NATIONAL QUALITY FORUM

RESOURCE USE BONE/JOINT TECHNICAL ADVISORY PANEL MEETING

July 7, 2011

Committee Members Participating: James Weinstein, DO, MS (Chair), Dartmouth Institute for Health Policy; Mary Kay O'Neill, MD, MBA, CIGNA HealthCare; Elizabeth Paxton, MA, Kaiser Permanente; John Ratliff, MD, FACS, Thomas Jefferson University; Catherine Roberts, MD, Mayo Clinic; Craig Rubin, MD, University of Texas Southwestern Medical School; Patricia Sinnott, PT, PhD, MPH, VA health Economics Resource Center, Menlo Park, CA

NQF Staff Participating:; Taroon Amin, MPH, MA, Senior Director; Ashlie Wilbon, MPH, BSN, Senior Project Manager; Lauralei Dorian, Project Manager; Sarah Fanta, Project Analyst; Carlos Alzola, NQF Statistical Consultant; Sally Turbyville, MA, MS, NQF Project Consultant.

Others present via telephone: Dan Dunn, PhD, Ingenix; Todd Lee, PharmD, PhD, American Board of Medical Specialties Research and Education Foundation (ABMS-REF); Lawrence Manheim, PhD, ABMS-REF; Howard Tarko, MD, Ingenix; Cheri Zielinski, Ingenix

MEETING PROCESS

Ms. Wilbon welcomed the Technical Advisory Panel (TAP) and thanked them for their participation. The purpose of the meeting was to discuss Bone/Joint measures submitted by Ingenix and ABMS-REF to the Resource Use Consensus Development Process (CDP).

The measure developers were available via telephone to respond to questions from the Committee as needed. A NQF Member and public comment period occurred at the end of the call; no comments were made at that time. The audio recordings and general project information can be found by clicking on the Resource Use project page.

MEASURE EVALUATION SUMMARY

The following summary includes a review of Bone and Joint measures submitted by Ingenix and the American Board of Medical Specialties Research and Education Foundation. The measure developer gave an overview of the General methods approach and the measure submitted to the project.

1586 Episode of care for acute/subacute lumbar radiculopathy with or without lower back pain
Description: Resource use and costs associated with management of an episode-of-care for acute/sub-acute lumbar radiculopathy with or without lower back pain. The episode is triggered by an initial ambulatory care visit for radiculopathy and lasts for 90 days following the initial visit. All individuals with a radiculopathy diagnosis within six months prior to initial radiculopathy visit are excluded. Measure radiculopathy-related resource use and costs during the three month measurement period following the initial visit, as well as 14 days prior to the initial visit that triggers the episode. Resource Use Type: Per episode Data Type: Administrative claims Resource Use Service Category: Inpatient services: Inpatient facility services, Inpatient services: Admissions/discharges, Ambulatory services: Outpatient facility services, Ambulatory services: Emergency Department, Ambulatory services: Pharmacy, Ambulatory services: Evaluation and management, Ambulatory services: Procedures and surgeries, Ambulatory services: Imaging and diagnostic, Ambulatory services: Lab services, Durable Medical Equipment (DME)

Care Setting: Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinic/Urgent Care, Ambulatory Care : Clinician Office, Emergency Medical Services/Ambulance, Home Health, Hospice, Hospital/Acute Care Facility, Imaging Facility, Laboratory Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility, Post Acute/Long Term Care Facility : Rehabilitation
Level of Analysis: Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinic/Urgent Care, Ambulatory Care : Clinician Office, Hospital/Acute Care Facility, Imaging Facility, Laboratory, Pharmacy
Measure Developer: American Board of Medical Specialties Research and Education Foundation, 222 N. LaSalle St., Suite 1500, Chicago, Illinois, 60601

Importance to Measure and Report:
1a.High Impact: H-7; M-0; L-0; N-0
TAP Discussion: The TAP deemed this measure to be a high cost/high impact area.
1b. Resource use/cost problems: H-2 ; M-4 ; L-1 ; I-0; N-0
TAP Discussion: The measure did not show as much variation in resource use as the TAP expected based on their experience in the field.
1c. Purpose clearly described: H-4 ; M-3 ; L- 0 ; I- 0; N-0
TAP Discussion: The TAP felt that the purpose was sufficiently described.
1d. Resource use service categories consistent and representative: H-3 ; M-4; L-0 ; I-0; N-0
TAP Discussion: The TAP felt that the resource use service categories were appropriate.

