NATIONAL QUALITY FORUM

TO: Neurology Steering Committee

FR: Suzanne Theberge, Project Manager
Karen Johnson, Senior Director

SU: Neurology Endorsement Maintenance—Post-Comment Call to Discuss Public and Member Comments for Phase I Measures

DA: August 23, 2012

The Neurology Steering Committee will meet via conference call on Monday, August 27. The purpose of this call is to:

- Review and discuss comments received during the public and member comment period.
- Provide input on responses to comments.
- Determine whether reconsideration of any measures or other courses of action is warranted.

Please let us know if you have any questions.

Steering Committee Action:

1. Review this briefing memo
2. Review the comments received and the proposed responses (see Excel and PDF files included with the call materials).
3. Be prepared to provide feedback and input on proposed comment responses.

Please use the following information to access the conference call line and online webinar:

Date/Time: Monday, August 27, 2012, 1:00-3:00 pm ET

Speaker dial-in #: 888-799-5160
Confirmation Code: 78373890

Webinar: http://nqf.commpartners.com/se/Rd/Mt.aspx?470566
All committee and speaker phone lines will be open. Please place your phone on mute when not speaking. Do not put your phone on hold during the call.

NQF received a total of 53 comments on the draft report from public and NQF members. In order to facilitate discussion, many of the comments have been categorized into major themes, although several other comments outside of the major thematic categories also were received and may require discussion by the Committee. Where possible, NQF staff has proposed draft responses for the Committee to consider. Although all comments and proposed responses are subject to discussion, we will not necessarily address each comment and response on the post-comment call. Instead, we will spend the majority of the time considering the major themes and/or those measures with the most significant issues that arose from the comments.

We have included all of the comments that we received in the Excel spreadsheet that is included with the call materials. This comment table contains the commenter's name, as well as the comment, associated measure, theme (if applicable), and draft responses for the Committee's consideration.

MAJOR THEMES

Three major themes were identified in the comments, as follows:

1. Feasibility
2. Harmonization of stroke rehabilitation measures
3. Inclusion of a stroke severity indicator in risk-adjustment models for stroke mortality and readmission measures

Theme 1: Feasibility

Description: We received 13 comments regarding the feasibility of several measures. Most of these comments noted that the measures "may require burdensome electronic health record data extraction or medical chart review."

Specifically, 12 of these comments addressed the following measures:

- 0240: Stroke and Stroke Rehabilitation: Deep Vein Thrombosis (DVT) Prophylaxis for Ischemic Stroke or Intracranial Hemorrhage (AMA-PCPI)
- 0241: Stroke and Stroke (AMA-PCPI)
- Rehabilitation: Anticoagulant Therapy Prescribed for Atrial Fibrillation at Discharge (AMA-PCPI)
- 0243: Stroke and Stroke Rehabilitation: Screening for Dysphagia (AMA-PCPI)
- 0244: Stroke and Stroke Rehabilitation: Rehabilitation Services Ordered (AMA-PCPI)
- 0325: Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy (AMA-PCPI)
- 0434: STK-01: Venous Thromboembolism (VTE) Prophylaxis (TJC)
- 0435: STK 02: Discharged on Antithrombotic Therapy (TJC)
- 0436: STK-03: Anticoagulation Therapy for Atrial Fibrillation/Flutter (TJC)
- 0437: STK 04: Thrombolytic Therapy (TJC)
- 0438: STK 05: Antithrombotic Therapy By End of Hospital Day Two (TJC)
- 0439: STK-06: Discharged on Statin Medication (TJC)

- 0441: STK-10: Assessed for Rehabilitation (TJC)

Also addressed in these comments was a concern that the measures will be difficult to implement from administrative claims because there are limitations in identifying relevant physician behavior in hospital claims. However, the AMA-PCPI measures that were addressed in these comments are clinician-level measures that use CPT-II codes to record the measure focus, and the TJC measures that were addressed are facility-level measures that are not specified for administrative claims.

> ***Proposed Committee Response***: While SC members recognize that the measure may require a fair amount of data abstraction, they agree that the measure meets NQF's feasibility criterion.

