NATIONAL QUALITY FORUM
Driving measurable health improvements together

Neurology Standing Committee – Spring 2021 Post-Comment Web Meeting

The National Quality Forum (NQF) hosted a web meeting for the Neurology Standing Committee on Wednesday, October 27, 2021, from 10:00 AM – 1:00 PM ET.

Welcome, Review of Meeting Objectives, and Attendance

Tamara Funk, NQF director, welcomed the Standing Committee to the web meeting. Standing Committee Chair Dr. David Tirschwell also provided a welcome message. Erin Buchanan, NQF manager, conducted the Standing Committee roll call. Ms. Funk then provided an overview of the objectives:

- Re-vote on the Consensus Not Reached (CNR) measure (NQF #0507 Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports [American College of Radiology])
- Review the reconsideration request for NQF #3614 Hospitalization After Release With Missed Dizzy Stroke (Johns Hopkins Armstrong Institute of Patient Safety and Quality)
- Review and discuss comments received during the post-evaluation public and member commenting period

Discussion and Re-vote on Consensus Not Reached (CNR) Measure NQF #0507

The first item discussed was NQF #0507 *Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports*, which is stewarded by the American College of Radiology. NQF #0507 assesses stenosis measurement in carotid imaging. Specifically, it assesses the percentage of final reports for carotid imaging studies (e.g., neck magnetic resonance angiography, neck computed tomography angiography, neck duplex ultrasound, and carotid angiograms) that include a direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement.

During the spring 2021 measure evaluation meeting, the Standing Committee did not reach consensus on validity for this measure; therefore, they did not vote on overall suitability for endorsement. The Standing Committee's previous concern was that the developer had not submitted empirical validity testing, which is required at the time of maintenance review. The developer previously attempted both construct and criterion validity testing. However, they were unable to complete either form of testing due to data limitations and a lack of gold standard comparators and instead submitted new face validity testing, which they conducted in November 2020. The face validity testing demonstrated that 82 percent of a Technical Expert Panel agreed or strongly agreed that the measure accurately discriminates good from poor quality.

During the spring 2021 discussion, a Standing Committee member asked whether the developer could audit a sample number of charts outside of the data set and compare it to the measure to assess validity. In response, the developer stated that they were unaware this was a possible way to assess empirical validity. Following that discussion, the developer completed this additional testing and submitted it as part of the public commenting period.

The developer conducted data element validity testing, which involved random audits of data that were submitted to the Qualified Clinical Data Registry (QCDR) of the Centers for Medicare & Medicaid Services' (CMS) Merit-Based Incentive Payment System (MIPS) over a four-year period. Data submitted to the QCDR were compared with a chart review and demonstrated a high level of concordance (98-100 percent) between the exam record data and registry data.

Dr. Tirschwell led the discussion. A Standing Committee member asked how the records were chosen for the audit; in response, the developer clarified that they used a random sample of charts that were submitted to CMS. Another Standing Committee member asked about the proportion of records in the sample that met the measure's criteria, which the developer did not submit. In response, the developer explained that examining the number of records which met the measure versus those that did not meet the measure was not part of the required validity testing. A question was raised regarding whether there was a stratification by severity, which the developer stated was also not part of the measure. The developer explained that the standard way in which radiologists measure carotid stenosis is being measured. The developer also stated that both the overall performance rate and the agreement for the validity testing were high.

Following this discussion, the Standing Committee re-voted on the measure's validity. NQF staff clarified that "high" was not an option because empirical validity testing was not conducted at the measure score level. Ultimately, the measure passed on validity. Following this vote, additional voting took place for overall suitability for endorsement. The Standing Committee voted to recommend the measure for endorsement.

Review of Reconsideration Request for NQF #3614 and Discussion of Public Comments

The Standing Committee's discussion then turned to the second item: a reconsideration request for NQF #3614 *Hospitalization After Release With Missed Dizzy Stroke (H.A.R.M. Dizzy-Stroke)*, which is stewarded by the Johns Hopkins Armstrong Institute for Patient Safety and Quality. NQF #3614 is a new measure that assesses the rate of patients admitted to the hospital for stroke within 30 days of being treated and released from the emergency department (ED) with either a nonspecific, presumed benign, symptom-only dizziness diagnosis or specific inner-ear/vestibular diagnosis. The measure accounts for the epidemiologic base rate of stroke in the population under study using a risk difference approach.

NQF Director Chelsea Lynch mentioned that during the original discussion during the spring 2021 measure evaluation meeting, the Standing Committee expressed concerns about the evidence presented for interventions to improve the measure, the appropriateness of a 30-day time frame, the unintended consequences of overdiagnosis, the sole use of dizziness to capture a missed stroke diagnosis, and whether many EDs would have neurologists available to aid in improving the measure at the bedside.

