - This measure may be subject to a small numbers problem raising reliability issues.

**STEERING COMMITTEE MEETING (August 27-28, 2012)**

**1. Importance to Measure and Report:**

**1a. High Impact: H-14; M-0; L-0; I-0**

**Discussion:** There is general agreement this measure focus addresses a high impact area. GI hemorrhage is a common problem.

**1c. Evidence**

**14:** Yes, body of evidence meets guidance for quantity, quality, consistency

**0:** No, body of evidence does not meet guidance for quantity, quality, consistency

**0:** No, inadequate information to rate quantity, quality, consistency of body of evidence

**Discussion:** Outcome measures do not require evidence; however, the Committee agreed that the developer did provide a rationale that supports the relationship of the health outcome to processes or structures of care.

**1b. Performance Gap: H-10; M-4; L-0; I-0**

**Discussion:**

- The odds ratio of bleeding ranges from 17 to 22 based on the type of hospital and from 14 to 25 based on insurance status.
- Risk adjusted using 3M APR-DRG's and it is publicly available to implement this measure.



<b>2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)</b>
<p>While gender is included in the risk adjustment model, race and ethnicity are not. This allows for stratification by race and ethnicity as the data submitted demonstrates significant differences in the outcomes among white, black, and Hispanic patients.</p> <ul style="list-style-type: none"> <li>The Committee agreed based on the above discussions, that there is a performance gap for this measure focus.</li> </ul>
<p><b>Recommendations to Developer for Stage 2:</b></p> <ul style="list-style-type: none"> <li>Numerator and denominator only include patients with primary diagnosis of GI bleed, consider how this might impact the capture of other patients with GI bleed who do not have it as a primary diagnosis.</li> <li>Consider stratifying by esophageal bleeds and lower GI bleeds.</li> </ul>
<p><b>Stage 1 Steering Committee Recommendation for Approval of Concept: Y-14; N- 0</b></p>
<p><b>STAGE 1 MEMBER &amp; PUBLIC COMMENT (September 26 – October 25, 2012)</b></p> <p><b>Member &amp; Public Comments:</b></p> <ul style="list-style-type: none"> <li>One comment in support of this concept.</li> </ul>
<p><b>STAGE 1 CSAC REVIEW (November 7-8, 2012)</b></p> <ul style="list-style-type: none"> <li><b>Decision:</b> Approved with the requirement that the Steering Committee recommendations must be addressed in the stage 2 submission.</li> </ul>
<p><b>STAGE 1 BOD REVIEW (November 29 – December 11, 2012)</b></p> <p><b>Decision:</b> Ratification of concept approval.</p>
<b>STAGE 2</b>
<p><b>STAGE 2 PRE-REVIEW MEMBER COMMENT (March 4-18, 2013)</b></p> <ul style="list-style-type: none"> <li>This measure received two supportive comments, noting that it measures an important health outcome that is both usable and feasible.</li> </ul>
<p><b>STAGE 2 STEERING COMMITTEE CHECKLIST REVIEW (March 2013)</b></p> <p>Checklist recommendations satisfactorily addressed: <b>Y-11; N-2; A-1</b></p> <p>Move forward to full Stage 2 review: <b>Y-12; N-1; A-1</b></p> <ul style="list-style-type: none"> <li>The Committee requested that the developer stratify the measure by esophageal bleeds and lower GI bleeds. They also requested the developer consider how only including patients with a primary diagnosis of GI bleed might impact the capture of other patients who have a GI bleed, but not as primary diagnosis. The developer stratified the measure into esophageal varices and all other cases. They explained their mortality measures use the principal diagnosis because using a secondary diagnosis results in a measure with a heterogeneous denominator, which makes it less meaningful to both providers and consumers, and also reduces the specificity of the measure. The Committee agreed this was a satisfactory answer.</li> </ul>
<p><b>STAGE 2 STEERING COMMITTEE REVIEW (April 3, 2013)</b></p> <p><b>2. Scientific Acceptability: The measure meets the Scientific Acceptability criteria</b> (2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)</p> <p>2a. Reliability: <b>H-0; M-13; L-0; I-0</b> 2b. Validity: <b>H-2; M-11; L-0; I-0</b></p> <ul style="list-style-type: none"> <li>This is a hospital-level measure based on administrative claims.</li> <li>In response to a question about validity of the diagnosis codes for GI bleeding, the developer advised the Committee that “we know that the accuracy of the principal diagnosis is on the order of 95 percent, at least at the level at which we are using it here, which is an aggregation of principal diagnoses.” Additionally a specific study comparing claims to chart or registry data showed 88% predictive value for GI hemorrhage in any diagnosis field in the claim.</li> </ul>



