MEETING OF THE RESOURCE USE STEERING COMMITTEE

August 29-30, 2011

Committee Members Participating: Bruce Steinwald, MBA (Co-Chair), Independent Consultant; Tom Rosenthal, MD, UCLA School of Medicine (Co-Chair); Paul Barnett, PhD, VA Palo Alto Health Care System; Jack Bowhan, Wisconsin Collaborative for Healthcare Quality; Jeptha Curtis, MD, Yale University School of Medicine; Lisa Grabert, MPH, American Hospital Association; Ann Hendrich, RN, MSN, Ascension Health; Jack Needleman, PhD, FAAN, University of California, Los Angeles School of Public Health; Mary Kay O'Neill, MD, MBA, CIGNA HealthCare; David Penson, MD, MPH, Vanderbilt University Medical Center; Doris Peter, PhD, Consumers Union; Steve Phillips, MPA, Ortho-McNeill-Janssen Pharmaceutical, Inc.; David Redfearn, PhD, WellPoint; Jeffrey Rich, MD, Mid-Atlantic Cardiothoracic Surgeons Ltd.; William Rich, MD, Northern Virginia Ophthalmology Associates; Barbara Rudolph, PhD, MSW, The Leapfrog Group; Joseph Stephansky, PhD, Michigan Health and Hospital Association; Dolores Yanagihara, MPH, Integrated Healthcare Association.

NQF Staff Participating: Helen Burstin, MD, MPH, Senior Vice President; Karen Pace, MSN, PhD, Senior Director; Taroon Amin, MPH, MA, Senior Director; Ashlie Wilbon, RN, MPH, Senior Project Manager; Lauralei Dorian, Project Manager; Sarah Fanta, Project Analyst; Sally Turbyville, MA, MS, NQF Consultant, Carlos Alzola, NQF Statistical Consultant

Others Present: Ben Hamlin, NCQA; Chad Heim, HealthPartners; Sue Knudson, HealthPartners; Todd Lee, ABMS-REF (Via Phone); Kevin Stroupe, ABMS-REF (Via Phone); Kevin Weiss, ABMS-REF (Via Phone) Arjun Venkatesh, Brigham and Women's Hospital

MEETING PROCESS

The purpose of this two-day meeting was to evaluate the measures reviewed by the Cardiovascular/Diabetes, Pulmonary, and Bone/Joint Technical Advisory Panels (TAPs) to date. The first day the committee considered and discussed the TAP's measure evaluations and made final recommendations for endorsement. The second day of the meeting was spent reviewing the process and techniques the committee used to evaluate the resource use measures. The committee also analyzed the "lessons learned" from this first phase of the project and reviewed the resource use measure evaluation criteria in order to develop guidance for measure developers who may submit resource measures to NQF in the future.

The measure developers were available in person and via phone to respond to questions from the Committee as needed. NQF Members and public were given the opportunity to comment period throughout the two-day meeting; no comments were made at that time. General project information can be found by clicking on the Resource Use project page.

COSTING APPROACH DISCUSSION

The following summary includes a review of the Steering Committee discussion of the various costing approaches used in the measures that were submitted to this project.

Each of the developers who submitted measures to the Resource Use project were asked to specify a costing method to apply to the measure. For the measures submitted, the costing approaches were either specified for the actual amount paid (i.e., cost of care measures) or for standardized prices (i.e., resource use measure). Standardized pricing allows users to compare the use and intensity of health services while holding actual paid amounts constant. Resource use measures that apply standardized prices allow for comparison of resource use units across regions and markets, while actual prices allow for comparison of prices paid. The Committee agreed that both approaches could be appropriate for different applications; however a measure used as a national consensus standard must select a single costing approach. Including both costing approaches within the same measure could reduce comparability and limit the user's ability to identify the source of variation. For this reason, developers that submitted a single measure with an option for the user to determine which costing method to apply were asked either to split the submission into two separate measures or select one of the approaches to apply to a single measure submission. At the Committee's request, measures submitted by Ingenix that were unknowingly evaluated and voted on with optional costing approaches were re-voted during the Cycle 2 Committee meeting based on the developer's selection of a single costing approach to be applied (actual prices paid) to all of their measures.

