NATIONAL QUALITY FORUM

Memo

TO: Agency for Healthcare Research and Quality

FR: NQF GI/GU Project Staff

RE: GI/GU Endorsement Maintenance Pilot Project: Stage two checklist

DA: September 28, 2012

GI/GU Endorsement Maintenance Pilot Project, 2012

Thank you for your participation and concept submission to the GI/GU Endorsement Maintenance Pilot Project. Please carefully review the instructions below for next steps.

Preparation for submission of recommended concepts to stage two

1. Keep in mind, while the measure submission forms for recommended concepts opens in early November, approval of concepts is finalized with Board of Directors approval on November 30.

2. Review all requirements for measure submission and criteria to be suitable for endorsement:
   - Ensure that evidence remains current and consistent with concept
     - Check if there have been any major changes in the evidence base supporting the approved concept. If yes, provide the citation and copy of the study or article and discuss the impact on the measure concept.
     - If there are any changes in the concept from that which was approved, identify those changes and discuss the relevance of the evidence to the approved concept and the updated concept.
   - Ensure that testing requirements have been satisfied
     - Testing requirements are available in the Measure Testing Task Force report

3. Review the Developer Guidebook for additional resources and information for

preparing your stage two measure submission. The updated guidebook will be available once stage two submission forms are opened and will also be distributed by NQF Technical Assistance Staff.

4. **Notify NQF project staff by October 25, 2012** if you plan to submit full specifications and testing for approved concepts by the December 19, 2012 stage two measure submission deadline.

5. You will be required to submit at least one of your fully specified and tested measures on or prior to the **technical assistance deadline on December 3, 2012**, for a technical review for completeness and responsiveness by the NQF staff.

6. Measure submissions must be complete and responsive to ALL questions in order to be advanced to the Steering Committee for consideration and evaluation.

GI Concept Recommended for Approval: AHRQ

Provide a response for EACH Committee recommendation describing your rationale for implementing (or not) the recommendation and any additional considerations.

Upload this document to your online measure submission form for review by the Committee in stage two.

C 2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)	
Committee Recommendations to Developer	**Developer Response**

C 2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)	
Numerator and denominator only include patients with primary diagnosis of GI bleed, consider how this might impact the capture of other patients GI bleed who do not have it has a primary diagnosis.	In general, our mortality measures use the principal diagnosis which is 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). However, we have identified specific situations where a gastrointestinal principal diagnosis that does not indicate hemorrhage is associated with a GI hemorrhage code reported as present on admission in a secondary diagnosis field. This appears to represent a heterogeneous mixture of clinical situations, including bleeding hemorrhoids, bleeding cancers, ischemic and inflammatory bowel disease, and ulcers outside the stomach and small bowel. The challenge is that, in many cases, the link between the principal diagnosis and the GI hemorrhage secondary diagnosis is not clear (e.g., irritable bowel syndrome, achalasia, other functional disorders or symptoms). In many of these patients, GI bleeding may have been an incidental finding rather than part of the reason for admission. In other patients, the principal diagnosis appears to have been an obstructing lesion that was NOT bleeding (e.g., gastric ulcer), and the bleeding lesion was actually in a different location (e.g., rectal or anal). In these cases, the patient was probably admitted for obstruction, not for bleeding. Therefore, the problem does not lend itself to a simple solution, and our analyses are ongoing. Esophageal varices represent a special situation, because coders are instructed to code the underlying cause of the varices in the principal

C 2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)	
	diagnosis field, and the associated variceal hemorrhage (456.20) in a secondary diagnosis field. Our analyses using both all-payer HCUP data and VA data confirm that principal diagnoses on the specified list of liver-related conditions must be used (in association with 456.20 in a secondary diagnosis field) to capture esophageal variceal hemorrhage. We have also added the capacity to stratify the measure (see below). We also note that the AHRQ QI already does have a measure that incorporates a secondary diagnosis code of GI Hemorrhage, which is the PSI #04 Death among Surgical Inpatients with Serious Treatable Complications.
Consider stratifying by esophageal bleeds and lower GI bleeds.	The denominator may be stratified into esophageal varices and all other cases. The definition of esophageal varices is all discharges, age 18 years and older, with a principal diagnosis code for gastrointestinal hemorrhage and a secondary diagnosis of esophageal varices with bleeding (456.0 and 456.20) 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) OR a principal diagnosis of esophageal varices with bleeding (456.0).

Measure Submission and Evaluation Worksheet 6.0

This form contains the information submitted by measure developers/stewards, organized according to NQF's measure evaluation criteria and process. The evaluation criteria, evaluation guidance documents, and a blank online submission form are available on the submitting standards web page.

NQF #: 2065 NQF Project: GI and GU Project

(for Endorsement Maintenance Review)
Original Endorsement Date: Most Recent Endorsement Date: Evaluation Form Created: March 22, 2013

BRIEF MEASURE INFORMATION

De.1 Measure Title: Gastrointestinal Hemorrhage Mortality Rate (IQI #18)

Co.1.1 Measure Steward: Agency for Healthcare Research and Quality

De.2 Brief Description of Measure: Percent of discharges with an in-hospital death among cases with a principal diagnosis of gastrointestinal hemorrhage

2a1.1 Numerator Statement: Number of in-hospital deaths among cases meeting the inclusion and exclusion rules for the denominator

2a1.4 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)

2a1.8 Denominator 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

1.1 Measure Type: Outcome
2a1. 25-26 Data Source: Administrative claims, State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality, Rockville, MD. The SID contain the universe of inpatient discharge abstracts in participating States, translated into uniform format to facilitate analyses. With data from 47 States, the SID, in the aggregate, currently encompass about 97 percent of all annual discharges in the U.S. (http://www.hcup-us.ahrq.gov/sidoverview.jsp#What). Data dictionary and code tables are available at http://www.qualityindicators.ahrq.gov/Downloads/Software/WinQI/V44/Software%20Instructions%20(WinQI)%20V4.4.pdf
A data dictionary for the source data, HCUP SID, is available at: http://www.hcup-us.ahrq.gov/db/state/siddist/sid_multivar.jsp
2a1.33 Level of Analysis: Facility
1.2-1.4 Is this measure paired with another measure? No

De.3 If included in a composite, please identify the composite measure (*title and NQF number if endorsed*): N/A

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1a. High Impact: H☐ M☐ L☐ I ☐

(*The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact aspect of healthcare.*)

De.4 Subject/Topic Areas *(Check all the areas that apply)*: Gastrointestinal (GI) : Bleeding, Gastrointestinal (GI)

De.5 Cross Cutting Areas (*Check all the areas that apply)*:

1a.1 Demonstrated High Impact Aspect of Healthcare:

A leading cause of morbidity/mortality

1a.2 If "Other," please describe: N/A

1a.3 Summary of Evidence of High Impact *(Provide epidemiologic or resource use data)*:

More people are admitted to the hospital for upper GI bleeding than for congestive heart failure or deep vein thrombosis.[1] In the United States, the annual rate of hospitalization for upper GI bleeding is estimated to be 165 per 100,000—equating to more than 300,000 hospitalizations per year, at a cost of $2.5 billion,[2,3] with a case-fatality rate of 7 to 10 percent.[1] However, costs are not constant across all bleed types. In a study using the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS), researchers reported a fourfold higher cost and length of stay (LOS) attributable to rebleeding for variceal upper gastrointestinal bleeding compared to nonvariceal upper gastrointestinal bleeding.[2] Hospitalization costs with and without complications were $5,632 and $3,402 for non-variceal upper gastrointestinal bleeding, versus $23,207 and $6,612 for variceal upper gastrointestinal bleeding, respectively. Mean length of stay was 4.4 and 2.7 days for nonvariceal bleeding, versus 15.2 and 3.8 days for variceal bleeding, respectively.[2] Acute, massive lower gastrointestinal bleeding has an incidence of 20 to 27 episodes per 100,000 persons annually, with a mortality rate of 4 to 10 percent.[3] Mortality rates increase in patients with advancing age and increasing number of associated underlying comorbidities, specifically renal and hepatic dysfunction, heart disease, and malignancies.[3-9]

Among community hospitals in the Healthcare Cost and Utilization Project (HCUP), the risk-adjusted rate of this indicator was 19.363 per 1,000 eligible admissions (1.94%) in 2008. This rate has steadily declined over the past 14 years, from 5.78% in 1994 to 4.57% in 2000 to 3.02% in 2005.

1a.4 Citations for Evidence of High Impact cited in 1a.3:

1. Albeldawi M., Qadeer MA, Vargo JJ. Managing acute upper GI bleeding, preventing recurrences. Cleveland Clin J Med. 2010; 77(2):131-142.
2. Viviane A, Alan BN. Estimates of costs of hospital stays for variceal and nonvariceal upper gastrointestinal bleeding in the United States. Value Health 2008; 11:1–3.
3. Yavorski RT, Wong RK, Maydonovitch C, Battin LS, Furnia A, Amundson DE. Analysis of 3,294 cases of upper gastrointestinal bleeding in military medical facilities. Am J Gastroenterol 1995; 90:568–573.
4. Manning-Dimmitt LL, Dimmitt SG,Wilson GR. Diagnosis of gastrointestinal bleeding in adults. Am Fam Physician. 2005; 71(7):1339-46.
5. Wilcox CM, Clark WS. Causes and outcome of upper and lower gastrointestinal bleeding: the Grady Hospital experience. South Med J. 1999;92:44–50.
6. Vreeburg EM, Snel P, de Bruijne JW, Bartelsman JF, Rauws EA, Tytgat GN. Acute upper gastrointestinal bleeding in the Amsterdam area: incidence, diagnosis, and clinical outcome. Am J Gastroenterol. 1997;92:236–43.
7. Hussain H, Lapin S, Cappell MS. Clinical scoring systems for determining the prognosis of gastrointestinal bleeding. Gastroenterol Clin North Am. 2000;29:445–64.
8. Zuccaro G Jr. Management of the adult patient with acute lower gastrointestinal bleeding. American College of Gastroenterology. Practice Parameters Committee. Am J Gastroenterol. 1998;93:1202–8.
9. Longstreth GF. Epidemiology and outcome of patients hospitalized with acute lower gastrointestinal hemorrhage: a population-based study. Am J Gastroenterol. 1997;92:419–24.

