



NATIONAL  
QUALITY FORUM

1030 15TH STREET, NW, SUITE 800  
WASHINGTON, DC 20005  
202 783-1300 MAIN 783-3434 FAX

**NATIONAL VOLUNTARY CONSENSUS STANDARDS  
FOR COST AND RESOURCE USE:  
TECHNICAL REPORT**

**APRIL 2012**

**NATIONAL VOLUNTARY CONSENSUS STANDARDS  
FOR COST AND RESOURCE USE, 2012  
Technical Report**

**TABLE OF CONTENTS**

<b>BACKGROUND.....</b>	<b>2</b>
<b>MEASURE EVALUATION.....</b>	<b>4</b>
<b>Evaluation of measurement approaches.....</b>	<b>5</b>
<b>Endorsed Consensus Standards.....</b>	<b>6</b>
(1557) Relative Resource Use for People with Diabetes (NCQA).....	7
(1558) Relative Resource Use for People with Cardiovascular Conditions (NCQA).....	9
(1560) Relative resource use for people with asthma (NCQA).....	12
(1561) Relative resource use for people with COPD (NCQA).....	15
(1611) ETG based pneumonia cost of care (Ingenix/OptumInsight).....	17
(1598) Total Resource Use Population-based PMPM Index (HealthPartners)....	19
(1604) Total Cost of Care Population-based PMPM Index (HealthPartners)....	23
(1609) ETG based hip/knee replacement cost of care (Ingenix/OptumInsight)...	26
<b>Candidate Consensus Standards Not Recommended For Endorsement.....</b>	<b>28</b>
(1591) ETG-based congestive heart failure (CHF) cost of care (Ingenix/OptumInsight).....	29
(1594) ETG-based coronary artery disease (CAD) cost of care (Ingenix/OptumInsight).....	32
(1595) ETG based diabetes cost of care (Ingenix/OptumInsight).....	35
(1599) ETG-based non-condition specific cost of care (Ingenix/OptumInsight)..	38
(1603) ETG-based hip fracture cost of care (Ingenix/OptumInsight).....	41
(1605) ETG-based asthma cost of care (Ingenix/OptumInsight).....	45
(1608) ETG-based chronic obstructive pulmonary disease (COPD) cost of care (Ingenix/OptumInsight).....	47
<b>Recurring Measure Evaluation Themes.....</b>	<b>49</b>
Costing Approaches.....	49
Feasibility.....	50
<b>WITHDRAWN FROM CONSIDERATION.....</b>	<b>51</b>
<b>ENDNOTES.....</b>	<b>52</b>

# NATIONAL VOLUNTARY CONSENSUS STANDARDS FOR COST AND RESOURCE USE Technical Report

## BACKGROUND

The United States' healthcare expenditures are unmatched by any country in the world.<sup>1</sup> This spending, however, has not resulted in better health for Americans. In fact, higher spending has not decreased mortality, increased patient satisfaction, or led to improvements in access or higher quality of care.<sup>2,3,4</sup> This phenomenon of high spending with disproportionate outcomes points to a system laden with waste. The contributing factors to this alarming trend are as complex as the health care system itself, with physician practice patterns, regional market influences, and access to care as major players. Meanwhile, the United States' healthcare spending continues to increase at a rate of seven percent per year and is largely focused on treating acute and chronic illness rather than preventive care.<sup>5</sup>

As ongoing health reform efforts focus on expanding coverage, increasing access to care, and reducing costs, it is important to understand how resources are currently being used in the system in the context of quality, preferably related to health outcomes. Linking resource use (or cost) and quality measures will enable the system to better evaluate efficiency of care. Several provisions in the Affordable Care Act (ACA), slated to be implemented over the next five years, require using resource use data to further support efforts to move toward a value-based purchasing (VBP) payment model. One such provision requires the Secretary of Health and Human Services to develop an episode grouper that combines separate but clinically related items and services into an episode of care for an individual.<sup>6</sup> Additionally, resource use data will be included on the physician compare website, as well as a physician value modifier that will be used to adjust fee-for-service (FFS) payments by combining physician performance on quality and resources use. While the ACA legislation is focused on the Medicare population, understanding resource use measurement as a building block of efficiency, even in the context of commercial-based measures, is a first step toward meeting these goals.

For the purposes of this project, resource use measures are defined as broadly applicable and comparable measures of health services counts (in terms of units or dollars) that are applied to a

population or event (broadly defined to include diagnoses, procedures, or encounters). A resource use measure counts the frequency of defined health system resources; some may further apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to each unit of resource use. Current approaches for measuring resource use range from broadly focused measures, such as per capita measures, which address total healthcare spending (or resource use) per person, to those with a more narrow focus, such as measures dealing with the healthcare spending or resource use of an individual procedure (e.g., a hip replacement).