**2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)**

- The Committee noted that this measure has been thoroughly tested. Reliability of the measure score was tested using signal-to-noise analysis, with reasonable results. The developer explained that the testing looked at how much of the total variation in performance is due to systemic variation across hospitals, and how much is random variation due to some hospitals having very small few cases. With small numbers the sampling variability must be taken into account for a more reliable measure.
- Empiric validity testing of the measure identified several hypotheses as to what the relationship should be between the score and the observations such as high volume will have better outcomes and transfer patients may have poorer outcomes. The results indicated a negative association between the hospital risk-adjusted mortality and the hospital volume and an opposite relationship between mortality and transfers out, which fit with their initial hypothesis.
- The measure is risk adjusted to account for comorbidities and severity, and transfers and excludes patients who transfer out, and stratifies patients who are transferred in (who tend to have worse outcomes than patients who complete all care at the same hospital). The c-statistic is 0.831.
- The only exclusion is the patients who are transferred out to an acute hospital because the outcome is not determined within the hospital where the patient presented.
- In an analysis to determine the ability to identify differences in performance, the information presented noted that “low volume hospitals are more likely to be identified as performing worse than the reference population rate primarily because of the volume/persistence effect. Although more hospitals are likely to be identified as performing worse than performing better, because patients are concentrated in high volume hospitals, about 5.1% of patients are in better performing hospitals, and 8.0% of patients are in worse performing hospitals.”

**3. Feasibility: H-10; M-2; L-0; I-0**

- The data source is claims data. The outcome of death is unlikely to be miscoded.

**4. Use & Usability: H-11; M-1; L-0; I-0**

- Though this measure is newly submitted to NQF, it has been in use for several years and is publicly reported.
- The submission lists four current uses of the measure for accountability that include a large number of hospitals across the nation.

**5. Related and Competing Measures**

- No related or competing measures noted.

**Stage 2 Steering Committee Recommendation for Endorsement: Y-12; N-0**

**Public & Member Comment**

**Comments received:**

- This measure received comments from six organizations or individuals. Four of these comments, those 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

2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)

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.

#### 2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)

- 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.

#### Developer response:

- We appreciate your concerns regarding reliability of the measure and want to be sure it is clear how we assess the psychometrics of a measure within the AHRQ QIs.
- 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.
- Health care providers should, and often do, have internal mechanisms for tracking outcomes on a nearly real-time basis, especially with current availability of electronic health information systems. However, public reporting and other accountability applications require a prior time period over which experience can be accumulated and compared across providers, with adequate reliability. "Looking in the rearview mirror," as the commenter describes it, allows stakeholders outside a provider organization to compare the performance of multiple providers, given available benchmarks, and to make appropriate decisions based on this assessment. "Looking in the rearview mirror" may not tell us where we are going in the future, but it does tell



#### 2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)

us how well we have negotiated the difficult terrain just behind us.

- In our original submission to NQF, we reported hospital level regression results that demonstrated that the prior year performance score was more predictive of current performance on the risk-adjusted rate than other hospital attributes (e.g. volume or transfer-out rate) with a coefficient of 0.65 (where a coefficient of 1.00 would be perfect persistence). We are currently in the process of updating the data to 2011, but the coefficient of 0.65 confirms the usefulness of using prior time-period data to inform current decision-making.
- 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:
  - Prompt recognition of gastrointestinal hemorrhage as the cause of a patient's symptoms, necessitating inpatient admission for further evaluation and treatment.
  - 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.
  - Appropriate stabilization of acutely ill patients with prompt but safe administration of fluids, blood products, vaspressors, and other resuscitative maneuvers.
  - Appropriate diagnostic and evaluation processes to identify the source of bleeding and to characterize the risk of rebleeding.
  - Appropriate monitoring by nurses, physicians, and other health professionals to identify early warning signs of clinical deterioration and to implement "rapid response" as appropriate.
  - 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.
  - 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



### 2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)

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.

- The denominator for this indicator, as for all of the risk-adjusted mortality indicators submitted by AHRQ and CMS, is defined using the principal diagnosis, which is defined in regulation as the condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital for care. The reason is that using a secondary diagnosis code results in a measure with a heterogeneous denominator, making the measure less useful for providers (in terms of allocating quality improvement resources) and less meaningful to consumers (in terms of knowing the likelihood of being in the population at risk). As previously noted, adding a denominator inclusion criterion for secondary diagnosis code of GI Hemorrhage would result in an increase in the denominator of 262% (from 458,307 cases to 1,660,884 cases). The denominator would be very heterogeneous, consisting of an additional 2,700 individual diagnosis codes in 538 different MS-DRG codes (with osteoarthritis as the most common). These concerns outweigh the potential benefit of increasing measure reliability in accountability applications, for which the topic of accountability (in this case, care of patients admitted with GI hemorrhage) must be clearly defined.
- The denominator of this indicator does not capture patients in whom GI hemorrhage was just a comorbid condition, and it does not capture patients who were admitted for unrelated reasons and developed stress ulceration during the hospital stay. Conversely, the numerator captures all inpatient deaths occurring among eligible admissions for GI hemorrhage, without regard to the final cause of death. This is a routine definitional practice for short-term mortality measures, due to the uncertainty in identifying causes of death among critically ill patients with multiple related conditions (e.g., GI hemorrhage may lead to hypotension and shock, which may lead to a myocardial infarction or stroke, leading to confusion about the cause of death). Usual practice in the quality measurement field is to link a death after this type of worsening trajectory back to the original cause of admission.

#### **Committee response:**

- 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.