Ms. Lynch then summarized the reconsideration request from the developer, which consisted of the following concerns: (1) The discussion was fragmented due to inconsistent Standing Committee attendance; (2) The lead developer was not allowed to present the measure to the Standing Committee; and (3) While an overwhelming amount of information was provided that did meet the evidence criteria, the Standing Committee did not fully understand it. Ms. Lynch also reminded the Standing Committee that NQF's preliminary analysis did yield a passing rating for evidence.

At this point in the meeting, quorum had been lost; therefore, the Standing Committee was unable to conduct a live vote on whether to reconsider the measure. Because there was no longer sufficient time

in the measure review cycle to reschedule this meeting, NQF moved to reconsider the measure and asked the Standing Committee to hold a complete discussion of the measure evaluation criteria. Since the meeting did not have a quorum, the Standing Committee would be provided with a recording of the discussion and an offline survey following the meeting in order to record their votes on the measure.

Ms. Buchanan described the four public comments received during the spring 2021 meeting for NQF #3614. One comment came from an NQF member who supported NQF's decision on this measure. A second comment came from a patient (i.e., a member of the public), that was in support of the measure. An additional comment came from another member organization that was in support of the measure concept but did raise concerns about the measure's exclusions, minimal sample size, and lack of risk adjustment. That comment was sent to the developer for a response. The developer addressed each concern in the comment and stated that the NQF Scientific Methods Panel (SMP) had reviewed and passed the measure on scientific acceptability. The final comment consisted of a new presentation of the evidence data from the developer.

Dr. Matt Pickering, NQF senior director, assisted Dr. Tirschwell in leading the Standing Committee through a full discussion of the criteria. Dr. Pickering reminded the Standing Committee that NQF staff gave this measure a passing vote for the evidence criterion in their preliminary analysis of the measure. The evidence suggested that dizziness was commonly misdiagnosed in the ED. In addition, patients hospitalized for stroke are more likely to have had a treat-and-release ED visit for so-called "benign" dizziness within the prior 14 days, and "benign" dizziness treat-and-release discharges from the ED (approximately 30,000 visits per year) are more likely to return for an inpatient stroke admission within the subsequent 30 days.

A Standing Committee member that there needs to be an action that is tied to the outcome. Dr. Pickering then redescribed the criteria for "Importance to Measure and Report," which also includes assessing the performance gap. The developer drew the Standing Committee's attention to a figure in the meeting materials that laid out the relationship between the measure and improvement in patient outcomes. The focus is on improving diagnostic accuracy in stroke, which will improve outcomes for patients in reducing morbidity and mortality. A question was raised about whether the measure was being assessed for either quality improvement (QI) or accountability. It was clarified that this measure would be assessed for both purposes, although NQF criteria should be assessed agnostic to planned use.

The Standing Committee chair questioned the developer as to whether there was an ongoing randomized trial about this measure. The developer reiterated the information that was present in the original submission, linking the use of the measure to improving stroke outcomes. The ongoing Acute Video-oculography for Vertigo in Emergency Rooms for Rapid Triage (AVERT) clinical trial is assessing diagnostic accuracy as an outcome of a care pathway for the evaluation of patients with dizziness. Preliminary results show that experts assessing eye movements do improve diagnostic accuracy, approximately doubling the detection rate. In addition, there is evidence to suggest that the quality of treatment improves with a better diagnosis. In particular, the developer stated that undergoing earlier treatment for minor stroke cuts the risk for major stroke by 34 percent in the next 21 days. The Standing Committee chair stated that this link is somewhat indirect, and a Standing Committee member expressed an additional concern that there is an unclear link between diagnostic accuracy and improving patient outcomes.

Ms. Funk reminded the Standing Committee that quorum had been lost and that voting would occur offline. Dr. Pickering continued to discuss the next criterion and presented the performance gap data. NQF staff's preliminary analysis identified a performance gap, as these data demonstrated a mean score of 17.7 across 967 hospitals, an interquartile range (IQR) of -7.3 to 31.4, and a standard deviation of 30.

In addition, disparities were present in diagnosis according to gender, age, and race. A Standing Committee member commented that differences may have to do with expertise of the clinician and that the measure does show a significant performance gap. In addition, the Standing Committee chair stated that the rate of stroke misdiagnosis is low. The developer clarified that the measure characterizes missed stroke, and there are more patients who have been diagnosed with missed stroke every year than patients who die of breast cancer.