These measures will serve as building blocks for efficiency of care measures and signal to the measure development industry the urgent need to develop resource use and efficiency measures that integrate quality domains. Phase one of this work, which began in 2009, was aimed at understanding resource use measures and identifying the important attributes to consider in their evaluation. During this phase, the current NQF Measure Evaluation Criteria used to evaluate quality measures was reviewed and refined by the Resource Use Steering Committee to address the unique aspects of resource use measures, resulting in the [NQF Resource Use Measure Evaluation Criteria](#). A single Steering Committee was used across both phases of work, with the addition of four Technical Advisory Panels (TAPs) in Phase two to assist the Committee in evaluating the measures' clinical and methodological aspects. The CDP project was divided into two review cycles, between which 14 focus areas were assigned:

### **Cycle 1**

#### **Cardiovascular**

- Congestive heart failure (CHF)
- Coronary artery disease (CAD)
- Acute myocardial infarction (AMI)

#### **Stroke**

#### **Diabetes**

### **Cycle 2**

#### **Pulmonary**

- Chronic obstructive pulmonary disease (COPD)
- Asthma
- Pneumonia

#### **Cancer**

- Breast cancer
- Colorectal cancer

#### **Bone/Joint**

- Hip or knee replacement
- Hip or pelvic fracture
- Low back pain

Eight measures were endorsed:

- (1557) Relative Resource Use for People with Diabetes (NCQA)
- (1558) Relative Resource Use for People with Cardiovascular Conditions (NCQA)
- (1560) Relative resource use for people with asthma (NCQA)
- (1561) Relative resource use for people with COPD (NCQA)
- (1611) ETG based pneumonia cost of care (Ingenix/OptumInsight)
- (1598) Total Resource Use Population-based PMPM Index (HealthPartners)
- (1604) Total Cost of Care Population-based PMPM Index (HealthPartners)
- (1609) ETG based hip/knee replacement cost of care (Ingenix/OptumInsight)

For more details on the overarching issues and recommendations for future work identified by the Committee in this initial effort to evaluate and endorse resource use measures, refer to the National Voluntary Consensus Standards for Cost and Resource Use Final Report.

## **MEASURE EVALUATION**

This report reflects the Committee's discussion of 15 cost and resource use measures evaluated for suitability as voluntary consensus standards for accountability across the two review cycles in Phase Two of this project. Candidate consensus standards were solicited through a Call for Measures on January 31, 2011. The measures were evaluated using [NQF Resource Use Measure Evaluation Criteria](#). Four condition-focused TAPs for pulmonary, cardiovascular and diabetes, bone and joint, and cancer conditions rated each candidate consensus standard according to the subcriteria and identified strengths and weaknesses to assist the Committee in making recommendations. The 23-member, multi-stakeholder Committee evaluated the subcriteria of the non-condition specific measures, provided final evaluations of the four main criteria—importance to measure and report, scientific acceptability of the measure properties, usability, and feasibility—and made endorsement recommendations for all measures. Measure developers were available during TAP and Committee discussions to respond to questions and clarify any issues or concerns.

	<b>TOTAL</b>
Measures under consideration	36
Withdrawn from consideration by developer	21
Endorsed	8
Not endorsed	7

### **Evaluation of Measurement Approaches**

The NQF measure evaluation process calls for each submitted measure to be evaluated individually, based on its own merit. This was also the approach used in this project.

Additionally in this project, given the various approaches by resource use measure developers, measures developed by a single developer shared many common underlying measure constructs and processes. By understanding the common constructs shared among a group of measures from a developer (i.e. general methods), it lays the foundation for understanding the nuances specific to each individual measure. During the measure evaluation process, the TAPs and Committees often identified some recurring themes within the criteria discussions that applied across measures from an individual developer, regardless of condition focus of the individual measure. Some of these recurring themes have been captured in several of the measure evaluation summaries and summarized at the end of this report.

**ENDORSED CONSENSUS STANDARDS**

Hyperlinks are provided from each summary table to the detailed measure specifications. To access the meeting transcripts and recordings in which these measures are discussed, refer to the [project web page](#).