1b. Opportunity for Improvement: H☐ M☐ L☐ I ☐

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

(*There is a demonstrated performance gap - variability or overall less than optimal performance*)

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
Providers may adopt the processes of care or structures of care of the best performing providers or consumers may select the best performing providers in order to improve overall outcomes.

1b.2 Summary of Data Demonstrating Performance Gap *(Variation or overall less than optimal performance across providers)*:
[*For Maintenance – Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.*]
In regard to figures below:
1st figure: estimate per 1,000, risk adjusted rates
2nd figure: standard error
3rd figure: p value relative to marked group (marked group = "c")
4th figure: p value: current year relative to prior year

Key:
"c": Reference for p-value test statistics
"*": Data do not meet criteria for statistical reliability, data quality, or confidentiality
HCUPNet: http://hcupnet.ahrq.gov

Hospital characteristic:				
Location of inpatient treatment:				
Northeastc	20.844	0.405	c.ccc	0.000
Midwest	17.614	0.367	0.000	0.000
South	19.539	0.288	0.009	0.000
West	19.688	0.437	0.052	0.000
Ownership/control:				
Private, not-for-profitc	18.405	0.208	c.ccc	0.000
Private, for-profit	21.746	0.495	0.000	0.001
Public	22.779	0.528	0.000	0.000
Teaching status:				
Teaching	17.346	0.333	0.000	0.000
Nonteachingc	20.192	0.214	c.ccc	0.000
Location of hospital (NCHS):				
Large central metropolitan	18.404	0.317	0.857	0.000
Large fringe metropolitanc	18.315	0.379	c.ccc	0.000
Medium metropolitan	18.742	0.377	0.424	0.000
Small metropolitan	22.087	0.578	0.000	0.008
Micropolitan	22.245	0.568	0.000	0.000
Noncore	24.739	1.193	0.000	0.000
Bed size of hospital:				
Less than 100	22.932	0.596	0.000	0.000
100 - 299c	20.285	0.284	c.ccc	0.000
300 - 499	18.548	0.347	0.000	0.000
500 or more	17.257	0.371	0.000	0.000

1b.3 Citations for Data on Performance Gap: [*For Maintenance – Description of the data or sample for measure results reported in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*]
Source: Agency for Healthcare Research and Quality (AHRQ), Center for Delivery, Organization, and Markets, Healthcare Cost and Utilization Project (HCUP), Nationwide Inpatient Sample (NIS), 2009, and AHRQ Quality Indicators, modified version of 4.1.
There are 630 hospitals with a least 1 denominator case for IQI 18 in the NIS; this represents all such hospitals in the NIS, which is a 20% stratified random sample of all community hospitals.

1b.4 Summary of Data on Disparities by Population Group *(for example by race/ethnicity, gender, age, insurance status,*

socioeconomic status, and/or disability, etc. If you do not have data on your specific measure, perform a literature search/review and report data for the measure or similar appropriate concept.): [*For Maintenance –Descriptive statistics for performance results for this measure by population group*]

In regard to figures below:
1st figure: estimate per 1,000, risk adjusted rates
2nd figure: standard error
3rd figure: p value relative to marked group (marked group = “c”)
4th figure: p value: current year relative to prior year

Key:
"c": Reference for p-value test statistics
"*": Data do not meet criteria for statistical reliability, data quality, or confidentiality
HCUPNet: http://hcupnet.ahrq.gov

Patient characteristic:				
Age groups for conditions affecting any age				
18-44c	04.395	0.308	c.ccc	0.111
45-64	18.300	0.345	0.000	0.000
65 and over	22.989	0.238	0.000	0.000
Age groups for conditions affecting primarily elderly				
65-69c	13.675	0.503	c.ccc	0.000
70-74	14.911	0.452	0.068	0.023
75-79	15.322	0.470	0.017	0.000
80-84	21.815	0.495	0.000	0.000
85 and over	38.483	0.597	0.000	0.000
Gender:				
Malec	22.153	0.259	c.ccc	0.000
Female	18.005	0.250	0.000	0.000
Median income of patient's ZIP code:				
First quartile (lowest income)	20.311	0.329	0.000	0.000
Second quartile	20.089	0.352	0.000	0.000
Third quartile	18.554	0.376	0.296	0.000
Fourth quartile (highest income)c	17.985	0.395	c.ccc	0.000
Location of patient residence (NCHS):				
Large central metropolitan	18.732	0.338	0.225	0.000
Large fringe metropolitanc	18.129	0.364	c.ccc	0.000
Medium metropolitan	19.466	0.405	0.014	0.000
Small metropolitan	21.911	0.639	0.000	0.013
Micropolitan	20.531	0.524	0.000	0.000
Noncore	21.007	0.662	0.000	0.000
Expected payment source:				
Private insurancec	21.381	0.510	c.ccc	0.176
Medicare	18.374	0.204	0.000	0.000
Medicaid	22.228	0.799	0.372	0.000
Other insurance	30.117	1.379	0.000	0.003
Uninsured / self-pay / no charge	25.447	1.094	0.001	0.020

Race/ethnicity (observed rates, not risk-adjusted):
White 0.14554
Black 0.09087
Hispanic 0.11465
Asian and NH/PI 0.19054
Amer Indian/AN 0.17424
Other 0.16857

1b.5 Citations for Data on Disparities Cited in 1b.4: [*For Maintenance – Description of the data or sample for measure results reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*]
Sources:
Agency for Healthcare Research and Quality (AHRQ), Center for Delivery, Organization, and Markets, Healthcare Cost and Utilization Project (HCUP), Nationwide Inpatient Sample (NIS), 2009, and AHRQ Quality Indicators, modified version of 4.1.
Race/ethnicity data are from the Healthcare Cost and Utilization Project (HCUP), State Inpatient Databases (SID), 2008, Agency for Healthcare Research and Quality, Rockville, MD.
There are 630 hospitals with a least 1 denominator case for IQI 18 in the NIS; this represents all such hospitals in the NIS, which is a 20% stratified random sample of all community hospitals.

1c. Evidence (*Measure focus is a health outcome OR meets the criteria for quantity, quality, consistency of the body of evidence.*)
Is the measure focus a health outcome? Yes☐ No☐ If not a health outcome, rate the body of evidence.

Quantity: H☐ M☐ L☐ I☐ Quality: H☐ M☐ L☐ I☐ Consistency: H☐ M☐ L☐ I☐

Quantity	Quality	Consistency	Does the measure pass subcriterion1c?
M-H	M-H	M-H	Yes☐
L	M-H	M	Yes☐ IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise No☐
M-H	L	M-H	Yes☐ IF potential benefits to patients clearly outweigh potential harms: otherwise No☐
L-M-H	L-M-H	L	No ☐

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service	Does the measure pass subcriterion1c? Yes☐ IF rationale supports relationship

SEE ATTACHED EVIDENCE SUBMISSION FORM

Was the threshold criterion, *Importance to Measure and Report*, met?
(*1a & 1b must be rated moderate or high and 1c yes*) Yes☐ No☐
Provide rationale based on specific subcriteria:

For a new measure if the Committee votes NO, then STOP.
For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

S.1 Measure Web Page *(In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained*). Do you have a web page where current detailed specifications for this measure can be obtained?
Not available at this time

2a1.1 Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, e.g., cases from the target population with the target process, condition, event, or outcome)*:

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

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

2a1.3 Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses*:
All discharges with a Disposition of Patient (DISP) coded as "died" (20)

2a1.4 Denominator Statement *(Brief, narrative description of the target population being measured)*:
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)

2a1.5 Target Population Category *(Check all the populations for which the measure is specified and tested if any)*:
Senior Care

2a1.7 Denominator Details (*All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
ICD-9-CM principal diagnosis code of Gastrointestinal hemorrhage (see below for detail). According to the ICD-9-CM Official Guidelines for Coding and Reporting (http://www.cdc.gov/nchs/data/icd9/icd9cm_guidelines_2011.pdf), the principal diagnosis is defined in the Uniform Hospital Discharge Data Set (UHDDS) as "that condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital for care." The UHDDS definitions are used by hospitals to report inpatient data elements in a standardized manner. These data elements and their definitions can be found in the July 31, 1985, Federal Register (Vol. 50, No, 147), pp. 31038-40.
The time window may be determined by the user, but is generally a calendar year.
ICD-9-CM Gastrointestinal hemorrhage diagnosis codes:
4560 ESOPHAG VARICES W BLEED
5307 MALLORY-WEISS SYNDROME
53021 ULCER ESOPHAGUS W BLEED
53082 ESOPHAGEAL HEMORRHAGE
53100 AC STOMACH ULCER W HEM
53101 AC STOMAC ULC W HEM-OBST
53120 AC STOMAC ULC W HEM/PERF
53121 AC STOM ULC HEM/PERF-OBS
53140 CHR STOMACH ULC W HEM
53141 CHR STOM ULC W HEM-OBSTR
53160 CHR STOMACH ULC HEM/PERF
53161 CHR STOM ULC HEM/PERF-OB
53200 AC DUODENAL ULCER W HEM
53201 AC DUODEN ULC W HEM-OBST
53220 AC DUODEN ULC W HEM/PERF
53221 AC DUOD ULC HEM/PERF-OBS
53240 CHR DUODEN ULCER W HEM
53241 CHR DUODEN ULC HEM-OBSTR
53260 CHR DUODEN ULC HEM/PERF
53261 CHR DUOD ULC HEM/PERF-OB
53300 AC PEPTIC ULCER W HEMORR
53301 AC PEPTIC ULC W HEM-OBST
53320 AC PEPTIC ULC W HEM/PERF
53321 AC PEPT ULC HEM/PERF-OBS
53340 CHR PEPTIC ULCER W HEM
53341 CHR PEPTIC ULC W HEM-OBS
53360 CHR PEPT ULC W HEM/PERF
53361 CHR PEPT ULC HEM/PERF-OB