Committee discussions on applying an actual price approach for national comparisons at an individual provider level identified additional concerns. Specifically, the Committee noted the potential for misinterpreting physician resource use in national reporting. This pricing approach includes environmental factors (i.e., local facility and labor costs) that may be outside of an individual provider's control. The Committee agreed that when actual prices paid are reported, utilization counts should be reported as well. The concern over the use of actual prices also was considered in the measure's usability. However, there was agreement that actual prices paid by health plans to providers is important to measure and report; for example, regional comparisons at the individual provider level where environmental factors may not be as prominent, or nationally at higher levels of measurement (i.e. health plan level). Measures based on actual prices paid are encouraged for endorsement, noting that the validity will be examined through the interaction of measure components including its specified level of analysis and risk adjustment model.

MEASURE EVALUATION SUMMARY

The following summaries include a review of the Steering Committee discussion of each measure. The Cardiovascular/Diabetes, Pulmonary and Bone/Joint TAP Co-Chairs were available to give a summary of the TAP evaluations of each measure. The measure developers were also available to give a brief overview of the overall approach and specific measures.

1611: ETG Based Pneumonia Cost of Care Measure (Ingenix)

Description: The measure focuses on resources used to deliver episodes of care for patients with pneumonia.
Pneumonia episodes are 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 pneumonia. A number of resource use measures are defined for pneumonia episodes, including overall cost of care, cost of care by type of service, and the utilization of specific types of services. Each resource use measure is expressed as a cost or a utilization count per episode and comparisons with internal and external benchmarks are made using risk adjustment to support valid comparisons. As requested by NQF, the focus of this submission is for pneumonia episodes and will cover both measures at the pneumonia base and severity level and also a pneumonia composite measure where pneumonia episode results are combined across pneumonia severity levels. At the most detailed level, the measure is defined as the base condition of pneumonia and an assigned level of severity (e.g., resources per episode for pneumonia, severity level 1 episodes). Composite measures can then be created using these measurement units to meet a specific need. For example, a composite measure for pneumonia is derived by combining pneumonia episode results across pneumonia severity levels. Appropriate risk adjustment is applied to support comparisons (e.g., for physician measurement, adjusting for a physician's mix of pneumonia episodes by severity level when supporting a pneumonia composite comparison). The focus of this measure is on pneumonia. However, pneumonia episode results could also be included in a "pulmonary" or other clinical composite for a physician, combining episodes in clinical areas similar to pneumonia. Further, an "overall" composite for a physician can be created, again by aggregating episode results across appropriate conditions and severity levels and applying proper risk adjustment when making comparisons.
Resource Use Type: Per episode
Data Type: Administrative claims, Other
Resource Use Service Categories:
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

Committee Recommendation for Endorsement: Y-12; N-4; Abstain-0

Conditions/Questions for Developer:

1. Would it be possible to break down the measure by bacterial versus non-bacterial to try to separate out pneumonia types?

Developer Response:

1. Yes, the measure is stratified. To the extent that administrative claims code the differences in pneumonia types, the measure can be stratified to evaluate resource use differences between pneumonia types.

Overall Importance: Y-14, N-1
Committee Discussion: The Steering Committee deemed the measure to be important.

Overall Scientifically Acceptable: Yes [Y-13; N-3 (Committee Vote)]
Overall Reliability: H-3; M-11; L-2; I-0
Overall Validity: H-1; M-13; L-2; I-0
Committee Discussion: The Steering Committee was concerned that the measure may not be clinically relevant due to its limited ability to differentiate between community and hospital acquired pneumonia. In general, the Committee also believed that the "start and stop rules" would be more readily apparent for acute procedure-oriented measures such as knee replacements, as compared with chronic illnesses, which has less clear cut start and stop dates.

Overall Usability: H-3; M-11; L-1; I-1
Committee Discussion: There were no additional concerns identified by the Steering Committee for this criterion.