Scientific Acceptability of Measure Properties:
2a. Reliability:
2a1.Measure well defined and precisely specified: H-2 ; M-5 ; L-0 ; I-0; N-0
TAP Discussion:
The inclusion of Physical Therapists and Chiropractors concerned the TAP since this inclusion criterion has the potential to include a large number of heterogeneous patients. The TAP agreed that including chiropractic services and physical therapy using the ICD-9 Code without any physical findings, such as confirmatory MRR has the potential to undermine the validity of the model. There was concern that the heterogeneous group of providers and how they code could introduce bias in the measure results. The TAP also questioned the decision to use the 90 day time window since the episode under evaluation may need a longer window.
2a2. The results are repeatable: H-0 ; M-0 ; L-4 ; I-3; N-0
TAP Discussion: The TAP wanted additional information on the reliability testing of the measure. Specifically, they requested additional information on whether independent programmers attempted to write the code and obtained the similar results.
Overall Reliability: H-1 ; M-2 ; L-2 ; I-1; N-0
TAP Discussion:
2b. Validity:
2b1. Evidence is consistent with intent: H-0 ; M-3; L-4 ; I-0; N-0
TAP Discussion:
The TAP was concerned with the 90 day time window for the episode since the condition may need a longer time interval. The Committee requested more information on how surgical and non-surgical patients were evaluated differently in the measure.
2b2.Score/Analysis: H-0 ; M-2 ; L-5 ; I-0; N-0
TAP Discussion:
The TAP agreed that while standardized pricing is helpful for comparison of resource use, it is also important to understand differences in actual prices charged. The TAP was concerned that the measure was only validated in the Market Scan database. The TAP also expressed concern that the measure was tested in Market Scan data which does not include injured workers who would be a large focus of this measure.
2b3. Exclusions: H-0 ; M-1 ; L-1 ; I-5; N-0
TAP Discussion:
The TAP agreed that patient preferences should be allowed to drive exclusions in the measure.
2b4. Risk Adjustment: H-0 ; M-2 ; L-5 ; I-0; N-0
TAP Discussion:
The TAP sought additional details on the risk adjustment model, specifically the goodness-of-fit details and how covariates were chosen for inclusion.
2b5. Identification of statistically significant/meaningful differences: H-0 ; M-0 ; L-7 ; I-0; N-0
TAP Discussion:
The TAP concluded that comparisons should be made of like-patients. The inclusion criteria does not make it clear if differences in resource use are attributed to the provider or the underlying severity of the patient since there is a large degree of heterogeneity in the patient population. Further, the TAP agreed that the O/E ratios would not be immediately actionable by providers or patients
2b6. Multiple data sources:
TAP Discussion: N/A (using all administrative data)
Overall Validity: H-0 ; M-0 ; L-6 ; I-0; N-0

TAP Discussion:

The measure was only evaluated in one database, thus the TAP felt that additional validation/testing of the measure was needed. However, the TAP agreed that the clarity in the measure construction was appealing and encouraged the developer to continue to current efforts.

2c. Stratification for disparities: H-0 ; M-1 ; L-1 ; I-5; N-0

TAP Discussion:

Administrative data used in by this measure is limited in its ability to stratify based on race.

Usability:

3a. Measure performance results are publicly reported: H-0 ; M-1 ; L-0 ; I-6; N-0

TAP Discussion:

The measure isn't currently being reported but agreed that the ability to understand the measure and its construction was appealing.

3b. Measure results are meaningful/useful for public reporting and quality improvement: H-0 ; M-2 ; L-0 ; I-5; N-0

TAP Discussion:

While the O/E result is informative, it is difficult to assess whether it would provide meaningful and actionable results for the accountable entity.