53400 AC MARGINAL ULCER W HEM
53401 AC MARGIN ULC W HEM-OBST
53420 AC MARGIN ULC W HEM/PERF
53421 AC MARG ULC HEM/PERF-OBS
53440 CHR MARGINAL ULCER W HEM
53441 CHR MARGIN ULC W HEM-OBS
53460 CHR MARGIN ULC HEM/PERF
53461 CHR MARG ULC HEM/PERF-OB
53501 ACUTE GASTRITIS W HMRHG
53511 ATRPH GASTRITIS W HMRHG
53521 GSTR MCSL HYPRT W HMRG
53531 ALCHL GSTRITIS W HMRHG
53541 OTH SPF GASTRT W HMRHG
53551 GSTR/DDNTS NOS W HMRHG
53561 DUODENITIS W HMRHG
53783 ANGIO STM/DUDN W HMRHG
53784 DIEULAFOY LES,STOM&DUOD
56202 DVRTCLO SML INT W HMRHG
56203 DVRTCLI SML INT W HMRHG
56212 DVRTCLO COLON W HMRHG
56213 DVRTCLI COLON W HMRHG
5693 RECTAL & ANAL HEMORRHAGE
56985 ANGIO INTES W HMRHG
56986 DIEULAFOY LES, INTESTINE
5780 HEMATEMESIS
5781 BLOOD IN STOOL
5789 GASTROINTEST HEMORR NOS
The following is the list of codes for “predisposing condition for esophageal varices”
07044 CHRNC HPT C W HEPAT COMA
07054 CHRNC HPT C WO HPAT COMA
5710 ALCOHOLIC FATTY LIVER
5711 AC ALCOHOLIC HEPATITIS
5712 ALCOHOL CIRRHOSIS LIVER
5713 ALCOHOL LIVER DAMAGE NOS
57140 CHRONIC HEPATITIS NOS
57141 CHRONIC HEPATITIS NOS
57142 AUTOIMMUNE HEPATITIS
57149 CHRONIC HEPATITIS NOS
5715 CIRRHOSIS OF LIVER NOS
5716 BILIARY CIRRHOSIS
5718 CHRONIC LIVER DIS NEC
5719 CHRONIC LIVER DIS NOS
5722 HEPATIC COMA
5723 PORTAL HYPERTENSION
5728 OTH SEQUELA, CHR LIV DIS
5738 LIVER DISORDERS NEC

2a1.8 Denominator Exclusions *(Brief narrative description of exclusions from the target population)*:
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

2a1.9 Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
• transfer to another short-term hospital (Disposition of Patient (DISP) coded as Transfer to Short-term Hospital (2))
• Major Diagnostic Category 14 (pregnancy, childbirth, and puerperium) - note that this exclusion is implied by the fact that the denominator is limited to patients with a principal diagnosis code for gastrointestinal hemorrhage, which maps to MDC 6 (digestive)
• missing discharge disposition (DISP=missing)
• missing gender (SEX=missing)
• missing age (AGE=missing)
• missing quarter (DQTR=missing)
• missing year (YEAR=missing)
• missing principal diagnosis (DX1=missing)

2a1.10 Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses)*:
The denominator may be stratified into two groups: 1) esophageal varices and 2) all other cases. Esophageal varices includes all discharges for patients age 18 years and older, with a principal diagnosis code for gastrointestinal hemorrhage and a secondary diagnosis of esophageal varices with bleeding (456.0 and 456.20), 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), OR a principal diagnosis of esophageal varices with bleeding (456.0).

2a1.11 Risk Adjustment Type *(Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13)*: Statistical risk model 2a1.12 If "Other," please describe: N/A

2a1.13 Statistical Risk Model and Variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.)*:
N/A

2a1.14-16 Detailed Risk Model Available at Web page URL (or attachment). Include coefficients, equations, codes with descriptors, definitions, and/or specific data collection items/responses. Attach documents only if they are not available on a webpage and keep attached file to 5 MB or less. NQF strongly prefers you make documents available at a Web page URL. Please supply login/password if needed:
Available in attached Excel or csv file

2a1.17-18. Type of Score:
Rate/proportion

If other: N/A

2a1.19 Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*:
better quality = lower score

2a1.20 Calculation Algorithm/Measure Logic*(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*:
The indicator is expressed as a rate, is defined as outcome of interest / population at risk, or numerator / denominator. The AHRQ Quality Indicators (AHRQ QI) software performs six steps to produce the rates. 1) Flag discharge-level records to identify the outcome of interest and 2) the population at risk. 3) Calculate observed rates as the sum of the records flagged in the numerator divided by the sum of the records flag in the denominator for user-specified combinations of stratifiers. 4) Calculate expected rates. Regression coefficients from a reference population database are applied to the discharge records to compute a predicted value.

For indicators that are not risk-adjusted, this is the reference population rate. The expected rate is computed as the sum of the predicted value for each record divided by the number of records flagged in the population at risk for the unit of analysis of interest (i.e., hospital). 5) Calculate risk-adjusted rate using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate. For indicators that are not risk-adjusted, this is the same as the observed rate. 6) Calculate smoothed rate using an Empirical Bayes shrinkage estimator (W) as the weighted average of the risk-adjusted rate and the reference population rate. The shrinkage estimate reflects a reliability adjustment unique to each indicator.

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:

2a1.24 Sampling (Survey) Methodology. If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):

2a1.25 Data Source *(Check all the sources for which the measure is specified and tested*). If other, please describe:
Administrative claims

2a1.26 Data Source/Data Collection Instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*):
N/A

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment:
State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality, Rockville, MD. The SID contain the universe of inpatient discharge abstracts in participating States, translated into uniform format to facilitate analyses. With data from 47 States, the SID, in the aggregate, currently encompass about 97 percent of all annual discharges in the U.S. (http://www.hcup-us.ahrq.gov/sidoverview.jsp#What).
Data dictionary and code tables are available at
http://www.qualityindicators.ahrq.gov/Downloads/Software/WinQI/V44/Software%20Instructions%20(WinQI)%20V4.4.pdf
A data dictionary for the source data, HCUP SID, is available at: http://www.hcup-us.ahrq.gov/db/state/siddist/sid_multivar.jspIncluded in attached appendix

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:
Available in attached Excel or csv file
Risk_Adjustment_Coefficients_for_GI_Hemorrhage_Mortality_Rate-634935176882353950.xlsx

2a1.33 Level of Analysis (*Check the levels of analysis for which the measure is specified and tested)*:
Facility

2a1.34-35 Care Setting *(Check all the settings for which the measure is specified and tested)*:
Hospital/Acute Care Facility

If other: N/A

2a. RELIABILITY. Precise Specifications and Reliability Testing: H☐ M☐ L☐ I ☐

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

SEE ATTACHED MEASURE TESTING FORM

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

If the Committee votes No, STOP

3. USABILITY

Extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations. (evaluation criteria)

3.1 Current and Planned Use *(NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported withi n 6 years of initial endorsement in addition to performance improvement.)*:

Current and Planned Use (check all the current and planned uses; for any current uses that are checked, provide a URL for the specific program)

Planned	Current	For current use, Provide URL
	Public Reporting;Public Health/ Disease Surveillance;Quality Improvement with Benchmarking (external benchmarking to multiple organizations);Quality Improvement (Internal to the specific organization)	http://www.oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/IQI/AHRQ_IMI_2009.pdf;http://www.ahrq.gov/qual/qrdr11.htm;https://www.uhc.edu/;https://www.premierinc.com/quest/

3a. Accountability and Transparency: H☐ M☐ L☐ I ☐
(Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.)

3a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Hospital Inpatient Morality Indicators for California
State of California Office of Statewide Planning and Development
For 2009, data are from 336 out of 466 general acute care California-licensed hospitals
National Healthcare Quality Report
Agency for Healthcare Research and Quality
2010 data are from 45 participating states in the Healthcare Cost and Utilization Project (HCUP) database
UHC Imperatives for Quality (IQ) Program
Measuring performance at over 200 academic medical centers across the country
Premier alliance
QUEST hospital collaborative
Including more than 340 hospitals across 40 states

3a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? *(e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?)*
N/A

3a.3 If not currently publicly reported OR used in at least one accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. *(Credible plan includes the specific program, purpose, intended audience, and timeline for*

implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.)
N/A

3b. Improvement: H☐ M☐ L☐ I ☐
(Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.6 If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.)

3b.1. Provide data that demonstrate improvement in performance and/or health. (Not required for initial endorsement unless available.)
Include:

- Source of Data
- Geographic area and number and percentage of accountable entities and patients included
- Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)

Data are available at the following link and recorded below:
http://hcupnet.ahrq.gov/HCUPnet.jsp?Parms=H4sIAAAAAAAAACtN8wz0DAly9XMJTktLC05KTcoBCiQVG6aBQSKQY2gBZgEAdjRn4SoAAAA724CC342D44837FD68021C3D98C79D5CF07CEC3B&JS=Y
Deaths per 1,000 hospital admissions with gastrointestinal hemorrhage, age 18 and over (IQI 18)
By Year
YEAR – RATE – STD ERR - P VS 1994 – P VS PREVIOUS YEAR
2010 17.468 0.173 0.000 0.000
2009 19.363 0.180 0.000 0.000
2008 22.294 0.187 0.000 0.000
2007 23.610 0.202 0.000 0.000
2006 27.634 0.210 0.000 0.000
2005 30.173 0.216 0.000 0.000
2004 33.356 0.220 0.000 0.000
2003 37.427 0.226 0.000 0.000
2002 40.457 0.234 0.000 0.000
2001 43.688 0.239 0.000 0.000
2000 45.694 0.248 0.000 0.000
1997 49.125 0.262 0.000 0.000

3b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations:
N/A

3c. Unintended Consequences: H☐ M☐ L☐ I ☐
(The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations)

3c.1. Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.
No unintended negative consequences identified.