A final comment related to measure feasibility concerns measure #1952 (Time to Intravenous Thrombolytic Therapy). Specifically, the comment addresses the difficulty in implementing this measure from administrative claims alone. This measure was mistakenly specified for administrative claims when originally submitted; however, the developer agrees that the measure cannot be captured by claims data only and has revised the submission to specify the measure for electronic registry data only.

Theme 2- Harmonization
Description: We received three comments suggesting that the numerators, denominator exclusions, and timeframe for measures #0244 and #0441 be harmonized.

These measures were identified by NQF staff as related measures because both address rehabilitation services for stroke patients. The measures differ in the following ways:

Number, Title, and Developer	**0244 Stroke and Stroke Rehabilitation: Rehabilitation Services Ordered (AMA-PCPI)**	**0441 STK-10: Assessed for Rehabilitation (The Joint Commission)**
Measure focus	Rehab services ordered OR documentation that no rehab needed	Assessed for or received rehab services
Patient population	Patients 18+, dx=ischemic stroke or intracranial hemorrhage	Patients 18+, dx=ischemic stroke or hemorrhagic stroke

Number, Title, and Developer	0244 Stroke and Stroke Rehabilitation: Rehabilitation Services Ordered (AMA-PCPI)	0441 STK-10: Assessed for Rehabilitation (The Joint Commission)
Denominator exclusions	None	Length of Stay > 120 days, comfort measures only documented, enrolled in clinical trials related to stroke, admitted for elective carotid intervention, discharged to another hospital, left against medical advice, expired, discharged to home for hospice care, discharged to a health care facility for hospice care
Timeframe	At/prior to discharge	Hospital admission to discharge
Level of analysis	Clinician	Facility
Data source	Administrative claims, electronic clinical data, EHR, registry	Electronic clinical data, EHR, paper medical records

In the discussion of these measures during the in-person meeting, the Committee did not identify any harmonization issues to be addressed by the developers.

Developer response (AMA-PCPI): The AMA PCPI uses measure exclusions and measure exceptions, where appropriate. However, as this measure numerator is constructed to capture patients for whom rehabilitation services were ordered and patients for whom the physician has documented that no rehabilitation services were indicated, there is no need for exclusions or exceptions. All patients that receive the appropriate care are captured in the numerator of the measure. Therefore, if rehabilitation services were not ordered and there is no reason documented for the services not being indicated, the physician will not meet the measure.

Developer response (TJC): Thank you for your comment and support of the Joint Commission's performance measure. The stroke core measures were developed in collaboration with the American Heart Association/American Stroke Association in 2003. Since that time, extensive efforts have been made to harmonize the measures, including STK-10: Assessed for Rehabilitation, with the American Medical Association Physician Consortium for Performance Improvement Stroke and Stroke Rehabilitation measures, The American Heart Association/American Stroke Association Get With the Guidelines Patient Management Tool, and the Centers for Disease Control and Prevention Paul Coverdell National Acute Stroke Registry. The Joint Commission will continue to work towards further harmonization with these organizations, and appreciates the feedback for future measure development.

Proposed Committee Response: While the Committee agrees that measures should be harmonized to the extent possible, members recognize that measures

specified for different levels of analysis (e.g., clinician vs. facility) may require different specifications.

Theme 3: Severity of stroke in risk-adjustment models

Description: We received 18 comments expressing the concern that an indicator of stroke severity (particularly, the value of the NIH Stroke Scale) is not included in the risk-adjustment models for stroke mortality and readmissions. These comments on stroke severity pertain to the following measures:

- 0467: Acute Stroke Mortality Rate (IQI 17) (AHRQ)
- 2026: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following an acute ischemic stroke hospitalization (CMS/Yale)
- 2027: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following an acute ischemic stroke hospitalization (CMS/Yale)

Most of these comments specifically cited a recent article by Fonarow[1] and colleagues. The conclusion of this article states:

> *Adding stroke severity assessed with the NIHSS score to a hospital 30-day mortality model based on claims data for Medicare beneficiaries with acute ischemic stroke is associated with substantial improvement in model discrimination and changes in mortality performance ranking for a considerable proportion of hospitals. These findings suggest that it may be critical to collect and include stroke severity for optimal hospital risk adjustment of 30-day mortality for Medicare beneficiaries with acute ischemic stroke.*

Responses from AHRQ: ***(****NOTE: The following text is taken from the responses to comments #2673 and #2707):*

