

# NATIONAL QUALITY FORUM

## CONFERENCE CALL OF THE CARDIOVASCULAR/DIABETES TECHNICAL ADVISORY PANEL

July 14, 2011

*Committee Members Participating:* Jeptha Curtis, MD (*Co-Chair*, Yale University School of Medicine); James Rosenzweig, MD (*Co-chair*, Boston Medical Center and Boston University School of Medicine); Mary Ann Clark, MHA (Neocure Group); Kay Reeder, PhD, RN (University of Kansas School of Nursing); William Weintraub, MD (Christiana Care Health System); Thomas Marwick, MBBS, PhD (Cleveland Clinic).

*NQF Staff Participating:* Heidi Bossley, MSN, MBA, Vice President; Taroon Amin, MPH, MA, Senior Director; Ashlie Wilbon, BSN, MPH, Senior Project Manager; Laurelei Dorian, Project Manager; Sarah Fanta, Project Analyst; Sally Turbyville, MA, MS, NQF Consultant; Carlos Alzola, NQF Statistical Consultant.

*Others Present:* Thomas Lynn, Ingenix; Robin Wagner, American Board of Medical Specialties-Research and Education Foundation (ABMS-REF); Dawn Alayon, National Committee for Quality Assurance (NCQA); Sophia Chan, Centers for Medicare and Medicaid (CMS).

### MEETING PROCESS

Ms. Wilbon welcomed the Cardiovascular/Diabetes Technical Advisory Panel (TAP) and thanked them for their participation. The purpose of this conference call was to discuss NQF# 1591 ETG Based Congestive Heart Failure (CHF) resource use measure and NQF# 1594 ETG Based Coronary Artery Disease (CAD) resource use measure; both of the measures were submitted by Ingenix.

The measure developers were available on the call 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. General project information can be found by clicking on the [Resource Use project page](#).

### ABMS-REF MEASURE UPDATE

At this time the American Board of Medical Specialties Research & Education Foundation has decided to withdraw all of their measures from this project. The measure developers feel that the measures need more testing before undergoing the rigorous evaluation process. Since a number of the measures have undergone the evaluation process thus far, NQF will provide feedback to the development team so they can continue to refine and test the measures. For the remainder of the measures, mainly the pulmonary measures, the TAP and Steering Committee will still be asked to provide feedback.

# NATIONAL QUALITY FORUM

## MEASURE EVALUATION SUMMARY

The following summaries include a preliminary review of two cardiovascular measures submitted by Ingenix. The measure developers were present to give a brief overview of each measure.

<p><b>1591 ETG Based Congestive Heart Failure (CHF) resource use measure</b></p> <p><b>Description:</b> The measure focuses on resources used to deliver episodes of care for patients with Congestive Heart Failure (CHF). CHF 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 CHF. A number of resource use measures are defined for CHF 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 NOF, the focus of this submission is for CHF episodes and will cover both measures at the CHF base and severity level and also a CHF composite measure where CHF episode results are combined across CHF severity levels. At the most detailed level, the measure is defined as the base condition of CHF and an assigned level of severity (e.g., resources per episode for CHF, 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 CHF is derived by combining CHF episode results across CHF severity levels. Appropriate risk adjustment is applied to support comparisons (e.g., for physician measurement, adjusting for a physician's mix of CHF episodes by severity level when supporting a CHF composite comparison). The focus of this measure is on CHF. However, CHF episode results could also be included in a "cardiology", "chronic care", or other clinical composite for a physician, combining episodes in clinical areas similar to CHF. 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.</p> <p><b>Resource Use Type:</b> Per episode</p> <p><b>Data Type:</b> Administrative claims, other</p> <p><b>Resource Use Service Category:</b> 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</p> <p><b>Care Setting:</b> 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</p> <p><b>Level of Analysis:</b> Clinician : Group/Practice, Clinician : Individual, Clinician : Team, Facility, Health Plan, Integrated Delivery System</p> <p><b>Population :</b> Community, Population : County or City, Population : National, Population : Regional, Population : states</p> <p><b>Measure Developer:</b> Ingenix, 950 Winter Street, Suite 3800, Waltham, Massachusetts, 02451</p> <p><b>Conditions/Questions for Developer:</b></p> <ul style="list-style-type: none"> <li>• Why are some of the codes, typically seen in congestive heart failure measures, excluded?</li> <li>• How are hospitalizations that occur during the course of the measure handled?</li> <li>• Does the episode include events that occur before or after the episode?</li> </ul> <p><b>Developer Response:</b></p> <ul style="list-style-type: none"> <li>• Ingenix excluded the codes that were specific to diastolic heart failure (as this is a systolic and diastolic/systolic mix measure); if those codes were included it would have created another episode. Ingenix includes codes that were both systolic and diastolic, and used them as a marker to increase the severity score for the episode.</li> <li>• Hospital admissions that occurred during the course of the measure that are coded for congestive heart failure are included in the measure; hospitalizations are not used for severity adjustment . If the hospital admission date occurs during the measurement year, then the admission is included in that measurement year.</li> <li>• No, this measure is insulated from events that occur before or after the episode.</li> </ul> <p><b>Steering Committee Follow-up:</b> N/A</p> <p><b>If applicable, Questions to the Steering Committee:</b> N/A</p> <p><b>1. Importance to Measure and Report</b></p> <p><b>1a. High Impact:</b> <i>Pending TAP's Ratings</i></p> <p><b>Discussion:</b> The TAP believes this is a high impact, high cost area that is important to measure and report.</p> <p><b>1b. Resource use/cost problems:</b> <i>Pending TAP's Ratings</i></p> <p><b>Discussion:</b> The TAP believes this subcriterion has been met.</p> <p><b>1c. Purpose clearly described:</b> <i>Pending TAP's Ratings</i></p> <p><b>Discussion:</b> The TAP believes this subcriterion has been met.</p> <p><b>1d. Resource use service categories consistent and representative:</b> <i>Pending TAP's Ratings</i></p>
--