3c. Data and results can be decomposed for transparency and understanding: H-2 ; M-5 ; L-0 ; I-0; N-0

TAP Discussion:

The TAP agreed that the clinical construction logic was clear and could be deconstructed to facilitate transparency and understanding.

3d. Harmonized or justification for differences:

TAP Discussion: N/A

4. Feasibility:

4a. Data elements routinely generated during care process: H-7 ; M-0 ; L-0 ; I-0; N-0

TAP Discussion: This measure would be rely on data elements that are routinely generated during the care process.

4b. Data elements available electronically: H-7 ; M-0 ; L-0 ; I-0; N-0

TAP Discussion: This measure would be rely on data elements that are available electronically.

4c. Susceptibility to inaccuracies/ unintended consequences identified: H-2 ; M-3 ; L-2 ; I-0; N-0

TAP Discussion:

The TAP felt that susceptibility to inaccuracies may not have been sufficiently addressed in the submission and there is potential for bias in this measure since there is a large amount of heterogeneity in the providers and thus the types of patients that would be included in the sample.

4d. Data collection strategy can be implemented: H-0 ; M-5 ; L-2 ; I-0; N-0

TAP Discussion:

1609 ETG Based hip/knee replacement Resource Use Measure

Description: The measure focuses on resources used to deliver episodes of care for patients who have undergone a Hip/Knee Replacement. Hip Replacement and Knee Replacement episodes are initially defined using the Episode Treatment Groups (ETG) methodology and describe the unique presence of the condition for a patient and the services involved in diagnosing, managing and treating the condition. The Procedure Episode Group (PEG) methodology uses the ETG results and further logic to creating a procedure episode that focuses on the Hip Replacement and Knee Replacement component of the care. Procedure episodes identify a unique procedure event as well as the related services performed before and after the procedure including workup and therapy prior to the procedure as well as post-op activities such as repeated surgery and patient follow-up. Together, the ETG and PEG methodologies identify the services involved in diagnosing, managing and treating patients with Hip/Knee Replacements. A methodology to assign a severity level to each episode is employed to group Hip and Knee Replacement episodes by level of risk.

Resource Use Type: Per episode

Data Type: Administrative claims

Resource Use Service Category: Inpatient services: Inpatient facility services, Inpatient services: Admissions/discharges, Ambulatory services: Outpatient facility services, Ambulatory services: Emergency Department, Ambulatory services: Pharmacy, Ambulatory services: Evaluation and management, Ambulatory services: Procedures and surgeries, Ambulatory services: Imaging and diagnostic, Ambulatory services: Lab services

Care Setting: Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinic/Urgent Care, Ambulatory Care : Clinician Office, Emergency Medical Services/Ambulance, Home Health, Hospice, Hospital/Acute Care Facility, Imaging Facility, Laboratory Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility, Post Acute/Long Term Care Facility : Rehabilitation

Level of Analysis: Clinician : Group/Practice, Clinician : Individual, Clinician : Team, Facility, Health Plan, Integrated Delivery System, Population : Community, Population : County or City, Population : National, Population : Regional, Population : State

Measure Developer: Ingenix, 950 Winter Street, Suite 3800, Waltham, Massachusetts, 02451

Conditions/Questions for Developer:
Developer Response:

Importance to Measure and Report –
1a.High Impact: H-6; M-1; L-0; N-0
TAP Discussion: The TAP deemed this measure to be a high cost/high impact area
1b. Resource use/cost problems: H-0 ; M-2 ; L-5 ; I-0; N-0
TAP Discussion: The TAP felt that the measure would be able to identify large variation in resource use and cost. However, the TAP felt that the developers could have provided more information specifically related to hip/knee replacement variation in resource use in the measure submission.
1c. Purpose clearly described: H-0 ; M-5 ; L-1 ; I-1; N-0
TAP Discussion: The TAP felt that the purpose was sufficiently described.
1d. Resource use service categories consistent and representative: H-2 ; M-5; L-0 ; I-1; N-0
TAP Discussion: The TAP felt that the resource use service categories were appropriate.