Overall Feasibility: H-1; M-8; L-7; I-0
Committee Discussion: All of the data is generated as a byproduct of care and is available electronically. There was concern around the measure's susceptibility to inaccuracies as Ingenix does not have a formal audit system to ensure that all of data is included and correct. Annually, for physician's the cost of the small package (less than 800 physicians) is $70,000, the medium package is $90,000 and the large package is $110,000 (over 2,000 physicians in the group). For health plans the package comes in small, medium and large; the small package is $90,115 (less than 400,000 covered lives) and the large package is $135,000 (over a million covered lives), all prices are annual rates. The Steering Committee came to the conclusion that this is no more expensive than other risk adjustment models another product of its caliber (for example, ACGs used by HealthPartners). However, the cost of the Ingenix product includes costs associated with the licensure of the proprietary software and the cost of all of their measures, over 558 ETGs. The Steering

Committee acknowledges that the system may be used without Ingenix's technical support, it is complicated, but if the documentation is thoroughly read, it is doable.

1609: ETG/PEG Based hip/knee replacement Cost of Care Measure (Ingenix)

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

Committee Recommendation for Endorsement: Y-9; N-7; Abstain-0

Conditions/Questions for Developer: N/A
Developer Response: N/A

Overall Importance: Y-17, N-0
Committee Discussion: The Steering Committee deemed this measure to be important.

Overall Scientifically Acceptable: Yes [Y-11; N-5 (Committee Vote)]
Overall Reliability: H-2; M-14; L-0; I-0
Overall Validity: H-1; M-9; L-6; I-0
Committee Discussion: The Steering Committee was concerned with the lack of specification regarding the measure's use of MSDRG's in the risk-adjustment methodology. Ingenix explained that among the population of patients who undergo knee or hip replacements, there is minimal variation in the underlying co-morbidities. Therefore, the methodology required to adequately risk adjust is much less stringent than it would be if looking at a more complicated condition such as coronary artery disease.

Overall Usability: H-0; M-12; L-4; I-1
Committee Discussion: The Steering Committee iterated their concern that, because the measure is used as part of a grouper, it is unclear if it is useful as a standalone measure. Additionally, based on the nature of the Ingenix product, hip and knee replacements had been combined into a single measure, which was not believed by some to be the most clinically relevant approach.

Overall Feasibility: H-1; M-8; L-7; I-0
Committee Discussion: All of the data is generated as a byproduct of care and is available electronically. There was concern around the measure's susceptibility to inaccuracies as Ingenix does not have a formal audit system to ensure that all of data is included and correct. Annually, for physician's the cost of the small package (less than 800 physicians) is $70,000, the medium package is $90,000 and the large package is $110,000 (over 2,000 physicians in the group). For health plans the package comes in small, medium and large; the small package is $90,115 (less than 400,000 covered lives) and the large package is $135,000 (over a million covered lives), all prices are annual rates. The Steering Committee came to the conclusion that this is no more expensive than other risk adjustment models another product of its caliber (for example, ACGs used by HealthPartners). However, the cost of the Ingenix product includes costs associated with the licensure of the proprietary software and the cost of all of their measures, over 558 ETGs. The Steering Committee acknowledges that the system may be used without Ingenix's technical support, it is complicated, but if the documentation is thoroughly read, it is doable.

1603: ETG/ PEG Based Hip Fracture Cost of Care measure (Ingenix)

Description: The measure focuses on resources used to deliver episodes of care for patients with Hip Fracture. Hip Fracture episodes are 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 Hip Fracture. A number of resource use measures are defined for Hip Fracture episodes, including overall cost of care, cost of care by type of service, and the utilization of specific types of services. Each resource use measure is expressed as a cost or a utilization count per episode and comparisons with internal and external benchmarks are made using risk adjustment to support valid comparisons. As requested by NQF, the focus of this submission is for Hip Fracture episodes and will cover both measures at the Hip Fracture base and severity level and also a Hip Fracture composite measure where Hip Fracture episode results are combined across Hip Fracture severity levels. At the most detailed level, the measure is defined as the base condition of Hip Fracture and an assigned level of severity (e.g., resources per episode for Hip Fracture, severity level 1 episodes). Composite measures can then be created using these measurement units to meet a specific need. For example, a composite measure for Hip Fracture is derived by combining Hip Fracture episode results across Hip Fracture severity levels. Appropriate risk adjustment is applied to support comparisons (e.g., for physician measurement, adjusting for a physician's mix of Hip Fracture episodes by severity level when supporting a Hip Fracture composite comparison).The focus of this measure is on Hip Fracture. However, Hip Fracture episode results could also be included in an "orthopedics", "acute care", or other clinical composite for a physician, combining episodes in clinical areas similar to Hip Fracture. Further, an "overall" composite for a physician can be created, again by aggregating episode results across appropriate conditions and severity levels and applying proper risk adjustment when making comparisons.
Resource Use Type: Per episode
Data Type: Administrative claims, Other
Resource Use Service Categories:
Inpatient services: Inpatient facility services; Admissions/discharged; Ambulatory services: Outpatient facility services; Emergency Department; Pharmacy; Evaluation and management; Procedures and surgeries; Imaging and diagnostic; 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

Committee Recommendation for Endorsement: This measure did not pass the scientific acceptability criterion, and is not recommended for endorsement.