Dr. Pickering then shifted the discussion to reliability. The measure passed the SMP's review of reliability with a moderate rating. A signal-to-noise analysis was conducted, with a median reliability score of 0.59 (IQR = 0.414-0.951). The SMP found this result to be tolerable despite being below the typical 0.7 threshold. The reliability was moderate when hospitals reported more than 250 ED discharges for dizziness, which would apply only to EDs with 40,000 discharges overall per year. Public comments were received that raised concerns regarding the measure's reliability, particularly about its failure to meet the 0.7 threshold; the comments also stated that the developer should add a minimum case count to account for this low result. In response, the developer explained that reliability was limited by the fact that CMS data were incomplete. With complete data, this measure would likely reliably measure hospitals with 10,000-15,000 visits per year, with a measurement window over three years. For lower volume sites, reliability could also be increased by expanding the time window.

For the discussion on validity, the SMP passed this measure on validity with a moderate rating. Dr. Pickering stated that the developer conducted validity testing at the data element level. Data element validity was assessed for two reasons: (1) to test whether stroke diagnoses were valid and (2) to test whether claims were coded as "benign dizziness" by the clinicians were intended to be coded as such. For denominator reliability for benign dizziness diagnoses, the developer conducted two studies focused on code-level validity. First, when an ED patient has a "benign dizziness" discharge diagnosis, what is the frequency of the charts suggesting the ED provider intended to code "benign dizziness"? This study was conducted using two academic hospitals. Positive predictive value (PPV) was calculated in a random sample of 64 charts in three cohorts (i.e., chief complaints of dizziness, otovestibular complaints, and other chief complaints). Negative predictive value (NPV) was calculated specifically if another diagnosis was coded. The developer reviewed a random subsample of 67 charts for high-risk subgroups to estimate NPV. The PPV was 100 percent for coding benign dizziness. The NPV was nearly 100 percent. The audit of discharged status demonstrated 100 percent accuracy, even for the highest risk cases. In addition, the observed rate of stroke in 30 days among cases was compared to an "expected rate" to calculate the measure, the latter being 91–360 days after the visit. The SMP was concerned that the expected rate was based on the assumption that the risk of stroke in 91–360 days is not associated with a misdiagnosis of benign dizziness. The SMP questioned whether this approach to calculate the "expected rate" fully accounts for the risk factors of patients. The lack of risk adjustment for social risk factors was only mentioned in the context of "risk-adjusting away" worse care for racial minorities, with no discussion of potential conceptual relationships. The SMP was also concerned that only a limited sample of hospitals (i.e., four hospitals within Johns Hopkins) were used for testing, which may not generalize well among other hospitals. Lastly, only a very small number of hospitals are extremely poor performers, specifically eight out of 927, which suggests that this is a rare event.

A Standing Committee member noted some additional concerns with validity, particularly that the diagnostic codes may not accurately capture the actual miss rate. In addition, when a low prevalence is present, the positive predictive value is low. Specifically, subtle differences between coding and local clinical practice could confound differences in quality between the hospitals observed in this measure. The developer noted these concerns.

Dr. Pickering then shifted the discussion to feasibility. He explained that the measure was based on data elements available in electronic claims data. The developer also stated that it presents no additional administrative burden. The Standing Committee had no concerns about feasibility.

Following this discussion, the Standing Committee discussed the measure's intended use. Because this is a new measure, the developer presented a plan for use for surveillance, public reporting, and program payment purposes. The developer also discussed various ways in which the measure is currently being used in specific quality improvement (QI) programs. The Standing Committee had no concerns with use.

Regarding usability, the developer reported no unexpected findings (positive or negative) during the relatively recent and small-scale deployment of this measure, including no unintended impacts on patients. In response, the Standing Committee commented that a balancing measure might be needed to measure the overuse of diagnostic imaging. Regarding usability, a concern was raised regarding neurologists being potentially unavailable, particularly neuro-otologists in EDs, to help implement this measure. A Standing Committee member commented that this measure could lead to unintended consequences, particularly with diametrically opposed incentives for emergency physicians to reduce diagnostic imaging and missed diagnoses. The Standing Committee did not have any further discussion on this measure.

Following the discussion, Ms. Funk reiterated the next steps and that the Standing Committee would have 48 hours to vote on the measure offline once the measure recording was available. Ms. Funk also mentioned that there were no related and competing measures for either measure discussed.

On Thursday, October 28, 2021, the meeting recording, transcript, and a voting survey were all shared with the Standing Committee so that they could vote on the criteria discussed during the post-comment meeting. The Standing Committee did not pass the measure on evidence (yes-5; no-10; total votes: 15); therefore, no other votes were recorded, and the measure was not recommended for endorsement.

Public Comment

Ms. Buchanan opened the call to accept public and member comments. No public or member comments were provided during this time.

Activities and Timelines

The Consensus Standards Approval Committee will meet on November 30 – December 1, 2021, to review all Standing Committee decisions. An Appeals period will be held from December 7, 2021 – January 5, 2022.