Overall, to what extent was the criterion, *Usability*, met? H☐ M☐ L☐ I ☐
Provide rationale based on specific subcriteria:

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)

4a. Data Generated as a Byproduct of Care Processes: H☐ M☐ L☐ I ☐

4a.1-2 How are the data elements needed to compute measure scores generated? *(Check all that apply).*
Data used in the measure are:
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

4b. Electronic Sources: H☐ M☐ L☐ I ☐

4b.1 Are the data elements needed for the measure as specified available electronically *(Elements that are needed to compute measure scores are in defined, computer-readable fields)*:
ALL data elements are in defined fields in electronic claims

4b.2 If ALL data elements are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources:
N/A

4d. Data Collection Strategy/Implementation: H☐ M☐ L☐ I ☐

4d.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues *(e.g., fees for use of proprietary measures)*:
Administrative data are collected as part of the routine operations. Some staff time is required to download and execute the software from the AHRQ website, which is available at no cost.

4d.2 Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*:
None. The software/programming necessary to calculate the measure is available at no cost from the AHRQ QI website. The Windows version requires only a PC running the Windows operating system and a public use version of the SQL Server and *.NET. The SAS version requires a license for Base SAS only. Both versions of the software include at no cost a limited license version of the APR-DRG grouper.

Overall, to what extent was the criterion, *Feasibility*, met? H☐ M☐ L☐ I ☐
Provide rationale based on specific subcriteria:

OVERALL SUITABILITY FOR ENDORSEMENT

Does the measure meet all the NQF criteria for endorsement? Yes☐ No☐
Rationale:

If the Committee votes No, STOP.
If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

5. COMPARISON TO RELATED AND COMPETING MEASURES

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure before a final recommendation is made.

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

5.1 If there are related measures *(either same measure focus or target population)* or competing measures *(both the same measure focus and same target population*), list the NQF # and title of all related and/or competing measures:

5a. Harmonization

5a.1 If this measure has EITHER the same measure focus OR the same target population as NQF-endorsed measure(s): Are the measure specifications completely harmonized?

5a.2 If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden:
N/A

5b. Competing Measure(s)

5b.1 If this measure has both the same measure focus and the same target population as NQF-endorsed measure(s): Describe why this measure is superior to competing measures *(e.g., a more valid or efficient way to measure quality);* OR provide a rationale for the additive value of endorsing an additional measure. *(Provide analyses when possible)*:
N/A

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner): Agency for Healthcare Research and Quality

Co.2 Point of Contact: Pamela | Owens | Pam.Owens@AHRQ.hhs.gov | 301-427-1412-

Co.3 Measure Developer if different from Measure Steward: Agency for Healthcare Research and Quality

Co.4 Point of Contact: Pamela | Owens | Pam.Owens@AHRQ.hhs.gov | 301-427-1412

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Multi-specialty Panel and Surgical Panel members are listed in the technical report:
http://www.qualityindicators.ahrq.gov/Downloads/Modules/IQI/iqi_development.zip

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.3 Year the measure was first released:
Ad.4 Month and Year of most recent revision:
Ad.5 What is your frequency for review/update of this measure?
Ad.6 When is the next scheduled review/update for this measure?

Ad.7 Copyright statement: N/A

Ad.8 Disclaimers: N/A

Ad.9 Additional Information/Comments:

Date of Submission (*MM/DD/YY*): Jul 16, 2012

NATIONAL QUALITY FORUM—Evidence (1c) Pilot Submission Form

Measure Title: Gastrointestinal Hemorrhage Mortality Rate (IQI #18)
Date of Submission: 7/9/2012

- Respond to all questions with answers immediately following the question.
- Maximum of 6 pages (*6 pages incudes questions/instructions in the form*); minimum font size 11 pt
- All information needed to demonstrate meeting the evidence criterion (1c) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- See NQF guidance on evaluating evidence. Contact NQF staff for examples, resources, or questions.

STRUCTURE-PROCESS-OUTCOME RELATIONSHIP

1c.1.This is a measure of:

Outcome

☒ Health outcome: Inpatient mortality

☐ Intermediate clinical outcome: 2T

☐ Process: 2T

☐ Structure: 2T

☐ Other: 2T

HEALTH OUTCOME MEASURE *If not a health outcome, skip to 1c.3*

If the measure focus identified in 1c.1 is a health outcome, answer 1c.2 and 1c.2.1.

1c.2. Briefly state or diagram how the health outcome is related to at least one healthcare structure, process, intervention, or service.

Admission for gastrointestinal hemorrhage is fairly common (circa 100/100,000 adults/year). Mortality is generally regarded as an undesirable outcome of hospital care for this condition, as for many other conditions and procedures (e.g., acute myocardial infarction, heart failure, pneumonia, stroke), although there is a small subset of patients for whom death may be the expected outcome.

Multiple care processes can influence the course of a patient during a hospital stay for gastrointestinal hemorrhage, including but not limited to:

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 rebleeding 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.

Mortality rates for GI hemorrhage vary greatly, and lower mortality has been associated with more use of treatments such as early endoscopy (within 24-48 hours of presentation), though the strength of this relationship has not been established, with some studies failing to find significant relationships. Mortality rates in large population based databases have not changed since the 1940s, though there have been increases in the ages and comorbidities of patients that may have offset mortality rate declines due to better quality of care.

1c.2.1. **State the rationale supporting the relationship between the health outcome and at least one healthcare structure, process, intervention, or service**.

A number of medical treatments have been shown to be associated with bleeding control among patients admitted with acute GI hemorrhage, although evidence on mortality is more limited. One meta-analysis showed a slight advantage for early endoscopy versus medical management among unselected patients with acute nonvariceal upper GI hemorrhage,[8] although some individual studies have failed to find significant associations in multivariate analyses.[2]

Recent attention has focused on patients with hemorrhage due to bleeding esophageal varices, who have a particularly high risk of death. A meta-analysis of 12 randomized trials of beta blockers showed a 21% improvement in the percentage of patients free of rebleeding (RR 1.42), a 5.4% improvement in the mean survival rate (RR 1.27), and 7.4% improvement in the mean percentage of patients free of bleeding death (RR 1.50).[9] Eight trials evaluated the effects of antibiotic prophylaxis compared with placebo or no antibiotic prophylaxis in 864 cirrhotic patients with upper gastrointestinal hemorrhage; significant beneficial effects on mortality (RR 0.73 [95% CI, 0.55 to 0.95]) and the incidence of bacterial infections (RR 0.40 [95% CI 0.32 to 0.51]) were observed.[10] Vasoactive agents such as terlipressin also significantly reduce mortality (RR 0.66 [95% CI, 0.49 to 0.88]) relative to placebo,[11] but not relative to endoscopic sclerotherapy.[12] A meta-analysis of 23 randomized trials, with 1860 patients, comparing endoscopic plus beta-blocker therapy with either therapy alone, showed that combination therapy reduced overall rebleeding, variceal bleeding, and variceal recurrence more than either endoscopic or beta-blocker therapy alone. Mortality after combination therapy was nonsignificantly lower than that after endoscopic (odds ratio, 0.78 [95% CI, 0.58 to 1.07] or drug therapy alone (odds ratio, 0.70 [95% CI, 0.46 to 1.06]).[13]

These findings from randomized controlled trials and meta-analyses have been incorporated into recent practice guidelines from the American College of Gastroenterology and the American Association for the Study of Liver Diseases.[14] Their Class 1 recommendations include:

1. Acute GI hemorrhage in a patient with cirrhosis is an emergency that requires prompt attention with intravascular volume support and blood transfusions, being careful to maintain a hemoglobin of 8 g/dL (Class I, Level B).

2. Short-term (maximum 7 days) antibiotic prophylaxis should be instituted in any patient with cirrhosis and GI hemorrhage (Class I, Level A)….
3. Pharmacological therapy (somatostatin or its analogues octreotide and vapreotide; terlipressin) should be initiated as soon as variceal hemorrhage is suspected and continued for 3-5 days after diagnosis is confirmed (Class I, Level A).
4. EGD, performed within 12 hours, should be used to make the diagnosis and to treat variceal hemorrhage, either with EVL or sclerotherapy (Class I, Level A).
5. TIPS is indicated in patients in whom hemorrhage from esophageal varices cannot be controlled or in whom bleeding recurs despite combined pharmacological and endoscopic therapy (Class I, Level C).
6. Balloon tamponade should be used as a temporizing measure (maximum 24 hours) in patients with uncontrollable bleeding for whom a more definitive therapy (e.g., TIPS or endoscopic therapy) is planned (Class I, Level B).

Similarly, from the World Gastroenterological Association's evidence-based guidelines:
"Management of Acute Variceal Hemorrhage in Patients with Cirrhosis"
Resuscitation measures include:
1. Intravenous (IV) volume support
2. Blood transfusion
3. Correct severe coagulation/platelet deficits
4. Antibiotic prophylaxis (up to 7 days):
5. Oral norfloxacin (400 mg twice daily [BID]), or
6. IV ciprofloxacin (400 mg BID), or
7. IV ceftriaxone (1 g/day) in advanced cirrhosis

Pharmacological therapy includes:
1. Continue 3-5 days after confirmed diagnosis
2. Somatostatin (terlipressin or octreotide, vapreotide)

Within 12 hours:
1. Confirm diagnosis with EGD
2. Treat variceal hemorrhage with EVL or sclerotherapy

In uncontrollable bleeding or recurrence:
1. TIPS indicated

In uncontrollable bleeding while waiting for TIPS or endoscopic therapy:
1. Balloon tamponade as temporizing measure for 24 hours maximum."[15]

Many of the deaths reported among GI hemorrhage are not associated with bleeding per se. One study found that only one such death was directly related to bleeding, and that patient had several severe comorbidities.[3] In many cases, deaths among patients with a principal diagnosis of gastrointestinal hemorrhage are due to infectious or cardiovascular complications of the hemorrhage or the underlying condition (e.g., chronic liver disease, cancer) and not primarily due to the acute hemorrhage itself. Among patients with bleeding from esophageal varices, death rates are higher and appear to be more closely related to blood loss and interventions to minimize blood loss.[7,16] However, appropriate risk stratification, early stabilization, ongoing monitoring, and measures to prevent infectious and cardiovascular complications (e.g., central line bundle to prevent central line associated bloodstream

infections, sepsis bundle to detect early signs of sepsis and respond appropriately) appear to have favorable effects on all hospitalized patients at risk, including patients with GI hemorrhage.