***Diabetes***

(1557) Relative Resource Use for People with Diabetes (NCQA).....7

***Cardiovascular***

(1558) Relative Resource Use for People with Cardiovascular Conditions (NCQA).....9

***Pulmonary***

(1560) Relative Resource Use for People with Asthma (NCQA).....12  
(1561) Relative Resource Use for People with COPD (NCQA).....15  
(1611) ETG-Based Pneumonia Cost of Care (Ingenix/OptumInsight).....17

***Non-Condition Specific***

(1598) Total Resource Use Population-based PMPM Index (HealthPartners).....19  
(1604) Total Cost of Care Population-based PMPM Index (HealthPartners).....23

***Bone/Joint***

(1609) ETG/PEG-Based Hip/Knee Replacement Cost of Care  
(Ingenix/OptumInsight).....26

**1557: Relative Resource Use for People with Diabetes**

**Description:** The risk-adjusted relative resource use by health plan members 18-75 years of age who were identified as having diabetes (type 1 and type 2) during the measurement year.

**Resource Use Measure Type:** Per capita (population- or patient-based)

**Data Source:** Administrative claims

**Resource Use Service Category:** 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: Clinic/Urgent Care; Ambulatory Care: Clinician Office; Hospital/Acute Care Facility; Imaging Facility; Laboratory; Pharmacy

**Level of Analysis:** Clinician: Group/Practice, Health Plan, Integrated Delivery System, Population : National, Population: Regional

**Measure Developer:** National Committee for Quality Assurance (NCQA)

**Committee Recommendation for Endorsement:** Y-17; N-0; Abstain-1

**If applicable, Conditions/Questions for Developer and Developer response:**

- In relation to criterion 2a.1, provide information on which maternity codes are included.
- In relation to criterion 2b.3, provide rationale for excluding patients >75 years old.

**TAP Evaluation:**

**1. Importance to Measure and Report:**

**1a. High Impact:** H- 9 , M-0, L-0, I-0, N/A-0

**TAP Discussion:** Developer provided sufficient evidence and support.

**1b. Resource use/cost problems:** H- 9, M-0, L-0, I-0, N/A-0

**TAP Discussion:** Developer provided sufficient evidence and support.

**1c. Purpose clearly described:** H- 8, M-1, L-0, I-0, N/A-0

**TAP Discussion:** Developer provided sufficient evidence and support.

**1d. Resource use service categories consistent and representative:** H- 7, M-2, L-0, I-0, N/A-0

**TAP Discussion:** The resource use service categories were sufficient.

**Overall Importance:** Y-17, N-0

**Committee Discussion:** While the measure is deemed important, the Committee pointed out the resources accounted for in the measure do not include important services provided to diabetic patients, including care coordination, and education by nurses and nutritionists. These services are typically not billed, services and so they are often left out of the cost calculations for measures using administrative claims data. This type of measurement is possible, but NCQA does not generally have access to this level of specificity in the data.

**TAP Evaluation**

**2. Scientific Acceptability of Measure Properties:**

**2a1. Well defined/precise specifications:** H- 8, M-0, L-0, I-0

**TAP Discussion:** The TAP had concerns about how are changing codes are handled. It was stated that this is very difficult to manage in all measures. Concern was also expressed related to adjusting away patients with lots of claims; conditions such as HIV and active cancer are excluded (this adjustment is made every year with a one year lag).

The intent of this measure is to capture all costs for a diabetic patient, including services that may not be related to a diabetes diagnosis. While counting all costs does add some noise to the measure, there is evidence that diabetics stay in hospital longer, even for stays triggered by non-diabetes related events. With a minimum sample size of 400, this measure has been specified for use at the health plan level; not for use at the physician level. TAP had concerns as to why conditions that are proven to be related to diabetes complications are not included, for example, amputations, ESRD, etc. The TAP wanted clarification on whether pregnancy/maternity codes were included in this measure.