Scientific Acceptability of Measure Properties:
2a. Reliability:
2a1.Measure well defined and precisely specified: H-0 ; M-3 ; L-4; I-0; N-0
TAP Discussion: The TAP wanted more information on how the developers handled right and left hip/knee replacement since there is limited ability to distinguish between right/left surgery in the administrative data used. The developer should consider including patient preferences in the measure since this procedure is largely driven by patient preferences. It is important to capture the rate of surgery at the provider level to ensure that the current measure construct does not penalize those providers who chose conservative treatment for low severity patients. The developer should provide more clear information on the clinical logic, including the specific codes that are used to create the episodes. Overall, the TAP wanted more clarity on the clinical construction logic of the episode such as severity level assignments, assignment of claims with two concurrent episodes (i.e. tie breaking logic). The TAP also wanted more information on the procedure definitions, handling of comorbidities and the weighting of multiple co-occurring comorbidities.
2a2. The results are repeatable: H-2 ; M-5 ; L-0 ; I-0; N-0
TAP Discussion: The TAP wanted additional information on how reliable the physician level scores were over time.
Overall Reliability: H-2; M-4 ; L-0 ; I-0; N-0
TAP Discussion:
2b. Validity:
2b1. Evidence is consistent with intent: H-2 ; M-4 ; L-1 ; I-0; N-0
TAP Discussion: The TAP felt that the evidence was consistent with the intent of the measure.
2b2.Score/Analysis: H-1 ; M-4 ; L-2 ; I-0; N-0
TAP Discussion: The TAP discussed the attribution of costs six months before the procedure as too long of a period for a physician based measure. With the current attribution method, it appears to be more appropriate at a plan or system-level rather than a single provider.
2b3. Exclusions: H-0 ; M-2 ; L-4 ; I-1; N-0
TAP Discussion: The TAP wanted more information on why low cost outliers were excluded but high cost outliers were windsorized; a sensitivity analysis of this decision was recommended by the TAP. The TAP also recommended that the measure should include a count of high cost outliers if they are going to be windsorized. Information about the high cost outliers might actually drive targeted interventions.
2b4. Risk Adjustment: H-0 ; M-0 ; L-6; I-1; N-0
TAP Discussion: The TAP wanted more information on severity levels on how they related to the risk adjustment model. The TAP agreed that not all of the comorbidities provided in the submission seem appropriate for the population in the measure.
2b5. Identification of statistically significant/meaningful differences:
TAP Discussion: There was general agreement that the complexities of the score may make it difficult to discern meaningful differences between providers.
2b6. Multiple data sources:
TAP Discussion: N/A (using all administrative data)
Overall Validity: H-0 ; M-1 ; L-5; I-0; N-0
TAP Discussion:
2c. Stratification for disparities: H-1 ; M-0 ; L-4; I-2; N-0
TAP Discussion: Administrative data is limited in its ability to stratify based on race.

Usability:
3a. Measure performance results are publicly reported: H-0 ; M-5 ; L-2 ; I-0; N-0
TAP Discussion: The TAP was concerned that this ETG was not currently being used as a stand-alone measure.

3b. Measure results are meaningful/useful for public reporting and quality improvement: H-0 ; M-4 ; L-3 ; I-0; N-0
TAP Discussion: The TAP was concerned that this ETG was not currently being used as a stand-alone measure.
3c. Data and results can be decomposed for transparency and understanding: H-0 ; M-3 ; L-4; I-0; N-0
TAP Discussion: The TAP expressed concern over the difficulty in understanding the clinical hierarchy and risk model. The lack of clarity in these aspects of the measure makes it difficult to deconstruct the measure for transparency and understanding.
3d. Harmonized or justification for differences:
TAP Discussion: N/A