References:

1. Rockall TA, Logan RF, Devlin HB, Northfield TC. Variation in outcome after acute upper gastrointestinal haemorrhage. The National Audit of Acute Upper Gastrointestinal Haemorrhage. Lancet. 1995;346(8971):346-350.
2. Cooper GS, Chak A, Way LE, Hammar PJ, Harper DL, Rosenthal GE. Early endoscopy in upper gastrointestinal hemorrhage: associations with recurrent bleeding, surgery, and length of hospital stay. Gastrointest Endosc. 1999;49(2):145-152.
3. Longstreth GF. Epidemiology and outcome of patients hospitalized with acute lower gastrointestinal hemorrhage: a population-based study. Am J Gastroenterol. 1997;92(3):419-424.
4. Rockall TA, Logan RF, Devlin HB, Northfield TC. Selection of patients for early discharge or outpatient care after acute upper gastrointestinal haemorrhage. National Audit of Acute Upper Gastrointestinal Haemorrhage. Lancet. 1996;347(9009):1138-1140.
5. Rockall TA, Logan RF, Devlin HB, Northfield TC. Incidence of and mortality from acute upper gastrointestinal haemorrhage in the United Kingdom. Steering Committee and members of the National Audit of Acute Upper Gastrointestinal Haemorrhage. Bmj. 1995;311(6999):222-226.
6. Hay JA, Lyubashevsky E, Elashoff J, Maldonado L, Weingarten SR, Ellrodt AG. Upper gastrointestinal hemorrhage clinical--guideline determining the optimal hospital length of stay. Am J Med. 1996;100(3):313-322.
7. Dy SM, Cromwell DM, Thuluvath PJ, Bass EB. Hospital experience and outcomes for esophageal variceal bleeding. Int J Qual Health Care. Apr 2003;15(2):139-146.
8. Cook DJ, Guyatt GH, Salena BJ, Laine LA. Endoscopic therapy for acute nonvariceal upper gastrointestinal hemorrhage: a meta-analysis. Gastroenterology. 1992;102(1):139-148.
9. Bernard B, Lebrec D, Mathurin P, Opolon P, Poynard T. Beta-adrenergic antagonists in the prevention of gastrointestinal rebleeding in patients with cirrhosis: a meta-analysis. Hepatology. Jan 1997;25(1):63-70.
10. Soares-Weiser K, Brezis M, Tur-Kaspa R, Leibovici L. Antibiotic prophylaxis for cirrhotic patients with gastrointestinal bleeding. Cochrane Database Syst Rev. 2002(2):CD002907.
11. Ioannou G, Doust J, Rockey DC. Terlipressin for acute esophageal variceal hemorrhage. Cochrane Database Syst Rev. 2003(1):CD002147.
12. D'Amico G, Pietrosi G, Tarantino I, Pagliaro L. Emergency sclerotherapy versus vasoactive drugs for variceal bleeding in cirrhosis: a Cochrane meta-analysis. Gastroenterology. May 2003;124(5):1277-1291.
13. Gonzalez R, Zamora J, Gomez-Camarero J, Molinero LM, Banares R, Albillos A. Meta-analysis: Combination endoscopic and drug therapy to prevent variceal rebleeding in cirrhosis. Ann Intern Med. Jul 15 2008;149(2):109-122.
14. Garcia-Tsao G, Sanyal AJ, Grace ND, Carey W. Prevention and management of gastroesophageal varices and variceal hemorrhage in cirrhosis. Hepatology. Sep 2007;46(3):922-938.
15. World Gastroenterology Organisation (WGO). Esophageal varices. Munich (Germany): World Gastroenterology Organisation (WGO); 2008 Jun. 17 p.
16. Qureshi W, Adler DG, Davila R, Egan J, Hirota W, Leighton J, Rajan E, Fanelli R, Wheeler-Harbaugh J, Baron TH, Faigel DO, Standards of Practice Committee. ASGE Guideline: the role of endoscopy in the management of variceal hemorrhage, updated July 2005. Gastrointest Endosc 2005 Nov;62(5):651-5.

Note: For health outcome measures, no further information is required

STRUCTURE, PROCESS, OR INTERMEDIATE OUTCOME MEASURE

If the measure focus identified in 1c.1 is a structure, process, or intermediate outcome answer all the following questions (except as indicated by skip pattern).

1c.3. Briefly state or diagram how the measure focus is related to desired health outcomes and proximity to desired health outcomes. *(Do not summarize the evidence here.)*

Not applicable

1c.4. Is there a guideline recommendation supporting the measure focus identified in 1c.1.? Yes☐ No☐
If no, skip to #1c.6

If yes, *answer 1c.4.1-1c.5.*

1c.4.1. **Guideline citation** (*including date*):

1c.4.2. **URL** (*if available online*):

1c.4.3. **Identify guideline number and/or page number**:

1c.4.4. **Quote verbatim, the specific guideline recommendation**:

1c.4.5. **Grade assigned to the recommendation with definition of the grade**:

1c.5. Did the guideline developer systematically review and grade the body of evidence for the specific guideline recommendation? Yes☐ No☐ *If no, skip to #1c.6*

If yes, *answer 1c.5.1. (**Note:** Findings of the systematic review of the body of evidence for the guideline recommendation must be reported in 1c.8-1c.13.)*

1c.5.1. **Grade assigned to the body of evidence with definition of the grade**:

1c.6. Is there another published systematic review of the body of evidence supporting the measure focus identified in 1c.1? *(other than from the guideline cited above, e.g., Cochrane, AHRQ, USPSTF)*
Yes☐ No☐ *If no, skip to #1c.7*

If yes, *answer 1c.6.1-1c.6.3. (**Note:** Findings of the systematic review of the body of evidence must be reported in 1c.8-1c.13.)*

1c.6.1. **Citation** (*including date*):

1c.6.2. **URL** (*if available online*):

1c.6.3. **Grade assigned to the body of evidence with definition of the grade**:

If a systematic review of the evidence was identified in either 1c.5 or 1c.6, skip to 1c.8

1c.7. If a systematic review of the body of evidence was not identifed and reported in 1c.5 or 1c.6, did the measure developer perform a systematic review of the body of evidence supporting the measure focus identified in 1c.1? Yes☐ No☐

If yes, *answer 1c.7.1-1c.7.3. (**Note:** Findings of the measure developer's systematic review of the body of evidence must be reported in 1c.8-1c.13 and unpublished evidence review products such as evidence tables provided in an appendix.)*

1c.7.1. **Who conducted the measure developer's systematic review of the body of evidence**?

1c.7.2. **Grade assigned to the body of evidence with definition of the grade**:

1c.7.3. **Describe the process used for the systematic review**:

If no systematic review of the body of evidence identified in 1c.5, 1c.6, or 1c.7, the evidence criterion can not be met.

FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE FOCUS

(Items 1c.8-1c.13 must be answered and should support the measure focus identified in 1c.1. If more than one systematic review was identified (1c.5, 1c.6, and 1c.7), provide a separate response for each.)

1c.8. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: 2T

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1c.9. How many and what type of study designs are inlcuded in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

1c.10. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect due to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1c.11. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

1c.12. What harms were studied and how do they affect the net benefit—benefits over harms?

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1c.13. Are there new studies that have been conducted since the systematic review(s) of the body of evidence? Yes☐ No☐ *If no, stop*

If yes,

1c.13.1. **For each new study provide: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review**.

Measure Testing to Demonstrate Scientific Acceptability of Measure Properties

Measure Title: Gastrointestinal Hemorrhage Mortality Rate (IQI #18)
Date of Submission: 1/7/2012
Type of Measure:

☐ Composite	☐ Outcome
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

This Word document template must be used to submit information for measure testing.

- **For all measures, sections 1, 2a2, 2b2, 2b3, 2b5 must be completed**
- **For outcome or resource use measures**, section **2b4** also must be completed
- If specified for **multiple data sources** (e.g., claims and medical records), section **2b6** also must be completed
- Respond to all questions with answers immediately following the question (*unless meet the skip criteria or those that are indicated as optional*).
- Maximum of 10 pages (*incuding questions/instructions; do not change margins or font size; contact project staff if need more pages*)
- All information on testing to demonstrate meeting the criteria for scientific acceptability of measure properties (2a,2b) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing, (e.g., reliability vs. validity) be sure to indicate the specific differences in question 7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the types of data specified and intended for measure implementation*)

Measure Specified to Use Data From:	Measure Tested with Data From:
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure implemented in electronic health record	☐ eMeasure implemented in electronic health record
☐ other: 3T	☐ other: 3T

1.2. If used an existing dataset, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality, Rockville, MD. The SID consists of approximately 30 million adult discharges and 4,000 hospitals.

1.3. What are the dates of the data used in testing? 2008 (Version 4.4; March 2012); 2010 (Version 4.5; March 2013)

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)
☐ **individual clinician** ☐ **group/practice** ☐ **hospital/facility/agency** ☐ **health plan**
☐ **other**: 3T

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

N = 4,033. The reference population for Version 4.4 of the AHRQ QI software is based on the 2008 State Inpatient Databases (SID). The 2008 SID includes all community hospitals from 42 states.