Conditions/Questions for Developer:

1. Why are different age groups assigned the same risk coefficients, when they will have extremely different risk factors?
2. How does the episode grouper work in terms of low and high outliers? Are you able to provide information on exactly how many episodes have been excluded?
3. Why do you cut the low cost episodes from being included in the measure?

Developer Response:

1. This represents a limitation of the data set. Due to the minimal number of people over 65 in commercial programs, we didn't have the numbers to further stratify.
2. Low cost cases are excluded.
3. The hypothesis that that these low cost episodes – ones under 2.5 percent of the average – are either mistakes or miscodes. They are probably incomplete episodes, therefore they are not counted.

Overall Importance: Y-10, N-6
Committee Discussion: The Committee agreed that hip fractures are a high impact area of healthcare. They were concerned, however, that the measure did not include populations of patients over 65, where the vast majority of hip fractures would occur, and where the nature of hip fractures is a significantly different than it is for younger populations. Ingenix reminded the Committee that the measure was tested in a commercial database, not a Medicare database, and would therefore be endorsed as such. The Committee ultimately questioned whether it was important to measure hip fractures in a younger population at all.

Overall Scientifically Acceptable: No [Y-7; N-10 (Committee Vote)]
Overall Reliability: H-1; M-11; L-3; I-2
Overall Validity: H-0; M-6; L-10; I-0
Committee Discussion: The Committee believed the measure was limited in its clinical construction logic as a result of its reliance upon commercial data, where the population of patients with hip fractures was notably low. Thus, the testing completed by Ingenix for this measure represented a fairly *uncommon* condition – hip fractures in under 65's – when the majority of hip fractures are much more common and different clinically. The Committee agreed, therefore, that significant and meaningful differences could not be produced by

this measure, particularly when reporting at an individual physician level. Furthermore, the Committee were concerned with the fact that the developer did not provide sufficient reliability and validity testing, and Ingenix provided no information comparing scoring of attribution over episodes of time

Overall Usability: This measure did not pass the scientific acceptability criterion. As a result, the Committee did not discuss usability.

Overall Feasibility: This measure did not pass the scientific acceptability criterion. As a result, the Committee did not discuss feasibility.

1605: ETG Based Asthma Cost of Care Measure(Ingenix)

Description: The measure focuses on resources used to deliver episodes of care for patients with Asthma.
Asthma episodes are 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 asthma. A number of resource use measures are defined for asthma episodes, including overall cost of care, cost of care by type of service, and the utilization of specific types of services. Each resource use measure is expressed as a cost or a utilization count per episode and comparisons with internal and external benchmarks are made using risk adjustment to support valid comparisons.
As requested by NQF, the focus of this submission is for Asthma episodes and will cover both measures at the Asthma base and severity level and also an Asthma composite measure where Asthma episode results are combined across Asthma severity levels. At the most detailed level, the measure is defined as the base condition of Asthma and an assigned level of severity (e.g., resources per episode for Asthma, severity level 1 episodes). Composite measures can then be created using these measurement units to meet a specific need. For example, a composite measure for Asthma is derived by combining Asthma episode results across Asthma severity levels. Appropriate risk adjustment is applied to support comparisons (e.g., for physician measurement, adjusting for a physician's mix of Asthma episodes by severity level when supporting an Asthma composite comparison). The focus of this measure is on Asthma. However, Asthma episode results could also be included in a "pulmonologist", "chronic care", or other clinical composite for a physician, combining episodes in clinical areas similar to Asthma. Further, an "overall" composite for a physician can be created, again by aggregating episode results across appropriate conditions and severity levels and applying proper risk adjustment when making comparisons.
Resource Use Type: Per episode
Data Type: Administrative claims, Other
Resource Use Service Categories:
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
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

Committee Recommendation for Endorsement: Y-7; N-9; Abstain-0

Conditions/Questions for Developer:

1. Can you give us more information on how repeatability and "consistency" were determined? The results don't appear consistent.
2. Are patients with COPD excluded?
3. How are results reported and interpreted?
4. How would a smaller health plan implement this measure? It seems it might be too complex and burdensome.