4. Feasibility:
4a. Data elements routinely generated during care process: H-5; M-2 ; L-0 ; I-0; N-0
TAP Discussion: This measure would be rely on data elements that are routinely generated during the care process.
4b. Data elements available electronically: H-6 ; M-1 ; L-0 ; I-0; N-0
TAP Discussion: This measure would be rely on data elements that are available electronically.
4c. Susceptibility to inaccuracies/ unintended consequences identified: H-0 ; M-3 ; L-4 ; I-0; N-0
TAP Discussion: The TAP agreed that much of this surgery is dependent on patient preferences thus the measure should account for these preferences in inclusion and exclusion in the measure. Additionally, providers who treat their patients conservatively can look like high cost users since the only patients who get surgery are those who are more severe.
4d. Data collection strategy can be implemented: H-1 ; M-5 ; L-1 ; I-0; N-0
TAP Discussion:

1585 Episode of care for simple, non-specific lower back pain (acute and subacute)

Description: Resource use and costs associated with management of an episode-of-care for simple nonspecific lower back pain. The episode is triggered by an initial ambulatory care visit for non-specific lower back pain (LBP). The episode lasts three months (90 days) from the time of the trigger ambulatory visit. An episode only begins if there are no LBP ambulatory care visits within 90 days prior to the initial LBP visit. Also, all individuals with a radiculopathy diagnosis during the measurement or prior period are excluded. LBP-related resource use and costs are measured during the episode, including 14 days prior to the initial visit that triggers the episode.
Resource Use Type: Per episode
Data Type: Administrative claims
Resource Use Service Category: Inpatient services: Inpatient facility services, Inpatient services: Admissions/discharges, Ambulatory services: Outpatient facility services, Ambulatory services: Emergency Department, Ambulatory services: Pharmacy, Ambulatory services: Evaluation and management, Ambulatory services: Procedures and surgeries, Ambulatory services: Imaging and diagnostic, Ambulatory services: Lab services, Durable Medical Equipment (DME)
Care Setting: Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinic/Urgent Care, Ambulatory Care : Clinician Office, Hospital/Acute Care Facility, Imaging Facility, Laboratory, Pharmacy
Level of Analysis: Clinician : Individual
Measure Developer: American Board of Medical Specialties Research and Education Foundation, 222 N. LaSalle St., Suite 1500, Chicago, Illinois, 60601

Conditions/Questions for Developer:
Developer Response:

Importance to Measure and Report –
1a.High Impact: H-7 ; M-0 ; L-0 ; I-0; N-0
TAP Discussion: The TAP deemed this measure to be a high cost/high impact area.
1b. Resource use/cost problems: H-7 ; M-0 ; L-0 ; I-0; N-0
TAP Discussion: The TAP agreed that the measure demonstrated a notable resource use/cost problem.
1c. Purpose clearly described: H-5 ; M-2 ; L-0 ; I-0; N-0
TAP Discussion: The TAP agreed that the purpose was sufficiently described.
1d. Resource use service categories consistent and representative: H-2 ; M-5 ; L-0 ; I-0; N-0
TAP Discussion: The TAP concluded that the resource use service categories were appropriate.

Scientific Acceptability of Measure Properties:
2a. Reliability:
2a1.Measure well defined and precisely specified: H-0 ; M-7 ; L-0 ; I-0; N-0
TAP Discussion: Considering the intent of the measure, the TAP agreed that the group of diagnoses chosen were appropriate for the measure.
2a2. The results are repeatable: H-0 ; M-1 ; L-3; I-3; N-0
TAP Discussion: The TAP wanted additional information on the reliability testing of the measure. There was no information provided if two independent programmers received the same results after running the product. More robust methods for reliability should be considered by the developer.