The hospital universe is defined as all hospitals located in the U.S. that are open during any part of the calendar year and designated as community hospitals in the AHA Annual Survey Database (Health Forum, LLC © 2011). The AHA defines community hospitals as follows: "All non-Federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions." Starting in 2005, the AHA included long term acute care facilities in the definition of community hospitals. These facilities provide acute care services to patients who need long term hospitalization (stays of more than 25 days). Consequently, Veterans Hospitals and other Federal facilities (Department of Defense and Indian Health Service) are excluded. Beginning in 1998, we excluded short-term rehabilitation hospitals from the universe because the type of care provided and the characteristics of the discharges from these facilities were markedly different from other short-term hospitals.
http://hcup-us.ahrq.gov/db/nation/nis/NISIntroduction2010.pdf

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

N = 458,307

Characteristic		Frequency	Characteristic		Frequency
Age	18 to 59	0.2725	APR-DRG	'2421' to '2423'	0.0447
Age	60 to 64	0.0766	APR-DRG	'2424'	0.0016
Age	65 to 85+	0.6509	APR-DRG	'2441' to '2442'	0.1294
APR-DRG	'2201'	0.0021	APR-DRG	'2443'	0.0150
APR-DRG	'2202'	0.0034	APR-DRG	'2444'	0.0020
APR-DRG	'2203'	0.0023	APR-DRG	'2531'	0.1323

APR-DRG	'2204'	0.0015	APR-DRG	'2532'	0.1750
APR-DRG	'2211'	0.0043	APR-DRG	'2533'	0.0682
APR-DRG	'2212'	0.0047	APR-DRG	'2534'	0.0114
APR-DRG	'2213'	0.0025	APR-DRG	'2541' to '2534'	0.0668
APR-DRG	'2214'	0.0010	APR-DRG	'2544'	0.0014
APR-DRG	'2411' to '2413'	0.2849	MDC	OTHER	0.0365
APR-DRG	'2414'	0.0089	TRNSFER	Transfer-in	0.0145

Descriptions of APR-DRG characteristics:
220 Major Stomach, Esophageal & Duodenal Procedures
221 Major Small & Large Bowel Procedures
241 Peptic Ulcer & Gastritis

Descriptions of Risk of Mortality Subclass:
1 Minor
2 Moderate
3 Major
4 Extreme

The full list of descriptors for the APR-DRG characteristics are listed here:
http://hcup-us.ahrq.gov/db/nation/nis/v261_aprdrg_meth_ovrview.pdf

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Not applicable

_____

2a2. RELIABILITY TESTING
Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – report validity of data elements in 2b2*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability*)
☐ **Performance measure score** (e.g., *signal-to-noise*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Our metric of reliability is the signal to noise ratio, which is the ratio of the between hospital variance (signal) to the within hospital variance (noise). The formula is signal / (signal + noise). There is hospital-specific signal to noise ratio, which is used as an Empirical Bayes univariate shrinkage estimator. The overall signal to noise ratio is a weighted average of the hospital-specific signal-to-noise ratio, where the weight is [1 / (signal+noise)^2]. The signal is calculated using an iterative method.

The analysis reports the reliability of the risk-adjusted rate (before applying the empirical Bayes univirate shrinkage estimator).

2a2.3. For each level checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis and association with case volume*)

Size Decile	Number of Hospitals	Ave. Number of Patients per Hospital in Decile	Ave. Signal-to-Noise Ratio for Hospitals in Decile	Percent of Signal Variance Explained by Performance Score
1	404	2.2	0.00784	0.25538
2	403	7.9	0.01840	0.26128
3	403	17.4	0.03937	0.27338
4	404	32.3	0.07589	0.29436
5	403	55.3	0.13402	0.32887
6	403	86.1	0.20778	0.37477
7	404	126.0	0.28276	0.42245
8	403	176.6	0.35913	0.47308
9	403	241.0	0.44316	0.53150
10	403	391.9	0.56323	0.62033
Overall	4,033	458,307	0.46902	0.55332

Note: The aveage "signal-to-noise" ratio for the hospitals in the decile. As with the overall ratio, the decile-specific ratio is weighted by 1 / (signal+noise)^2.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The ratio itself is only a diagnostic for the degree of variance in the risk-adjusted rate systematically associated with the provider. Therefore, what matters is the magnitude of the variance in the "smoothed" rate (that is, the variance in the risk-adjusted rate after the application of the univariate shrinkage estimator based on the signal ratio).

We do not consider reliability a threshold criterion, where measures with a ratio above X are consdiered relilable, and measures with a ratio below X are considered not reliable. Rather the reliablity ratio is a continuous value used as a weight W, and what matters is the amount of variation remaining in the performance score after the weight is applied, where the performance score = W * risk-adjusted rate + (1-W) * reference population rate

We also report the percentage of the signal variance "explained" by the variance in the performance score. The interpretation of a "low" percentage is that there is room to improve the reliablity of the measure through more cases per hosptial or "borrowing" strength from related process or outcome measures or an improved prior probability.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements**

☐ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance quality or resource use and can distinguish performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

We conduct construct validity testing to examine the association between the risk-adjusted mortality rate and hospital structural characteristics potentially associated with quality of care, including prior performance, using regression analysis.

The hypothesized relationship is as follows:
1) Volume – higher volume is associtaed with better outcomes, either because practice makes perfect (volume causes outcome) or referral (outcome causes volume)
2) Reservation quality – higher reservation quality is associated with better outcomes because reservation quality is associated with excess capacity
3) Transfer out – higher transfer out rate is associated with worse outcomes because transferred cases have higher risk of mortality
4) Diagnosis codes – More reported diagnosis codes are associated with more reported comorbidities, therefore higher expected rates, there fore better outcomes

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test, ANOVA*)

Table 1. Structure Measures Used to Estimate Prior Probability

Measure	How it is measured	Rationale
Ln(Volume)	Natural log of the denominator	Practice makes perfect or referral
Reservation Quality	Inverse of average daily census (ADC)	Reflects the excess capacity in the inputs of production (e.g. nurse staffing)
Transfer Out	Overall percent transfer out	Routine transferring of particular categories of patients
Maximum DX	Maximum reported diagnosis codes	Higher prevalence and co-morbidities
Prior Performance	Prior year smoothed rate	Share of performance likely to persist

Table 2. Hospital Level Regression without Prior Performance

Variable	Label	Coef.	Std. Err	. t	P>|t|	[95% Conf.	Interval]
lnvol	Ln(Volume)	-0.0022	0.0005	-4.1500	0.0000	-0.0032	-0.0012
adcinv	Reservation Quality	0.0025	0.0112	0.2200	0.8230	-0.0195	0.0245
trnsout	Transfer Out	0.0338	0.0151	2.2400	0.0250	0.0042	0.0633
maxdx	Maximum DX	-0.0001	0.0000	-2.5100	0.0120	-0.0001	0.0000
_cons	Constant	0.0370	0.0032	11.6500	0.0000	0.0308	0.0432

Note: the dependent variable in the regression is a rate/proportion

Table 3. Hospital Level Regression with Prior Performance

Variable	Label	Coef.	Std. Err	t	P>|t|	[95% Conf.	Interval]
lnvol	Ln(Volume)	-0.0007	0.0005	-1.4200	0.1550	-0.0018	0.0003
adcinv	Reservation Quality	0.0118	0.0112	1.0500	0.2920	-0.0101	0.0337
trnsout	Transfer Out	0.0163	0.0149	1.0900	0.2750	-0.0130	0.0455
maxdx	Maximum DX	-0.0001	0.0000	-2.0500	0.0410	-0.0001	0.0000
prior2	Prior Performance	0.6576	0.0516	12.7500	0.0000	0.5565	0.7587
_cons	Constant	0.0134	0.0034	3.9200	0.0000	0.0067	0.0201

Note: the dependent variable in the regression is a rate/proportion

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

There is a significant (negative) association between the hospital risk-adjusted mortality rate and the hospital volume, indicative of a volume-outcome relationship. The direction of causality is unknown. There is also a (positive) association between the hospital risk-adjusted mortality rate and the hospital transfer-out rate (overall) meaning that hospitals that transfer-out a higher proportion of their patients overall have worse quality of care. Finally, there is a (negative) association between the hospital risk-adjusted mortality rate and the number of diagnosis codes reported meaning potentially that hospitals that report more codes have more reported co-morbidities (and therefore have a higher expected rate and a lower risk adjusted rate).

With the exception of the results for the number of reported diagnosis codes, the volume effect and the transfer out effect are not statistically significant once prior performance is taken into account. This means that structural characteristics and the associated relationship with mortality are highly persistent, although the mortality rate itself is not that persistent.

The volume-outcome relationship supports the validity of the measure; as does the persistence of performance over time.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to #2b5*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The only exclusion is for cases transferred to an acute care hospital (there are other listed exclusions for MDC 14 and missing data but these exclusions impact a non material number of cases). The rational for this exclusion is that we do not observe the outcome of interest (i.e. in-hospital death). However, we do observe the outcome of interest for cases transferred from an acute care hospital. We conducted a regression analysis to determine whether cases transferred from an acute care hospital have higher in-hospital mortality rates than cases not transferred.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
TRNSFER	Transfer-in	1	0.6387	0.0633	101.67	0.0000

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Even after accounting for demographics, severity and co-morbidities the cases transferred in have significantly higher risk-adjusted in-hospital mortality rates (OR=1.894). Therefore the transfer-in status is included in our risk-adjustment model.

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used)*

We calculate the posterior probability distribution for each hospital parameterized using the Gamma distribution. We then calculate the probability that the hospital is better or worse than the reference population rate at a 95 percent probability overall and by hospital size decile.

The analysis is with the computed performance scores for the measure as specified (including shrinkage estimator).

2b5.2. What were the statistical results from testing the ability to identify differences in performance measure scores across measured entities? (*at a minimum, the distribution of performance measure scores for the measured entities by decile/quartile, mean, std dev; preferably also number and percentage statistically different from mean or some benchmark, different form expected, etc.*)

Size Decile	Hospitals	Ave. Patients	Percent Better	Percent Worse
1	404	2.2	0.00000	0.56188
2	403	7.9	0.00000	0.49380
3	403	17.4	0.00000	0.20099
4	404	32.3	0.00000	0.10396
5	403	55.3	0.00000	0.09429
6	403	86.1	0.00000	0.07692
7	404	126.0	0.00248	0.07673
8	403	176.6	0.01489	0.06203
9	403	241.0	0.03722	0.05955
10	403	391.9	0.10670	0.07692
	4,033	458,307	0.01612	0.18076

Size Decile	Hospitals	Ave. Patients	Percent Better	Percent Worse
Patient weighted			0.05083	0.07969

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean and what are the norms for the test conducted?*)

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. Therefore the measure provides useful information on hospital performance for selection and change.

_____

If not an intermediate or health outcome or resource use measure, this section can be deleted

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

2b4.1. What method of controlling for differences in case mix is used?