AHRQ acknowledges that optimal risk-adjustment would include clinical markers of stoke severity, such as the NIH Stroke Scale, which may vary across hospitals in association with socioeconomic factors (Kleindorfer D, et al. Stroke 2012;43:2055-9). However, the recent paper by Fonarow et al. is likely to exaggerate the magnitude of this problem, for reasons described fully below. AHRQ will continue to work with the "Get With The Guidelines" team, the VA, and other interested entities that have linked clinical and administrative data to test and improve risk-adjustment modeling. The only currently available data set that links clinical and administrative data on a large population of stroke patients in the US is potentially useful , but it has historically suffered from two disadvantages: (1) it includes only Medicare fee-for-service beneficiaries aged 65 year s or older, and thus underrepresents younger and healthier patients, and patients from states with high managed care penetration; and (2) up until 4/1/12 it has it only had 9 diagnosis fields, as opposed to the 25 or more diagnosis fields available in most data sets used to estimate this AHRQ measure. It will be some time before a year of data will be available

[1] Fonarow, et al. (July 18, 2012). Comparison of 30-day mortality models for profiling hospital performance in acute ischemic stroke with vs without adjustment for stroke severity. *JAMA, 308(3),* 257-264.

with expanded diagnosis codes. AHRQ will continue to collaborate with other interested parties to improve the data and take advantage of recently improved data that are available for testing and validation of risk-adjustment models. Analysis of a relatively small linked data set from the Veterans Health Administration suggests that the NIH Stroke Scale may not have as much impact on risk-standardized mortality rates in the VA setting as among hospitals participating in "Get With The Guidelines" (Keyhani S, et al. Circ Cardiovasc Qual Outcomes 2012; 5:508-13).

AHRQ acknowledges that optimal risk-adjustment would include clinical markers of stoke severity, such as the NIH Stroke Scale. AHRQ has carefully reviewed Fonarow's findings and held two meetings with his team. However, the applicability of their findings to the AHRQ measure is uncertain, because the risk-adjustment model that Fonarow et al. estimated using Medicare administrative data is markedly inferior to AHRQ's model using all-payer administrative data. Specifically, Fonarow et al. initially (J Am Heart Assoc 2012; 1:42-50) reported c statistics for ischemic stroke mortality of 0.71 (95% CI, 0.70-0.72), 0.82 (0.81-0.83), and 0.84 (0.84-0.85) using demographic and comorbidity information from administrative data, the NIH Stroke Scale alone, and both data sources, respectively. In a more recent paper (JAMA 2012; 308:257-64), the same authors built a more robust risk-adjustment model with 87 covariates derived from longitudinal claims data, and reported c statistics of 0.772 (0.769-0.776) and 0.864 (0.861-0.867) for models without and with the NIH Stroke Scale, respectively. By comparison, the AHRQ model, fully stratified for ischemic stroke, has a c statistic of 0.866, which is similar to that of Fonarow et al's combined model and much higher than their model based only on administrative data. The superiority of AHRQ's risk-adjustment model is not due to combining ischemic and hemorrhagic stroke, and it is also not due to adjustment for procedures performed after admission. Re-estimating the AHRQ ischemic stroke model without procedure-related APR DRGs, the c statistic dropped slightly from 0.866 to 0.858, and the weighted hospital-level correlation of adjusted rates between models with and without procedure-related APR DRGs was 0.977. The superiority of AHRQ's risk-adjustment model appears to be attributable to: (1) more complete data, with 25 or more available diagnosis fields instead of 9; (2) inclusion of a wider age spectrum, with adjustment for age; and (3) adjustment for markers of stroke severity that are present on admission and codable in ICD-9-CM, such as coma, other alteration of consciousness, convulsions, and hemiplegia. For example, among patients with ischemic stroke (APR DRG 045), we are able to stratify patients into four risk of mortality categories, with the following numbers of patients and death rates:

Minor (referent) 112,533 0.0038 (0.38%)
Moderate (OR=2.92) 160,536 0.0282 (2.82%)
Major (OR=10.99) 53,457 0.0883 (8.83%)
Extreme (OR=98.15) 23,077 0.3916 (39.2%)

Responses from CMS/YALE: (*NOTE: The following text is taken from a detailed letter from Susannah Bernheim, MD, MHS of Yale, in response to the recent JAMA paper by Fonarow, et al.; the full letter has been made available to the Committee):*

A few important considerations limit the interpretability of the Fonarow paper with reference to our measure.