**2a2. Reliability testing:** H- 9, M-0, L-0, I-0, N/A-0

**TAP Discussion:** Reliability testing was acceptable.

**2b1. Specifications consistent with resource use/cost problem:** H- 5, M-4, L-0, I-0, N/A-0

**TAP Discussion:** Measure captures all costs for a diabetes patient.

**2b2. Validity testing:** H- 5, M-4, L-0, I-0, N/A-0

**TAP Discussion:** Adequate validity testing information provided.

**2b3. Exclusions:** H- 6, M-3, L-0, I-0, N/A-0

**TAP Discussion:** The TAP expressed concern over the age limit criteria; Age 75 may be too low.

**2b4. Risk-adjustment :** H- 9, M-0, L-0, I-0, N/A-0



<p><b>1557: Relative Resource Use for People with Diabetes</b></p> <p><i>TAP Discussion:</i> Measure uses HCC's for the risk-adjustment. The TAP agrees this is acceptable methodology.</p> <p><b>2b5. Identification of statistically significant/meaningful differences:</b> H- 9 , M-0, L-0 , I-0,N/A-0</p> <p><i>TAP Discussion:</i> Minimum sample size at 400 allows for increased statistical stability.</p> <p><b>2b6. Multiple data sources:</b> H- 0, M-0, L-0 , I-0, N-9, N/A-0</p> <p><i>TAP Discussion:</i> N/A</p> <p><b>2c. Stratification for disparities:</b> H- 2, M-5 , L-1, I-0,N-, N/A-0</p> <p><i>TAP Discussion:</i> Can only be stratified only for age, gender and region, as with most of the measures submitted.</p>
<p><b>Overall Scientifically Acceptable: Yes [Y-18; N-0 (Committee Vote)]</b></p> <p><b>Committee Discussion:</b> There was acknowledgement that certain types of claims and clinicians are invisible in these types of measures because administrative claims data does not capture all resource use or recognize the resources used by of all types of clinicians. The Committee also pointed out that a broad scope of cost codes is important, and the thinking about measuring resources should be expanded beyond intermediate care (e.g., consider home health costs, skilled nursing facilities, etc.) There was discussion on the use of the standardized pricing tables and how they are applied within the measures. These pricing tables are now publicly available on the NCQA website and can be used by anyone for their own purposes. A number of resources have been used to develop the tables, including the Medicare fee schedule and data from thousands of pharmacy prescriptions.</p> <p>The Committee discussed the TAP's concern over the exclusion of patients over the age of 75 and the mandatory exclusions for active cancer, transplantation, ESRD, and HIV that are applied to all NCQA measures, but are particularly relevant to the diabetes population. The developers are going back to re-examine these exclusions for future versions of the measure.</p> <p>The final concern the Committee was related to the logic of the truncation scheme. In order to avoid a small proportion of members driving up the standardized costs, the developers identified cap levels at which members would be capped and truncated once costs reach that high level; however, they are not excluded. This prevents skewing of the results. The timeframes used in the measure logic were in attempt to focus on a group of patients who are not newly diagnosed.</p>
<p><b>TAP Evaluation:</b></p> <p><b>3. Usability:</b></p> <p><b>3a. Measure performance results are publicly reported:</b> H- 9, M-0 , L-0 , I-0, N/A-0</p> <p><i>TAP Discussion:</i> Measure is currently in use by large number of health plans.</p> <p><b>3b. Measure results are meaningful/useful for accountability and quality improvement:</b> H- 8, M-1 ,L-0 , I-0, N/A-0</p> <p><i>TAP Discussion:</i> Accountability mechanism sufficient.</p> <p><b>3c. Data and results can be decomposed for transparency and understanding:</b> H- 8, M-1, L-0, I-0, N/A-0</p> <p><i>TAP Discussion:</i> Specifications adequate for transparency.</p> <p><b>3d. Harmonized or justification for differences:</b> N/A</p> <p><i>TAP Discussion:</i> Developers were not asked to harmonize prior to submissions.</p>
<p><b>Overall Usability: H-12; M-6; L-0; I-0</b></p> <p><b>Committee Discussion:</b> The Committee did not identify any additional issues for this criterion.</p>
<p><b>TAP Evaluation:</b></p> <p><b>4. Feasibility:</b></p> <p><b>4a. Data elements routinely generated during care process:</b> H- 9, M-0, L-0 , I-0, N/A-0</p> <p><i>TAP Discussion:</i> Measures rely on administrative data.</p> <p><b>4b. Data elements available electronically:</b> H- 9, M-0, L-0 , I-0, N/A-0</p> <p><i>TAP Discussion:</i> Administrative data are in electronic format.</p> <p><b>4c. Susceptibility to inaccuracies/ unintended consequences identified:</b> H- 6,M-3, L-0, I-0, N/A-0</p> <p><i>TAP Discussion:</i> Users of NCQA are subject to a data audit process. Susceptibility to errors/inaccuracies is low.</p> <p><b>4d. Data collection strategy can be implemented:</b> H- 9, M-0, L-0 , I-0, N/A-0</p> <p><i>TAP Discussion:</i> Barriers to use are low.</p>
<p><b>Overall Feasibility: H-11; M-7; L-0; I-0</b></p> <p><b>Committee Discussion:</b> There were no additional concerns identified by the Committee.</p>

**1558 Relative Resource Use for People with Cardiovascular Conditions**

**Description:** The risk-adjusted relative resource use by health plan members with specific cardiovascular conditions during the measurement year.