☐ **Statistical risk model with 25 risk factors**
☐ **Stratification by 3T risk categories**
☐ **No risk adjustment or stratification**
☐ **Other,** 3T

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select factors used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher*)

The Inpatient Quality Indicators (IQI) use a standard set of risk factors based on gender, age in 5-year age groups, severity as defined by APR-DRG and comorbidity as defined by the APR-DRG risk-of-mortality subclass. There are additional covariates tested such as transfer-in status and the availability of data elements such as point of origin or procedure days.

The selection criteria are whether there are at least 30 cases in the numerator for the covariate, and then whether the coefficient on the covariate is statistically significant (p<.05). If not statistically significant, then covariates are pooled along the risk gradient (e.g. proximate age categories or risk-of-mortality subclass). If there is no gradient (e.g. APR-DRG) then covariates are pooled in a hierarchy. For example, APR-DRG are pooled into Major Diagnosis Category (MDC).

Variables are eliminated if not statisically significant (p<.05). If the covariate has a risk gradient (e.g. the risk increases with age) then cojoining age categories (e.g. 60-64, 65-69) that are not significant are

combined with age categoires that are significant (or with the omitted category). If the category is hiearhical (e.g. risk-of-mortality subclass are grouped within APR-DRG, APR-DRG are grouped within MDC) then covariates that are not significant are pooled "up" the hierachy

2b4.4. What were the statistical results of the analyses used to select risk factors?

Parameter	Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Odds Ratio
Intercept		1	-5.2153	0.0697	5595.93	0	0.005
Age	18 to 59	1	-0.1733	0.0494	12.31	0.0004	0.841
Age	65 to 85+	1	0.1454	0.047	9.59	0.002	1.157
APR-DRG	'2201'	1	1.4052	0.2493	31.77	0	4.076
APR-DRG	'2202'	1	2.5315	0.1298	380.17	0	12.572
APR-DRG	'2203'	1	3.3609	0.118	811.71	0	28.815
APR-DRG	'2204'	1	5.0554	0.1156	1911.36	0	156.867
APR-DRG	'2211'	1	0.7506	0.2236	11.27	0.0008	2.118
APR-DRG	'2212'	1	2.0235	0.1373	217.22	0	7.565
APR-DRG	'2213'	1	3.0474	0.1224	619.54	0	21.061
APR-DRG	'2214'	1	5.1812	0.1305	1576.83	0	177.896
APR-DRG	'2411' to '2413'	1	0.3547	0.0618	32.92	0	1.426
APR-DRG	'2414'	1	3.7847	0.0734	2662.12	0	44.022
APR-DRG	'2421' to '2423'	1	0.6748	0.0884	58.22	0	1.964
APR-DRG	'2424'	1	4.2938	0.1081	1577.52	0	73.244
APR-DRG	'2441' to '2442'	1	-1.0253	0.1055	94.51	0	0.359
APR-DRG	'2443'	1	1.2019	0.1083	123.08	0	3.326
APR-DRG	'2444'	1	3.3347	0.12	772.1	0	28.070
APR-DRG	'2532'	1	0.9608	0.0621	239.43	0	2.614
APR-DRG	'2533'	1	2.5538	0.0612	1742.2	0	12.856
APR-DRG	'2534'	1	4.4473	0.0696	4085.73	0	85.396
APR-DRG	'2541' to '2534'	1	0.8521	0.0764	124.47	0	2.345
APR-DRG	'2544'	1	4.3364	0.1069	1645.74	0	76.432
MDC	OTHER	1	1.5317	0.0659	540.39	0	4.626
TRNSFER	Transfer-in	1	0.6387	0.0633	101.67	0	1.894

Descriptions of APR-DRG characteristics:

220 Major Stomach, Esophageal & Duodenal Procedures
221 Major Small & Large Bowel Procedures
241 Peptic Ulcer & Gastritis

Descriptions of Risk of Mortality Subclass:

1 Minor
2 Moderate
3 Major
4 Extreme

The full list of descriptors for the APR-DRG characteristics are listed here:
http://hcup-us.ahrq.gov/db/nation/nis/v261_aprdrg_meth_ovrview.pdf

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

When developing the model, we split the denominator into a development and validation sample. We select the covariates on the development sample, and conduct the empirical testing for discrimination and calibration on the validation sample.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics:

c-statistic 0.831

One calculates the c-statistic by taking all possible pairs of subjects consisting of one subject who experienced the event of interest and one subject who did not experience the event of interest. The c-statistic is the proportion of such pairs in which the subject who experienced the event had a higher predicted probability of experiencing the event than the subject who did not experience the event.

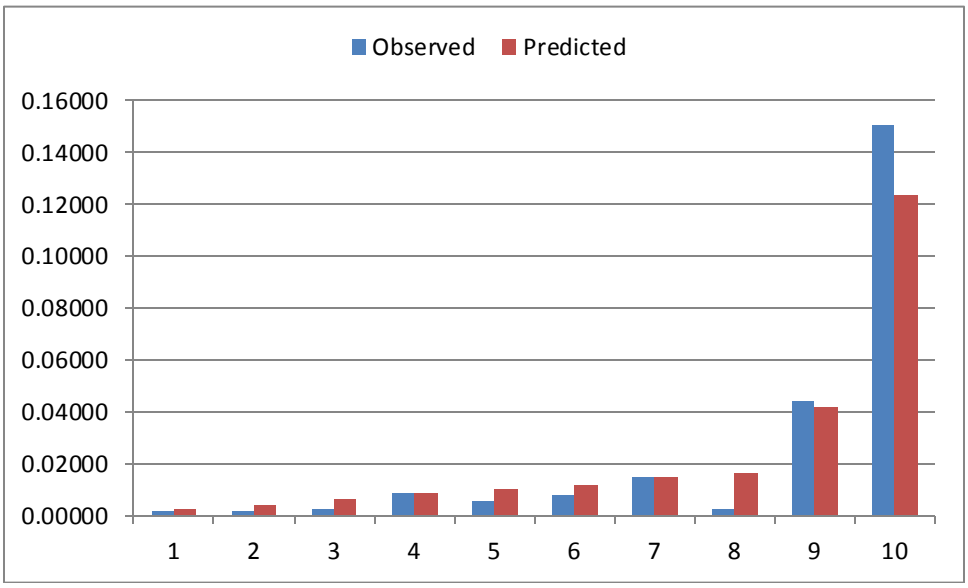
A model with a c-statistic above 0.80 is generally considered to provide good discrimination

2b4.7. Statistical Risk Model Calibration Statistics:

Decile	Patients	Observed	Predicted
1	45,831	0.00207	0.00251
2	45,831	0.00216	0.00446
3	45,831	0.00314	0.00698
4	45,830	0.00930	0.00924
5	45,831	0.00609	0.01030
6	45,831	0.00823	0.01227
7	45,830	0.01499	0.01485
8	45,831	0.00251	0.01664
9	45,831	0.04429	0.04210
10	45,830	0.15062	0.12405

A model that is well calibration will have observed values similar to predicted values across the predicted value deciles. Although there are statistical tests of such "goodness of fit" the tests generally are not informative for datasets with large sample sizes.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:



This is a visual representation of the data presented in the above section

2b4.9. Results of Risk Stratification Analysis:

The rate is also stratified into esophageal varices and all other cases for the purpose of facilitating quality improvement. Esophageal varices cases have higher mortality rates than other cases included in the denominator.

	Numerator	Denominator	Observed Rate
All Other Cases	9,861	431,024	0.02288
Esophageal Varices	1,949	27,282	0.07143

Although stratification is useful for quality improvement purposes, stratification is less useful for comparative reporting. The number of cases per hospital is small, with 70% of hospitals having 10 or fewer cases. In addition, the signal variance (the degree of systematic variance in the risk adjusted rate) is low, with a signal-to-noise ratio close to zero, and fewer than 1% of hospitals have rates above or below the reference population rate at 95% probability.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Based on the empirical testing, the model performs well on discrimination and calibration, which suggests that the model is adequate in terms of controlling for differences in patient characteristics. One potential area for concern is the association between the risk-adjusted rate and the number of diagnosis codes reported by the hospital.

***2b4.11. Optional Additional Testing** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods*)

With respect to the persistence of the hospital risk adjusted rate, we conducted a descriptive analysis to examine the distribution of the current year risk-adjusted rate by the prior year smoothed rate performance quintile.

Prior Year Quintile	Prior Year Smoothed Rate	Current Year Risk-adjusted Rate
1	0.01888	0.02043
2	0.02385	0.02592
3	0.02645	0.02663
4	0.02928	0.02989
5	0.03393	0.04160

The results demonstrate that performance is persistent over time, but there is some movement toward the mean among better performing hospitals. Worse performing hospitals tend to remain worse performing.