# NATIONAL QUALITY FORUM

*Discussion:* The TAP believes this subcriterion has been met.

## 2. Scientific Acceptability of Measure Properties:

### 2a. Reliability: *Pending TAP's Ratings*

#### 2a1. Well defined/precise specifications:

*Discussion:* There was a bit of confusion around the term, "congestive heart failure", it was brought up that not all "heart failure" is necessarily "congestive" and that there needs to be more clarification around the use of this term. The TAP agrees that this measure is targeting systolic and then a mix of systolic/diastolic heart failure. Ingenix also has a diastolic heart failure measure, but it has not been submitted for NQF endorsement. When the ICD9 code exists for systolic and diastolic – it's a marker for severity adjustment. Overall, the TAP believes that the clinical and construction logic of the measure was described in sufficient detail and the measure users should be able to implement it using the description provided.

#### 2a2. Reliability testing: *Pending TAP's Ratings*

*Discussion:* This measure has extensive benchmarking and comparisons; the TAP would have liked to see more external comparisons. The testing data submitted was from across nine health care organizations, all from large commercial insurers and were geographically varied. Ingenix demonstrated reliability by performing parallel development of the data by using two independent approaches. These two different approaches led to the same results, nearly 99.9% - this proves very strong reliability. The data was tested on a mostly commercial database, there were some Part C plan Medicare patients included also, however this measure was submitted for use in the commercial, less than 65 years old population.

#### 2b. Validity: *Pending TAP's Ratings*

#### 2b1. Specifications consistent with resource use/cost problem: *Pending TAP's Ratings*

*Discussion:* The TAP agrees that the specifications are consistent with the resource use.

#### 2b2. Validity testing: *Pending TAP's Ratings*

*Discussion:* The TAP believes Ingenix has sufficiently demonstrated face validity.

#### 2b3. Exclusions: *Pending TAP's Ratings*

*Discussion:* There are no exclusions within this measures, the TAP believes this subcriterion has been met.

#### 2b4. Risk adjustment: *Pending TAP's Ratings*

*Discussion:* The TAP believes that this risk adjustment appears to be somewhat circular – the measure is risk adjusted if the individual was hospitalized during the year – if the provider is using a large amount of resources, inevitably there will be more diagnoses in that measurement period, which would in turn also affect severity level category. Ingenix has made it clear that they are not using utilization to directly risk-adjust the cost of the episode. There is a lack of information in terms of the inclusion in the calibration- the risk groups are selected in terms of a cutoff for the severity score, there is not a rationale presented for why this cutoff point has been chosen.

#### 2b5. Identification of statistically significant/meaningful differences: *Pending TAP's Ratings*

*Discussion:* There is little information to compare statistical versus practical significance. The measure allows the user to determine what is clinically significant and use formula to determine the confidence intervals. The sample size appears sufficient enough to obtain a confidence interval that it will be useful to establish differences that are clinically and statistically significant. Ingenix is creating confidence intervals around the observed to expected ratio and stating that they are statistically different if they do not cross the threshold. The minimum sample count depends upon the case mix of the physicians and how different they are from the expected value.