1. The first concern is the high percent of patients missing National Institutes of Health Stroke Scale (NIHSS) -- over half of the patients in the study do not have a measured NIHSS. The authors provide little information on the potential bias that could be introduced by the missing stroke scales –such as how the degree of missing NIHSS scores relates to median NIHSS for a hospital. If the hospitals with low percentage of completed NIHSS scores also have particularly high NIHSS median scores, this may account for the handful of hospitals whose profile changes with the addition of the score in the model.

2. Secondly, the measure described within the JAMA paper, though described as being modeled after our measure, differs in important respects from ours: 1) the cohort includes hemorrhagic patients, 2) the risk-adjustment includes different variables and is much less parsimonious (including 87 variables in total), and most importantly, 3) the measure does not risk adjust for transfers from Emergency Departments (ED).

   The inclusion of a risk-adjustment variable indicating that a patient transferred into their index admission from an outside ED is important in our measure for two reasons. First, it will helps to account for the increased severity of cases at hospitals which frequently accept such patients (because the ED transfer patients are often more acutely ill). Second, it removes incentives to turn away transferred patients because the severity of such patients is accounted for within the model.

3. Finally, there are a number of issues about the modeling strategy used and the comparisons provided within the JAMA paper that affect our interpretation of the results of the paper.

   a. Unlike our medical record validation which directly compares hospital rates estimated by measures developed using two independent data sources (clinical versus administrative data), the JAMA paper compares a primary administrative model with a second model that includes one additional clinical predictor. Therefore, the administrative primary model is nested within this bigger model. As a mathematical certainty, adding an additional covariate will reduce overall variance at the hospital level and have the effect of pulling some outliers in as seen in the reclassification analysis.

   b. Furthermore, the re-classification analysis provided in the paper is based solely on the hospital random intercept, rather than comparing hospital results based on risk-standardized rates. This approach is, in essence, comparing hospitals' performance on one standard patient (in this case a patient with no comorbid disease). We find interpretation of such results is uncertain because they comparison of intercepts does not capture the full case mix of the hospitals as a risk-standardized rate would.

   c. Finally, and perhaps most importantly, the article does not allow evaluation of the degree of differences between the two models. The reported results in addition to reclassification refer to changes in ranking (based on hospital random intercepts and a

standard patient) rather than actual rate estimates. All estimates have a degree of uncertainty. A small perturbation in the estimates may change ranking without meaningfully changing hospital estimates. The paper does not provide information about how similar the new estimates are to the original estimates, or whether the new estimates fall within the uncertainty of the original estimates. Nor does it present the correlation between the original model results and new results for hospitals.

In summary, although the stated goal of the Fonarow paper is to as assess the additional value of inclusion of stroke severity in our 30-day mortality measure, we find that the model used differs substantially from the measure we have put forward at NQF. The paper does not address critical questions about the impact of missing NIHSS on the majority of patients, nor do the final analyses fundamentally answer the question of whether hospital profiles differ meaningfully with inclusion of the severity score (for all the specific reasons described above).

Proposed Committee Response: **TBD, based on discussion.**

ADDITIONAL DISCUSSION ON COMMENTS/RESPONSES
No other comments require additional discussion unless specifically desired by the Committee.

ADDITIONAL AREAS FOR MEASURE DEVELOPMENT
Several comments included suggestions for additional measure development, as follows:

New Gaps

- An outcome measure that is a combined endpoint of death and severe disability (i.e. Rankin Score 4-6), for a patient centered approach that would incorporate a patient's values on quality of life.
- A measure to document patient and family training and education in acute and post acute settings to reduce disability, burden of care, and primary and secondary prevention.

Suggested edit

- Change "*Measures of* ***post hospital care*** *(prescriptions use at timed intervals after stroke, whether health problems are controlled over time, etc.)"* to "*Measures of* ***post-acute care and rehabilitation care*** *(prescriptions use at timed intervals after stroke, whether health problems are controlled over time, etc.)"* (see page 13 of the project report).

Proposed Committee Response: The Committee agrees with your suggestions for future measure development and the report was updated to include this suggestion.