_____

If only one set of specifications, this section can be deleted

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical records and a different set of specifications for claims). It does not apply to measures that use more than one type of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).*

2b6.1. Describe the method of testing conducted to demonstrate equivalence of performance scores for the same entities across the different specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different specifications? (*e.g., correlation, rank order*)

Not applicable

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

Risk Adjustment Coefficients for IQI 18 - Gastrointestinal Hemorrhage Morality Rate

Parameter	Label	Code	Risk of Mortality Level	Descriptor	DF	Estimate	Standard Error	Wald Chi-Square	PR > Chi-Square
INTERCEPT					1	-5.2153	0.0697	5595.93	0.0000
AGE	18 to 59			AGE IN YEARS AT ADMISSION	1	-0.1733	0.0494	12.31	0.0004
AGE	65+			AGE IN YEARS AT ADMISSION	1	0.1454	0.0470	9.59	0.0020
APR-DRG	'2201'	220	1	MAJOR STOMACH, ESPOHAGEAL AND DUODENAL PROCEDURES	1	1.4052	0.2493	31.77	0.0000
APR-DRG	'2202'	220	2	MAJOR STOMACH, ESPOHAGEAL AND DUODENAL PROCEDURES	1	2.5315	0.1298	380.17	0.0000
APR-DRG	'2203'	220	3	MAJOR STOMACH, ESPOHAGEAL AND DUODENAL PROCEDURES	1	3.3609	0.1180	811.71	0.0000
APR-DRG	'2204'	220	4	MAJOR STOMACH, ESPOHAGEAL AND DUODENAL PROCEDURES	1	5.0554	0.1156	1911.36	0.0000
APR-DRG	'2211'	221	2	MAJOR SMALL AND LARGE BOWEL PROCEDURES	1	0.7506	0.2236	11.27	0.0008
APR-DRG	'2212'	221	3	MAJOR SMALL AND LARGE BOWEL PROCEDURES	1	2.0235	0.1373	217.22	0.0000
APR-DRG	'2213'	221	4	MAJOR SMALL AND LARGE BOWEL PROCEDURES	1	3.0474	0.1224	619.54	0.0000
APR-DRG	'2214'	221	5	MAJOR SMALL AND LARGE BOWEL PROCEDURES	1	5.1812	0.1305	1576.83	0.0000
APR-DRG	'2411' to '2413'	241	1 to 3	PEPTIC ULCER AND GASTRITIS	1	0.3547	0.0618	32.92	0.0000
APR-DRG	'2414'	241	4	PEPTIC ULCER AND GASTRITIS	1	3.7847	0.0734	2662.12	0.0000
APR-DRG	'2421' to '2423'	242	1 to 3	MAJOR ESOPHAGEAL DISORDERS	1	0.6748	0.0884	58.22	0.0000
APR-DRG	'2424'	242	4	MAJOR ESOPHAGEAL DISORDERS	1	4.2938	0.1081	1577.52	0.0000
APR-DRG	'2441' to '2442'	244	1 to 2	DIVERTICULITIS AND DIVERTICULOSIS	1	-1.0253	0.1055	94.51	0.0000
APR-DRG	'2443'	244	3	DIVERTICULITIS AND DIVERTICULOSIS	1	1.2019	0.1083	123.08	0.0000
APR-DRG	'2444'	244	4	DIVERTICULITIS AND DIVERTICULOSIS	1	3.3347	0.1200	772.1	0.0000
APR-DRG	'2532'	253	2	OTHER AND UNSPECIFIED GASTROINTESTINAL HEMORRHAGE	1	0.9608	0.0621	239.43	0.0000
APR-DRG	'2533'	253	3	OTHER AND UNSPECIFIED GASTROINTESTINAL HEMORRHAGE	1	2.5538	0.0612	1742.20	0.0000
APR-DRG	'2534'	253	4	OTHER AND UNSPECIFIED GASTROINTESTINAL HEMORRHAGE	1	4.4473	0.0696	4085.73	0.0000
APR-DRG	'2541' to '2534'	254	1 to 4	OTHER DIGESTIVE SYSTEM DIAGNOSES	1	0.8521	0.0764	124.47	0.0000
APR-DRG	'2544'	254	4	OTHER DIGESTIVE SYSTEM DIAGNOSES	1	4.3364	0.1069	1645.74	0.0000
MDC	Other			MAJOR DIAGNOSTIC CATEGORY	1	1.5317	0.0659	540.39	0.0000
TRANSFER	Transfer-in				1	0.6387	0.0633	101.67	0.0000

For more detailed information on the APR-DRG codes and descriptions, please go to www.aprdrgassign.com and login with UserID: NQFUser.

Technical Review of Scientific Acceptability of Measure Properties

#2065 Gastrointestinal Hemorrhage Mortality Rate (IQI #18)

03/25/13 Review by K. Pace
Overall, the measure submission for testing is very good. It provides the information requested in the appropriate fields. The only thing that is missing (probably related to how we asked the question in this pilot form) is the distribution of final risk-adjusted performance scores. Some additional explanation of scoring is needed for the specifications.

RELIABILITY
For which NQF rating is the measure eligible?
Testing was conducted for one level – performance score, so the measure is eligible for a moderate rating on reliability IF the testing is determined to be an appropriate method, in an adequate and representative sample, with adequate results. Specifications also must be precise and unambiguous to achieve a moderate rating.

2a1. Are the specifications complete and precise?
2a1.10 Is the measure stratified or not?
2a1.13 should not be NA – hierarchical logistic regression? What are the risk factors?
2a1.20 Need more explanation of "smoothed" rate – how are weights determined and what is the reference population – overall average or average by volume category?

2a2. Reliability Testing
1) Is there any requested information missing from the measure submission form that prevents rating reliability testing?
The submission is complete and information presented well.

2) Method, level of testing, and whether it was appropriate and conducted with a representative and adequate sample.
Reliability was tested using signal-to-noise analysis for the performance measure score. This method is appropriate and testing included essentially all hospitals.

3) Results
The average reliability ranged from .00784 for the smallest size decile to 0.56323 for the largest decile.
This analysis results in a reliability estimate for each hospital so the results were presented by decile (10 equal size groups) based on number of cases. Reliability is about the confidence in ability to distinguish performance of one hospital from another. There are three main drivers of reliability: sample size, differences between providers, and measurement error Signal-to-noise reliability = (provider-provider variance)/ (provider-provider variance)+ (provider-specific error variance). . A reliability of one implies that all the variability is attributable to real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error.

******It is important to note that the measure is specified to make up for low reliability through the empirical Bayes shrinkage estimator. Therefore, the final risk-adjusted rates are a more reliable estimate of performance.*** When there is too little information from a hospital to produce a strong

enough signal, information is "borrowed" from the average performance, which is an accepted statistical practice.

Questions –

Is it possible to identify reliability of final scores?

Please explain last column - Percent of signal variance explained by performance score.

VALIDITY

For which NQF rating is the measure eligible?

Testing was conducted for one level – performance score, so the measure is eligible for a moderate rating on validity IF testing is determined to be with appropriate method, in an adequate and representative sample, with adequate results. Adequately addressing all threats to validity (exclusions, risk adjustment, etc.) also is required for a moderate rating.

1.Is there any requested information missing from the measure submission form that prevents rating validity testing, exclusions, risk adjustment, discrimination, comparability of multiple data sources?

The submission is complete and information presented well.

2b2. Validity testing

1) Method, level of testing, and whether it was appropriate and conducted with a representative and adequate sample.

Construct validity to examine the relationship between the performance score and other indicators hypothesized to be related to quality of care is an appropriate method of validity testing.

2) Results

From the regression analysis, the coefficient for volume is negative and statistically significant (higher volume associated with lower adjusted mortality score), coefficient for transfer out is positive and statistically significant (more transfers associated with higher adjusted mortality score, number of dx codes is negative and statistically significant (more codes associated with lower adjusted mortality score).

Testing for construct validity requires understanding and hypothesizing conceptual relationships within the constraints of available data. The results indicate support for the conceptualized/hypothesized relationships.

2b3. Exclusions

1) Method, level of testing, and whether it was appropriate and conducted with a representative and adequate sample.

The frequency of exclusions was not provided – rationale for transfer out because outcome of in-hospital death not relevant once transferred. Stated that MDC 14 and missing data had "non-material number of cases. Analysis of transfer in supports inclusion in risk factors (also addressed in risk model).

2) Results.

2b5. Meaningful differences in performance

1) Method and whether it was appropriate.

Identified for each decile by case volume, the percentage of hospitals identified as better or worse than the reference population. This does provide information on identifying differences in performance.

We ask for at a minimum to provide the distribution of performance scores by quartile or decile, which was not provided (probably due to some confusion with how the question was worded.) This could be provided across all hospitals. The average performance rate for the size deciles could be added in 2b5.2, which also would be informative.

2) Results.
Overall, approximately 13.1 % of hospitals were identified a performing better or worse than the reference population (5.1% of hospitals were identified as performing worse and 8.0% performing better than the reference population).

2b4. Risk adjustment
1) Methods for identifying risk factors and testing the model and whether it was appropriate.
Standard risk factors used in APR-DRGs and other measures are applied and analyzed on statistical criteria ($p<0.5$). The dataset was divided into development and validation samples as considered standard practice.

2) Results
Contribution of risk factors: Odds ratios provided in 2b4.4. All are >1 except age 18-59 and APR-DRG2441-2442.
Discrimination: *c* statistic = 0.831
Calibration: see table and graph in 2b4.7

Description of APR-DRGs was incomplete. What are 242, 243, 244, 253, 254?

Odds ratios compare the odds of the outcome (death) in those with the risk factor and those without the risk factor. Odds ratio = (probability of death with factor present/probability of survival with factor present)/ (probability of death without factor/probability of survival without factor).

Model discrimination represents the extent to which a model predicts higher probabilities of an outcome (e.g., death) for patients who had the outcome (e.g., died) than for those without the outcome (e.g., lived).

In the context of healthcare performance assessment, the purpose of the risk model is to reduce bias due to case mix characteristics present at the start of care (i.e. to risk adjust). The purpose of the model is not to totally explain variation in outcomes, which would require also including variables about quality of care. Variables related to quality of care are purposely not included in risk models for performance measures used to assess quality. From a theoretical standpoint, it is possible for a regression model to exhibit low discrimination and still perform well at reducing bias due to case mix.

The *c* statistic represents the proportion of pairs with different observed outcomes for which the model correctly predicts a higher probability for observations with the event outcome than the probability for nonevent observations. The *c* statistic ranges from 0.5 to 1. A value of 0.5 means that the model is no better than assigning observations randomly into outcome categories (however, that does not necessarily mean it is useless – see above). A value of 1.0 means that the model assigns higher probabilities to all observations with the event outcome compared with nonevent observations. A *c* statistic of 0.831 means that for 83.1% of all possible pairs of patients—one who dies and the other who

lived—the model correctly assigned a higher probability to those who died. The *c* statistic is based only on rank, not the accuracy of the predicted probability.

Model calibration refers to fitting a model to a data set so that the average expected outcome equals the average actual outcome for subgroups at various levels of risk. The table and graph indicate the largest differences between observed and predicted for deciles 5, 6, 8 and 10. The largest difference was .02657 in decile 10. NQF staff are not aware of norms for calibration.

2b6. Comparability of multiple data sources
Not applicable – only one data source.