#### 2b6. Multiple data sources:

*Discussion:* N/A

### 2c. Stratification for disparities:

*Discussion:* N/A

## 3. Usability:

### 3a. Measure performance results are publicly reported: *Pending TAP's Ratings*

*Discussion:* The TAP was concerned with the availability of this data to the public and requested clarification from NQF on what is required for "public reporting". The NQF Consensus Standards Approval Committee (CSAC) and the Board of Directors (BOD) continue to discuss this issue. The measures are widely used by providers to compare to one another. The results of this measure also allow for provider profiling, provider report cards and there is a cost base analysis for the members to estimate what the cost of the service would be, including the out of pocket expense. Since this measure is reported within a suite of measures, it has not been broken out individually for reporting or use in quality improvement.

### 3b. Measure results are meaningful/useful for public reporting and quality improvement: *Pending TAP's Ratings*

*Discussion:* The TAP agreed that more information would be needed to explain the results of this measure to the public and to be used for internal quality improvement.

### 3c. Data and results can be decomposed for transparency and understanding: *Pending TAP's Ratings*

*Discussion:* The TAP agrees there are challenges for the use of this measure, which include its complexity and lack of clarity in the specifications. TAP also agrees it is difficult to assess the extent to which the measure can be decomposed as it is currently specified.

### 3d. Harmonized or justification for differences:

*Discussion:* N/A

# NATIONAL QUALITY FORUM

<p><b>4. Feasibility:</b></p> <p><b>4a. Data elements routinely generated during care process:</b> <i>Pending TAP's Ratings</i>  <b>Discussion:</b> The TAP believes that this sub criterion has been met; all of the data elements are generated during the care process.</p> <p><b>4b. Data elements available electronically:</b> <i>Pending TAP's Ratings</i>  <b>Discussion:</b> The TAP believes that this sub criterion has been met; all of the data is available electronically.</p> <p><b>4c. Susceptibility to inaccuracies/ unintended consequences identified:</b> <i>Pending TAP's Ratings</i>  <b>Discussion:</b> Ingenix does not have a formal audit system to ensure that all of the numbers are included &amp; correct. In general, when dealing with any measure that uses administrative data there are various inaccuracies, pertaining particularly to coding inaccuracies and variation.</p> <p><b>4d. Data collection strategy can be implemented:</b> <i>Pending TAP's Ratings</i>  <b>Discussion:</b> 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 is \$110,000 (over 2000 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 another product of its caliber (for example, ACGs used by HealthPartners). The cost is not only for the licensure of the proprietary software, the cost includes all of the measure, 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.</p>
---

<p><b>1594 ETG Based Coronary Artery Disease (CAD) resource use measure</b></p> <p><b>Description:</b> The measure focuses on resources used to deliver episodes of care for patients with CAD. CAD 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 CAD. A number of resource use measures are defined for CAD 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 CAD episodes and will cover both measures at the CAD base and severity level and also a CAD composite measure where CAD episode results are combined across CAD severity levels. At the most detailed level, the measure is defined as the base condition of CAD and an assigned level of severity (e.g., resources per episode for CAD, 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 CAD is derived by combining CAD episode results across CAD severity levels. Appropriate risk adjustment is applied to support comparisons (e.g., for physician measurement, adjusting for a physician's mix of CAD episodes by severity level when supporting a CAD composite comparison). The focus of this measure is on CAD. However, CAD episode results could also be included in a "cardiology", "chronic care", or other clinical composite for a physician, combining episodes in clinical areas similar to CAD. 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.</p> <p><b>Resource Use Type:</b> Per episode</p> <p><b>Data Type:</b> Administrative claims, other</p> <p><b>Resource Use Service Category:</b> 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</p> <p><b>Care Setting:</b> 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</p> <p><b>Level of Analysis:</b> Clinician : Group/Practice, Clinician : Individual, Clinician : Team , Facility, Health Plan, Integrated Delivery System,          Population : Community, Population : County or City, Population : National, Population : Regional, Population : states</p> <p><b>Measure Developer:</b> Ingenix, 950 Winter Street, Waltham, Massachusetts, 02154</p>
<p><b>Conditions/Questions for Developer:</b> N/A</p> <p><b>Developer Response:</b> N/A</p> <p><b>Steering Committee Follow-up:</b> N/A</p> <p><b>If applicable, Questions to the Steering Committee:</b> N/A</p>
<p><b>1. Importance to Measure and Report</b></p> <p><b>1a. High Impact:</b> <i>Pending TAP's Ratings</i>  <b>Discussion:</b> The TAP believes this is a high impact, high cost area; this sub criterion has been met.</p> <p><b>1b. Resource use/cost problems:</b> <i>Pending TAP's Ratings</i>  <b>Discussion:</b> The TAP believes this subcriterion has been met.</p>