Developer Response:

1. Repeatability was demonstrated by programming the measure in SAS code and the Ingenix software and comparing results. Because there are differences in what geographies these health plans are pulling from, variation is expected. But while differences across HCO's are expected, whether the differences are too high or low is difficult to know.
2. Patients are excluded from the asthma episode if they have more costs attributable to COPD than asthma.
3. The main measurement is the O/E ratio metric - the numerator of which is the cost of all the episodes of asthma, and the denominator which is the expected costs.
4. The burden depends on the plan's familiarity with ETGs and similar products, and for those who are just starting out, there is great deal of training involved (i.e. help desk support, etc.). There is another option where Ingenix takes the data and runs it themselves - or uses their PCQ Connect product that prepared the data into report-ready formats.

Overall Importance: Y-16, N-0

Committee Discussion: The Steering Committee agreed that asthma constitutes a high impact healthcare area.

Overall Reliability: H-1; M-14; L-1; I-0
Overall Validity: H-0; M-8; L-8; I-0
Overall Scientifically Acceptable: Yes [Split vote [Y-8; N-8 (Committee Vote)]
Committee Discussion: The Committee struggled with the circuitous reasoning behind asthma with acute exacerbation being a condition status and then having that condition status factor into the assignment of severity levels. Ingenix defended this methodology by explaining that for all measures, everything related to severity is based on utilization, which, although circular, is the best possible option. The Committee reiterated the TAP's concern that over half of asthma resource use costs are not captured in this measure since pharmacy data is not collected. They expressed unease about the incomparability of entities that have pharmacy data to those that do not.

Overall Usability: H-0; M-9; L-6; I-1
Committee Discussion: Several Steering Committee members challenged the idea that asthma should be thought of in terms of "episodes," as it is a chronic condition.

Overall Feasibility: H-1; M-8; L-7; I-0
Committee Discussion: All of the data is generated as a byproduct of care and is available electronically. There was concern around the measure's susceptibility to inaccuracies as Ingenix does not have a formal audit system to ensure that all of data is included and correct. Annually, for physician's the cost of the small package (less than 800 physicians) is $70,000, the medium package is $90,000 and the large package is $110,000 (over 2,000 physicians in the group). For health plans the package comes in small, medium and large; the small package is $90,115 (less than 400,000 covered lives) and the large package is $135,000 (over a million covered lives), all prices are annual rates. The Steering Committee came to the conclusion that this is no more expensive than other risk adjustment models another product of its caliber (for example, ACGs used by HealthPartners). However, the cost of the Ingenix product includes costs associated with the licensure of the proprietary software and the cost of all of their measures, over 558 ETGs. The Steering Committee acknowledges that the system may be used without Ingenix's technical support, it is complicated, but if the documentation is thoroughly read, it is doable.

1608: ETG Based Chronic Obstructive Pulmonary Disease Cost of Care Measure (COPD) (Ingenix)

Description: The measure focuses on resources used to deliver episodes of care for patients with COPD.
COPD episodes are 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 COPD. A number of resource use measures are defined for COPD episodes, including overall cost of care, cost of care by type of service, and the utilization of specific types of services. Each resource use measure is expressed as a cost or a utilization count per episode and comparisons with internal and external benchmarks are made using risk adjustment to support valid comparisons.
As requested by NQF, the focus of this submission is for COPD episodes and will cover both measures at the COPD base and severity level and also a COPD composite measure where COPD episode results are combined across COPD severity levels. At the most detailed level, the measure is defined as the base condition of COPD and an assigned level of severity (e.g., resources per episode for COPD, severity level 1 episodes). Composite measures can then be created using these measurement units to meet a specific need. For example, a composite measure for COPD is derived by combining COPD episode results across COPD severity levels. Appropriate risk adjustment is applied to support comparisons (e.g., for physician measurement, adjusting for a physician's mix of COPD episodes by severity level when supporting a COPD composite comparison). The focus of this measure is on COPD. However, COPD episode results could also be included in a "pulmonary" "chronic care", or other clinical composite for a physician, combining episodes in clinical areas similar to COPD. Further, an "overall" composite for a physician can be created, again by aggregating episode results across appropriate conditions and severity levels and applying proper risk adjustment when making comparisons.
Resource Use Type: Per episode
Data Type: Administrative claims, Other
Resource Use Service Categories:
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
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