**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 Category:** Inpatient services: Inpatient facility services; Inpatient services: Evaluation and management; Inpatient services: Procedures and surgeries; 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: Clinic/Urgent Care; Ambulatory Care: Clinician Office; Hospital/Acute Care Facility; Imaging Facility; Laboratory; Pharmacy

**Level of Analysis:** Clinician: Group/Practice, Health Plan, Integrated Delivery System, Population : National, Population: Regional

**Measure Developer:** National Committee for Quality Assurance (NCQA)

**Committee Recommendation for Endorsement:** Y-13; N-3; Abstain-1

**TAP Conditions/Questions for Developer:**

1. Are other conditions similar to Coronary Artery Disease included, such as ischemic heart disease?
2. How does the stratification discern between high- and low -risk patients?
3. What is the time frame for exclusions?
4. How would a provider know how to improve based on the report?

**Developer Response:**

1. This measure is based on the HEDIS measure, covering both acute and sub-acute, ischemic heart disease, cardiovascular unspecified, angina, atherosclerosis of extremity, etc. CAD-related codes diverged into family history, etc. The measure does not try to account for anything other than what CAD is described as in the code set. The developer is going to look into including code sets that are non -CAD -specific for non-traditional patients.
2. In terms of stratification for the risk-adjustment, it is dependent on the number of comorbidities. Section 10.1 includes additional information on the risk -adjustment methods, identifies based on qualifying and HCC rankings.
3. The time frames align with the eligible population period; patients are looked at a year prior to the measurement year and are looked at the year prior to and during the eligibility period.
4. The reports are divided up by resource categories; user would need to look into measure specifications, which are fairly broad.

**Committee Follow-up:**

- Has this type of risk -adjustment model been validated in the past?
  - **Response:** HCC's are well validated. RTI evaluated this in April 2011, and it continues to be a valid stratification method.
- The Committee wanted additional follow-up on the time period for eligibility for risk-adjustment/ exclusions.

**TAP Evaluation:**

**1.Importance to Measure and Report**

**1a.High Impact:** H-5; M-0; L-0; I-0

**TAP Discussion:** The TAP agreed that this subcriterion has been met.

**1b. Resource use/cost problems:** H-5; M-0; L-0; I-0

**TAP Discussion:** The TAP agreed this subcriterion has been met and is supported by the evidence.

**1c. Purpose clearly described:** H-3; M-3; L-0; I-0

**TAP Discussion:** Inclusion criteria for this measure are very broad – PCI and CABG, but not other codes are associated with chronic conditions. It would be difficult for this measure to be actionable by an individual provider because of the broad nature of the category. The costs of carotid disease are included in the category. It does capture costs, but there is the issue of which costs are incorporated and which costs are not. Given the broad category, the calculation of costs is difficult for a user to understand. This measure covers all costs across all procedures and excludes those who were screened and had plaque in their carotid paired equally as with those with PCIs and that early detection may become a preponderance of those grouped in cardiovascular disease.

**1d. Resource use service categories consistent and representative:** H-2; M-3; L-0; I-0

**TAP Discussion:** The TAP agreed that this subcriterion has been met.

**Overall Importance:** Y-14; N-1; Abstain-1

**Committee Discussion:** There were no additional concerns identified by the Committee for this criterion.

**TAP Evaluation:**

**2.Scientific Acceptability of Measure Properties:**

**2a. Reliability:**

**2a1. Well- defined/precise specifications:** H-2; M-1; L-1, I-0

*TAP Discussion:* The TAP this to be a relevant way to approach the measure, as each grouping and person is stratified according to risk. It is unclear which risk-adjustment is used for each patient. This measure is calculated by using databases from insurers, up to age 75, and only reports only on organizations with more than 400 people in the measure. This measure is restricted in use for larger groups.

**2a2. Reliability testing:** H-2; M-2; L-0; I-0

*TAP Discussion:* The reliability testing uses data for 15 months. The results are consistent with other models.

**2b. Validity:**

**2b1. Specifications consistent with resource use/cost problem:** H-1; M-2; L-2; I-0

*TAP Discussion:* Discussion similar to 2a1. It is unclear which risk-adjustment is used for which patient.

**2b2. Validity testing:** H-2; M-2; L-0; I-0

*TAP Discussion:* NCOA publicly reported the results annually and continues to publicly report publicly. The costs are standardized and are good measures of the resources being used. There is a track record of data being clean, including resource use not what was actually charged.