# NATIONAL QUALITY FORUM

<p><b>1c. Purpose clearly described:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP believes this subcriterion has been met.</p> <p><b>1d. Resource use service categories consistent and representative:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP believes this subcriterion has been met.</p>
<p><b>2. Scientific Acceptability of Measure Properties:</b></p> <p><b>2a. Reliability:</b></p> <p><b>2a1. Well defined/precise specifications:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The diagnosis codes for this measure are the 410s through 414s and then the 429s, all of which represent complications of myocardial infarction. These codes seem comprehensive for identifying patients with coronary artery disease; however, it raises the question if the populations are similar enough that the user can reasonably make inferences about the resource use needed for each type of cardiac episode. Overall, the measure is very well specified and is being used across different health plans.</p> <p><b>2a2. Reliability testing:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The measure is specified in a way that it has been used over a long period of time, Ingenix demonstrated that if the user uses the same measure in the same population then the result will be the same. The TAP believes this subcriterion has been met.</p> <p><b>2b. Validity:</b></p> <p><b>2b1. Specifications consistent with resource use/cost problem:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP believes this subcriterion has been met; a specific population is defined and measured.</p> <p><b>2b2. Validity testing:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP believes Ingenix has sufficiently demonstrated face validity.</p> <p><b>2b3. Exclusions:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> There are no exclusions within this measure, the TAP believes this subcriterion has been met.</p> <p><b>2b4. Risk adjustment:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP requested that the developer demonstrate proof of the concept that this is accurately accounting for differences in the population – the risk adjustment method does not appear to be robust. Additional information the model's goodness of fit was requested. NQF staff is working with Ingenix to supply this information to the Steering Committee.</p> <p><b>2b5. Identification of statistically significant/meaningful differences:</b> <i>Discussion:</i> The Steering Committee believes that this measure did not identify statistically significant or meaningful differences across groups. There was general concern that something may be classified as statistically significant, when it really is not clinically significant.</p> <p><b>2b6. Multiple data sources:</b> <i>Discussion:</i> N/A</p> <p><b>2c. Stratification for disparities:</b> <i>Discussion:</i> N/A</p>
<p><b>3. Usability:</b></p> <p><b>3a. Measure performance results are publicly reported:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP was concerned with the availability of this data to the public and requested clarification from NQF on what is required for "public reporting". The NQF Consensus Standards Approval Committee (CSAC) and the Board of Directors (BOD) continue to discuss this issue. The measures are widely used by providers to compare to one another. The results of this measure also allow for provider profiling, provider report cards and there is a cost base analysis for the members to estimate what the cost of the service would be, including the out of pocket expense. Since this measure is reported within a suite of measures, it has not been broken out individually for reporting or use in quality improvement.</p> <p><b>3b. Measure results are meaningful/useful for public reporting and quality improvement:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP agreed that more information would be needed to explain the results of this measure to the public and to be used for internal quality improvement.</p> <p><b>3c. Data and results can be decomposed for transparency and understanding:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP agreed there are challenges for the use of this measure, which include its complexity and lack of clarity in the specifications. TAP also agreed it is difficult to assess the extent of which the measure can be decomposed as it is currently specified.</p> <p><b>3d. Harmonized or justification for differences:</b> <i>Discussion:</i> N/A</p>
<p><b>4. Feasibility:</b></p> <p><b>4a. Data elements routinely generated during care process:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP believes that this sub criterion has been met; all of the data elements are generated during the care process.</p> <p><b>4b. Data elements available electronically:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> The TAP believes that this sub criterion has been met; all of the data is available electronically.</p> <p><b>4c. Susceptibility to inaccuracies/ unintended consequences identified:</b> <i>Pending TAP's Ratings</i> <i>Discussion:</i> Ingenix does not have a formal audit system to ensure that all of the numbers are included &amp; correct. In general, when dealing with any measure that uses administrative data there are various inaccuracies, pertaining particularly to coding inaccuracies and variation.</p>

# NATIONAL QUALITY FORUM

## 4d. Data collection strategy can be implemented: *Pending TAP's Ratings*

**Discussion:** 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 is \$110,000 (over 2000 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 another product of its caliber (for example, ACGs used by HealthPartners). The cost is not only for the licensure of the proprietary software, the cost includes all of the measure, 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.

## **PUBLIC COMMENT**

There were no public comments.

## **NEXT STEPS**

Ms. Wilbon thanked the TAP for their time and effort throughout this process. The TAP results will be forwarded on to the Steering Committee as they review the measures.