Committee Recommendation for Endorsement: This measure did not pass the scientific acceptability criterion, and is not recommended for endorsement.

Conditions/Questions for Developer: N/A
Developer Response: N/A

Overall Importance: Y-16, N-0
Committee Discussion: There was unanimous agreement that asthma constitutes a high impact area of healthcare.

Overall Reliability: H-3; M-10; L-2; I-0
Overall Validity: H-1; M-5; L-9; I-0
Overall Scientifically Acceptable: Yes [Y-5; N-10 (Committee Vote)]
Committee Discussion: The Steering Committee appreciated the change Ingenix made to the measure's timeframe per the TAP's suggestion, from 180 to 365 days, to remain consistent with the asthma measure. It was felt the analysis of scientific acceptability for this measure would generally reflect the same analysis for measure NQF#1560. The Committee agreed that the measure did not provide sufficient validity and reliability testing.

Overall Usability: This measure did not pass the scientific acceptability criterion. As a result, the Committee did not discuss usability.

Overall Feasibility: This measure did not pass the scientific acceptability criterion. As a result, the Committee did not discuss feasibility.

1560: Relative Resource Use for People with Asthma (NCQA)

Description: This measure addresses the resource use of members identified as having asthma. Both encounter and pharmacy data are used to identify members for inclusion in the eligible population, and the results are adjusted to account for age, gender, and HCC-RRU risk classifications that predict cost variability (Refer to Attachment S8_Clinical Logic for additional information).
Resource Use Type: Per capita (population- or patient-based)
Data Type: Administrative claims; Electronic Clinical Data : Electronic Health Record; Electronic Clinical Data : Imaging/Diagnostic Study; Electronic Clinical Data : Laboratory; Electronic Clinical Data : Pharmacy Paper Records
Resource Use Service Categories: Inpatient services: Inpatient facility services, Inpatient services: Evaluation and management, Inpatient services: Procedures and surgeries, Inpatient services: Imaging and diagnostic, Inpatient services: Lab 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, Hospital/Acute Care Facility, Imaging Facility, Laboratory, Pharmacy, Post Acute/Long Term Care Facility: Nursing Home/Skilled Nursing Facility
Level of Analysis: Clinician: Group/Practice, Health Plan, Integrated Delivery System, Population : National, Population: Regional
Measure Developer: National Committee for Quality Assurance (NCQA), 1100 13th Street NW, STE 1000,Washington, District Of Columbia, 20005

Committee Recommendation for Endorsement: Y-13; N-0; Abstain-1

Conditions/Questions for Developer:

1. Could this measure be improved by including other diagnostic criteria to ensure all appropriate asthma patients are captured?
2. How have you come up with the age strata in your risk-adjustment?
3. Can secondary diagnosis be taken into account within the measurement year?
4. Is cost during the measurement year part of the risk-adjustment strategy?
5. Are your measure results published publically?

Developer Response:

1. Using asthma as a principal diagnosis will make it difficult to identify most patients, especially those who are acute and come into the ER and are diagnosed with bronchitis first, and then asthma.
2. The age strata for risk-adjustment are designed around known utilization patterns and clinical treatment patterns.
3. *All* costs for anyone with asthma are counted.
4. The HCC uses any services during the year to appropriately categorize patients into those 13 risk cohorts by severity of comorbidity. They also look at ICD-9 and procedural codes to categorize them and then go back and look at the number of times those services were offered to that population. Therefore, if a patient has multiple co-morbidities, that factors into the risk-adjustment, and will put a patient into a more severe risk-adjustment category.
5. Results are published through NCQA's Quality Compass module which contains the individual plan results by detailed service category along with a quality score.

Overall Importance: Y-16, N-0
Committee Discussion: The Steering Committee agrees this criterion has been met.