**2b3. Exclusions:** H-1; M-2; L-1; I-1

*TAP Discussion:* The measure is unclear regarding the time period for exclusions.

**2b4. Risk-adjustment:** H-1; M-2; L-1; I-0

*TAP Discussion:* It is difficult to discern what is included in risk -adjustment criteria. Unclear how stratification is working and if the groups produced is are legitimate.

**2b5. Identification of statistically significant/meaningful differences:**

*TAP Discussion:* The Committee has agreed this subcriterion has been met.

**2b6. Multiple data sources:** H-1; M-4; L-0; I-0; N/A-0

*TAP Discussion:* N/A

**2c. Stratification for disparities:** H-0; M-4; L-0; I-0; N/A-1

*TAP Discussion:* This measure stratifies for age and gender.

**Overall Scientifically Acceptable : Yes [Y-13; N-4 (Committee Vote)]**

*Committee Discussion:* Concerns with comparing like plans (e.g., Medicaid to Medicaid plans). Developer acknowledged that there has been testing of the measure at the group practice level; however, it was only tested with over 400 patients. Subsequently, the Committee clarified that the measure can be used at the group practice level with a minimum sample size of 400 patients. The Committee was interested in the exclusions for end stage renal disease (ESRD). The Committee was concerned with the peer group comparison of "like plans" because there might be correlations with socioeconomic status (SES) across plans. Further, the Committee was concerned over the appropriateness of excluding patients who are >75 years old.

The risk -adjustment model used in this measure includes HCCs where risk-adjustment takes into account the resource use from within the measurement year. The Committee agreed that a better title for the measure might be "Measure of Patients with Chronic Cardiac Conditions."

While the Committee was concerned with the level of measurement, the developer clarified that it would only be used at population level, and reported with quality measures. Purchasers and health plan representatives agreed that this measure would be useful.

**TAP Evaluation:**

**3. Usability:**

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

*TAP Discussion:* This measure has been utilized for a short amount of time (since 2007); it is difficult to assess if the manner in which they are reporting is useful.

**3b. Measure results are meaningful/useful for accountability and quality improvement:** H-2; M-1; L-1; I-0

*TAP Discussion:* There is no data on how consumers are utilizing the data and making changes based on this measure. It is unclear what would or would not affect the score and change practices in the long run. It may not be extremely useful for accountability as it's it is not easily interpreted.

**3c. Data and results can be decomposed for transparency and understanding:** H-2; M-1; L-1; I-0

*TAP Discussion:* The measure is very broad and it's unclear how providers can change behavior.

**3d. Harmonized or justification for differences:** N/A

*TAP Discussion:* N/A

**Overall Usability: H-6; M-9; L-2; I-0**

*Committee Discussion:* The Committee was comfortable with the measure since it has been in use for 5 years. They did express concern over how the results will be used for consumers. The breakdown within the service categories was found to be more useful information than the overall score. There are currently 800 out of 1100 plans reporting RRU/quality measures with less than 1% of the health plans as outliers.

The Committee was not as concerned with "carve -outs" since pharmacy costs are reported separately from medical costs. There was

interest in how to make this kind of data meaningful for consumers as well. The developer clarified that the major users are employer groups and business groups, and it helps to inform their decisions for the following year. However, skepticism had been expressed regarding the usability at the plan level.

**TAP Evaluation**

**4. Feasibility:**

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

*TAP Discussion:* All administrative data is generated as a byproduct of care.

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

*TAP Discussion:* All data is available electronically.

**4c. Susceptibility to inaccuracies/ unintended consequences identified:** H-1; M-4; L-0; I-0

*TAP Discussion:* This subcriterion has been met.

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

*TAP Discussion:* This subcriterion has been met.

**Overall Feasibility:** H-7; M-6; L-3; I-1

**Committee Discussion:**

The developer explained that health plans calculate observed measure scores but NCQA calculates the expected for the final measure score. The Committee was interested in how carve-outs and capitated arrangements were addressed. Data within the measure are reported out into each service categories with pharmacy benefits measured separately.

**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)

**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.

**1. Importance to Measure and Report**

**1a.High Impact:** H-9; M-0; L-0; I-0

**TAP Discussion:** The TAP agrees that asthma is an important area of healthcare to measure due to its high cost and the potential for improvements in care.

**1b. Resource use/cost problems:** H-7; M-2; L-0; I-0

**TAP Discussion:** The TAP agrees that asthma represents a resource use problem and noted that there is a well-documented opportunity for improvement.