Overall Reliability: H-0 ; M-1; L-3 ; I-2; N-0
TAP Discussion:
2b. Validity:
2b1. Evidence is consistent with intent: H-1 ; M-6 ; L-0 ; I-0; N-0
TAP Discussion: The Committee felt that the evidence was consistent with the intent of the measure.
2b2.Score/Analysis: H-0 ; M-1 ; L-4 ; I-2; N-0
TAP Discussion: The TAP continued to be concerned about the interpretability of the measure score. The group was concerned that O/E ratios would not be actionable by accountable entities.
2b3. Exclusions: H-0 ; M-4 ; L-3 ; I-0; N-0
TAP Discussion: The TAP agreed that not all of the exclusions in this measure submission have a clinical basis for exclusion (examples include active cancer patients have been excluded while prostate, melanoma, skin cancer patients are included). The exclusion of compression fractures also requires additional explanation.
2b4. Risk Adjustment: H-0 ; M-3 ; L-4 ; I-0; N-0
TAP Discussion: The TAP sought additional details on the performance of the risk adjustment model include information the goodness-of-fit. Details were provided on the inclusion of covariates however clinical justification on why they were chosen needs more explanation. The TAP expressed concern over the use of the hierarchical condition categories (HCC) on a non-Medicare population
2b5. Identification of statistically significant/meaningful differences: H-0 ; M-3 ; L-3 ; I-1; N-0
TAP Discussion: The TAP agreed that O/E ratios may not be immediately actionable by providers or patients. Since the condition is a symptom condition it would be difficult to ascertain what O/E variation would imply. These limitations reduce the ability of the measure to identify statistically significant and meaningful differences in performance.
2b6. Multiple data sources:
TAP Discussion: N/A (using all administrative data)
Overall Validity: H-0 ; M-3 ; L-2 ; I-1; N-0
TAP Discussion: The TAP agreed that additional validation/testing of the measure was needed. However, the TAP applauded the effort thus far since the philosophical measure construction was appealing and encouraged the developer to continue to current efforts.
2c. Stratification for disparities: H-1 ; M-1 ; L-2 ; I-3
TAP Discussion: The underlying administrative data is limited in its ability to stratify based on race.

Usability:
3a. Measure performance results are publicly reported: H-0 ; M-3 ; L-3 ; I-1; N-0
TAP Discussion: The measure isn't currently being reported.
3b. Measure results are meaningful/useful for public reporting and quality improvement: H-0 ; M-1 ; L-4; I-2; N-0
TAP Discussion: While the O/E result is informative, it is difficult to assess whether it would provide meaningful results.
3c. Data and results can be decomposed for transparency and understanding: H-1 ; M-3 ; L-2 ; I-0; N-0
TAP Discussion: The clinical construction logic was clear however more detailed information on the clinical logic was requested specifically exclusions.
3d. Harmonized or justification for differences:
TAP Discussion: N/A

4. Feasibility:
4a. Data elements routinely generated during care process: H-5; M-1; L-0; I-0; N-0
TAP Discussion: This measure would be rely on data elements that are routinely generated during the care process.
4b. Data elements available electronically: H-6 ; M-0 ; L-0 ; I-0; N-0
TAP Discussion: This measure would be rely on data elements that are available electronically.
4c. Susceptibility to inaccuracies/ unintended consequences identified: H- 0; M-1 ; L-4 ; I-1; N-0
TAP Discussion: The TAP agreed that for this particular diagnosis group there are very different coding practices among providers thus there are susceptibility to inaccuracies in both the coding and classification of patients in the measure.
4d. Data collection strategy can be implemented: H-0 ; M-1 ; L-2 ; I-3; N-0
TAP Discussion:

PUBLIC COMMENT

Mr. Martin from the American Academy of Orthopedic Surgeons commented that he was supportive of the work of the TAP and thanked the group for their time on this effort.

NEXT STEPS

The TAP agreed the clinical course and treatment of hip and pelvic fractures are vastly different and the resources/costs of these episodes should not be combined in one measure as submitted in measure #1603 from Ingenix (ETG-based hip/pelvic fracture resource use measure). As such, the TAP decided not to review this measure during the meeting and to defer discussion until the hip and pelvic codes can be separate into separate episodes. NQF Staff will follow up with the Ingenix development team regarding the resubmission of individual episodes for hip and pelvic fracture resource use for review by the TAP at a future date. Staff will also follow up with the TAP to schedule an additional conference call to discuss the new measures for the separate episodes once they have been re-submitted.