Overall Scientifically Acceptable: Yes [Y-12; N-2 (Committee Vote)]
Overall Reliability: H-12; M-3; L-0; I-0
Overall Validity: H-4; M-9; L-1; I-0
Committee Discussion: The Committee agreed with the TAP's analysis of reliability and raised no additional concerns. There was further discussion around missing pharmacy data, and confirmation that plans submit separate components (total medical, quality, and pharmacy, for example) to NCQA and are allowed to have a certain number of missing components. NCQA then holds the plans accountable for ensuring that they have the complete data required to report the measure, and any plans that are missing a major component of the measure specification would not end up in the NCQA reporting product. The Committee asked the developers to defend the measure's use of indirect standardization in creating standardized costs.

Overall Usability: H-9; M-5; L-0; I-0
Committee Discussion: The Steering Committee was concerned about the ability of small groups to implement this measure.

Overall Feasibility: H-10; M-4; L-0; I-0
Committee Discussion: No additional concerns were raised by the Steering Committee regarding feasibility.

1561: Relative Resource Use for People with COPD (NCQA)

Description: This measure addresses the resource use of members identified with COPD.
Clinical diagnosis of COPD during the measurement year is used to identify members for inclusion in the eligible population and the results are adjusted to account for age, gender, and HCC-RRU risk classifications that predict cost variability (Refer to Attachment S8_Clinical Logic for additional information).
Resource Use Type: Per capita (population- or patient-based)
Data Type: Administrative claims, Electronic Clinical Data: Electronic Health Record, Electronic Clinical Data: Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy, Paper Records
Resource Use Service Categories: Inpatient services: Inpatient facility services, Inpatient services: Evaluation and management, Inpatient services: Procedures and surgeries, Inpatient services: Imaging and diagnostic, Inpatient services: Lab 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, Hospital/Acute Care Facility, Imaging Facility, Laboratory, Pharmacy, Post Acute/Long Term Care Facility: Nursing Home/Skilled Nursing Facility
Level of Analysis: Clinician : Group/Practice, Health Plan, Integrated Delivery System, Population: Community, Population: National, Population : Regional
Measure Developer: National Committee for Quality Assurance (NCQA), 1100 13th street NW, STE 1000,Washington, District Of Columbia, 20005

Committee Recommendation for Endorsement: Y-13; N-0; Abstain-1

Conditions/Questions for Developer:

1. If the goal is to eventually link these measures with quality measures and stratification is different, how will that be plausible?
2. What is the upper age limit to be included in this measure?
3. How do you ensure similar populations are compared?

Developer Response:

1. The resource use strata are different than they are for clinical quality strata, which are not risk-adjusted. As the quality measures further increase and perhaps in the future become risk-adjusted, there will be more room for comparability.
2. There is no upper age limit to this measure.
3. By risk adjusting to the specified level using the HCC's and the 13 different cohorts, NCQA end up comparing relatively similar plan populations. The quality index for this measure is use of diagnostic spirometer and exacerbations measures. There is no attribution of specific procedures to COPD yet.

Overall Importance: Y-14, N-0
Committee Discussion: The Steering Committee agreed the measure focused on an important area of healthcare.

Overall Scientifically Acceptable: Yes [Y-13; N-1 (Committee Vote)]
Overall Reliability: H-11; M-3; L-0; I-0
Overall Validity: H-4; M-10; L-0; I-0
Committee Discussion: The Steering Committee was satisfied by the appropriateness of the risk-adjustment methodology employed to address the multiple co-morbidities associated with COPD. They agreed with the TAP's assessment of Scientific Acceptability and raised no new concerns.

Overall Usability: H-7; M-7; L-0; I-0 Committee Discussion: The Steering Committee valued NCQA's rigorous auditing processes and the transparency with which the developers construct their measures. In addition to being used by health plans, the Committee acknowledged the usefulness of measures for purchasers/providers, giving them much more leverage during negotiations for their annual purchasing agreements.
Overall Feasibility: H-10; M-4; L-0; I-0 Committee Discussion: No additional concerns were raised by the Steering Committee regarding feasibility.