**1c. Purpose clearly described:** H-9; M-0; L-0; I-0

**Discussion:** The TAP believes the purpose and objective are clear; this subcriterion has been met.

**1d. Resource use service categories consistent and representative:** H-9; M-0; L-0; I-0

**TAP Discussion:** The TAP believes this subcriterion has been met; there were not issues raised.

**Overall Importance:** Y-16, N-0

**Committee Discussion:** The Steering Committee agrees this criterion has been met.

**2. Scientific Acceptability of Measure Properties:**

**2a. Overall Reliability:** H-8; M-1; L-0; I-0

**2a1.Measure well defined and precisely specified:** H-9; M-0; L-0; I-0

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

**2a2. The results are repeatable:** H-8; M-1; L-0; I-0

**TAP Discussion:** There was general agreement from the TAP that following a methodology of including all costs avoids having to consider what costs should or should not be associated with asthma. The developer reaffirmed that the measures are valid for any health plan; they are population-based measures and have been tested and can be used in physician groups with a sufficient number of

patients. A population of at least 400 members is needed for the methodology to be valid, so it consequently tends to be larger physician groups that can use the measures.

**2b. Overall Validity:** H-5; M-4; L-0; I-0

**2b1. Evidence is consistent with intent:** H-6; M-3; L-0; I-0

**TAP Discussion:** The TAP agrees there is good overall evidence of face validity, but also a general desire to see more specific discussion around the face validity of the use of HCC's in this population.

**2b2. Score/Analysis:** H-6; M-3; L-0; I-0

**TAP Discussion:** The face validity of HCC's was found to be clear, but the logic behind the age stratification was unclear.

**2b3. Exclusions:** H-6; M-3; L-0; I-0

**TAP Discussion:** The TAP had an in-depth discussion regarding measure exclusions. The measure developer explained that cardiovascular conditions are not specifically excluded, but are used in the risk adjustment model. Patients with COPD are excluded. Exclusions affect the denominator population over either year within the two-year criteria, which is similar to the HEDIS asthma measure. There was agreement that the exclusion of COPD (which resulted in 38% of the initial population being eliminated) seems appropriate, particularly in light of the age range increasing to 64. The TAP did express concern that excluding acute respiratory failure could exclude poorly managed asthma patients. However, NCQA noted that acute respiratory failure only accounted for 3% of the population, so it doesn't meet their 5% threshold of concern.

**2b4. Risk Adjustment:** H-7; M-2; L-0; I-0

**TAP Discussion:** The TAP believes the risk-adjustment strategy seems appropriate. Several strategies are tested by NCQA, and the same methodology is used for all of their measures. The developer stratifies the population by age and gender and uses HCC's to risk adjust the population.

**2b5. Identification of statistically significant/meaningful differences:** H-8; M-1; L-0; I-0

**TAP Discussion:** There was general agreement that the distribution of the scores' detail score was appropriate. There was concern regarding whether the measure score could differentiate statistically significant and clinically significant variation.

**2b6. Multiple data sources:** N/A

**2c. Stratification for disparities:** H-5; M-3; L-0; I-1

**TAP Discussion:** The TAP believes stratification is needed although the data isn't available at this time.

**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 be included in the NCQA reporting product. The Committee asked the developers to explain the measure's use of indirect (vs. direct) standardization in creating standardized prices.

**Usability:**

**3a. Measure performance results are publicly reported:** H-8; M-1; L-0; I-0

**TAP Discussion:** The TAP was satisfied that NCQA publically reports measure results and provides support to enable understanding of those results. Purchasers are using this information, along with NCQA quality measures, to improve value for their employees. Asthma is a bit more difficult because there is only one NCQA quality measure to associate with this cost measure, however there are more quality measures in the pipeline.

**3b. Measure results are meaningful/useful for public reporting and quality improvement:** H-6; M-3; L-0; I-0

**TAP Discussion:** The measure is straightforward and easy to interpret. NCQA uses standardized pricing tables, which are reviewed annually. Health plans are the main users for this data. However, purchasers and the large employers will also drive a need for this information. The TAP questioned how smaller businesses would implement this measure, and NCQA explained that they provide help through their annual conferences, webinar services and a dedicated webpage.

**3c. Data and results can be decomposed for transparency and understanding:** H-8; M-1; L-0; I-0

**TAP Discussion:** The TAP believes the methodology was transparent and appropriate.

**3d. Harmonized or justification for differences:** N/A

**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.

**4. Feasibility:**

**4a. Data elements routinely generated during care process:** H-9; M-0; L-0; I-0

**TAP Discussion:** The TAP agrees this subcriterion has been met; the data is a byproduct of care.

**4b. Data elements available electronically:** H-9; M-0; L-0; I-0

**TAP Discussion:** The TAP agrees this subcriterion has been met; the data is available electronically.

**4c. Susceptibility to inaccuracies/ unintended consequences identified:** H-7; M-2; L-0; I-0

**TAP Discussion:** There was agreement that NCQA did a sufficient job recognizing where the challenges with data inaccuracies are and have adequately addressed these challenges.