HARMONIZATION AND BEST-IN-CLASS

In Phase One of this resource use measurement project, the Committee agreed that because this is NQF's first effort focused on evaluating resource use measures, identifying "best-in-class" and requiring harmonization among resource use measures would be premature. In the context of resource use measures, similar measures may share the same measure type (e.g., per episode, per capita), or measure the same costs/resources (e.g., actual cost vs. standard prices, resource service categories), or address the same population (e.g., people with diabetes). Competing measures would share all of the characteristics previously listed. Among the eight measures recommended for endorsement, there were no competing measures. Recommended measures that were the same measure type were submitted from the same developer and were already harmonized. With the exception of the two non-condition-specific total cost of care measures (submitted by the same developer and recommended in Cycle 1), which employ different costing methodologies; all recommended measures addressed different populations. Future resource use measure endorsement efforts should explore the potential ways in which harmonization among similar measures might be achieved. Specifically, identifying which measure constructs (e.g., condition-specific episode trigger and end mechanisms, age ranges), if any, could be harmonized for standard measurement is needed in this measurement area. Also, exploring the implications of harmonization for the resource use measure development community in which proprietary measure components are common would be useful as the portfolio of endorsed resource use measures expands.

ADDITIONAL RECOMMENDATIONS

As the first NQF resource use measure review and evaluation process concludes, there is a great opportunity to reflect on and provide recommendations for future efforts in this area. While resource use measurement has been used in the commercial sector for many years, the emerging interest in using these measures for public reporting and payment initiatives further highlights the need for efforts such as this to explore the complexities and potential challenges for multiple applications. The Committee was asked to provide guidance to the field on how measure of resource use can be improved and provide special considerations for developing measures for the Medicare program. Additionally, members provided insight on how measures of resource use may be linked with quality measures to provide a true assessment of efficiency. Through this

exercise, the Committee offered recommendations in several areas. In doing so, several new principles emerged.

1) **Submitting and Evaluating Resource Use Measures**

Emerging Principle 1: While guidelines in measure components may be acceptable for internal quality improvement, to promote measurement for comparison across entities nationally, the entire resource use measure construct should be standardized in the form of specifications.

Emerging Principle 2: The risk adjustment model applied to the measure should be specific to the intended population.

Emerging Principle 3: The factors in the risk adjustment model and severity model should be confirmed to be a contributor to the outcome of the measure.

Emerging Principle 4: In addition to statistical significance, justification of the variables used in the risk-adjustment model should be provided based on either clinical relevance or evidence in the literature.

2) **Reliability and Validity Testing**

Emerging Principle 5: To demonstrate reliability of a resource use measure, developers can focus on precision of the measure score.

Emerging Principle 6: The gold standard approach to determining the validity of data elements based on administrative claims data in resource use measures is to assess the agreement of claims data with source of the data elements in the chart.

3) **Data Quality and Comprehensiveness**

Emerging Principle 7: Data sets used to measure resources should be as comprehensive as possible. Efforts to obtain clinical and carved-out data (e.g., pharmacy, behavioral health) should be made to ensure the data set used to calculate resource use is robust, complete, and representative.

Emerging Principle 8: Measure scores calculated and reported using data with carve-outs should be labeled as such.

Emerging Principle 9: Comparisons of entities with and without carved-out data is inappropriate.

Emerging Principle 10: If a measure is intending to measure a clinical condition that encompasses a predominant portion of its costs in pharmacy claims, consider whether costs

should be measured at all in the absence of these data. It is the developers' responsibility to conduct an analysis to determine whether the lack of these data invalidates the measure score or comparisons.

4) Measuring Cost and Resource Use in the Medicare Population

Emerging Principle 11: A patient-centered approach should be used to describe the interaction of conditions (and episodes) in the development of resource use measures for the Medicare population.

5) Linking Quality and Cost to Develop Measures of Efficiency and Value

Emerging Principle 12: Efficiency measurement approaches should be patient-centered, building upon previous efforts such as the NQF Patient-Centered Episodes of Care (EOC) Efficiency Framework.

NEXT STEPS

The NQF member and public comment period for Cycle I will take place August 30-September 28, 2011. The Steering Committee will reconvene via conference call on October 10 to review and respond to NQF member and public comments that were received. Cycle II will be posted for NQF member and public comment October 19-November 17, 2011. A conference call will be scheduled at a later date to discuss the comments received.