**4d. Data collection strategy can be implemented:** H-8; M-1; L-0; I-0

**TAP Discussion:** All the data submitted to NCQA must go through a certified auditor before it's reported to NCQA. As part of their annual analysis, NCQA reviews outliers, but currently the outliers are less than half a percent for 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)

**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.

**1. Importance to Measure and Report**

**1a.High Impact:** H-9; M-0; L-0; I-0

**TAP Discussion:** The TAP was in agreement that this is an important area of measurement.

**1b. Resource use/cost problems:** H-9; M-0; L-0; I-0

**TAP Discussion:** The TAP believes while there is variation in resource use was identified in other parts of the submission, the information submitted in the form for this item only discussed the variations in clinical care provided.

**1c. Purpose clearly described:** H-8; M-1; L-0; I-0

**TAP Discussion:** The TAP was concerned that the measure submission applied only to newly diagnosed patients. The developer clarified that it is supposed to apply to anyone with a diagnosis with COPD. Otherwise, the purpose of the measure is to evaluate the total cost of care for COPD patients within a 1 year timeframe was clear.

**1d. Resource use service categories consistent and representative:** H-9; M-0; L-0; I-0

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

**Overall Importance:** Y-14, N-0

**Committee Discussion:** The Steering Committee agreed the measure focused on an important area of healthcare.

**2. Scientific Acceptability of Measure Properties:**

**2a. Overall Reliability:** H-7; M-2; L-0; I-0

**2a1.Measure well defined and precisely specified:** H-9; M-0; L-0; I-0

**TAP Discussion:** The TAP believes the specifications provided are clear and precise. The developer provided clarification on age stratification for resource use categories indicating that they are based on utilization patterns in the data-set, not clinical factors.

**2a2. The results are repeatable:** H-8; M-1; L-0; I-0

**TAP Discussion:** A similar methodology was used for this measure as for NCQA measure #1560, the primary difference being in the selection of the population. The TAP was concerned about the multiple populations being studied including commercial, Medicare, and Medicaid, due to the age range (unlike Measure 1560, where the age range cut off at 64). There was also concern that NCQA did not distinguish the fee-for-service versus the beneficiaries in Medicare Advantage plans.

**2b. Overall Validity:** H-4; M-5; L-0; I-0



<p><b>2b1. Evidence is consistent with intent:</b> H-8; M-1; L-0; I-0  <b>TAP Discussion:</b> The TAP believes the measure is clearly defined; however, one of the challenges will be the fact that COPD has multiple co-morbidities, particularly when compared to asthma. It will therefore be difficult to know if you are actually measuring COPD resource use. Specifications should be explored on how to develop disease severity; however, this is difficult to do with administrative datasets.</p> <p><b>2b2.Score/Analysis:</b> H-6; M-3; L-0; I-0  <b>TAP Discussion:</b> The TAP believes that overall the validity testing was appropriate. Outliers are identified by tagging O/E ratios below .3 or above 3.</p> <p><b>2b3. Exclusions:</b> H-4; M-5; L-0; I-0  <b>TAP Discussion:</b> The TAP agrees the exclusions are well stated and are similar to the asthma measure.</p> <p><b>2b4. Risk Adjustment:</b> H-6; M-3; L-0; I-0  <b>TAP Discussion:</b> Cardiovascular disease maybe a major driver of the severity of COPD. The risk adjustment approach appears reasonable for the data available. The intent is to compare across populations.</p> <p><b>2b5. Identification of statistically significant/meaningful differences:</b> H-5; M-4; L-0; I-0  <b>TAP Discussion:</b> The TAP believes NCQA did a sufficient job presenting their data in a transparent manner.</p> <p><b>2b6. Multiple data sources:</b>  <b>TAP Discussion:</b> N/A (using all administrative data)</p> <p><b>2c. Stratification for disparities:</b> H-5; M-4; L-0; I-0  <b>TAP Discussion:</b> Examining differences in racial disparities for this data set is not yet possible, but there is stratification by gender. Race is not a required field for most provider systems and is usually unavailable except in the Medicare population.</p>
<p><b>Overall Scientifically Acceptable: Yes [Y-13; N-1 (Committee Vote)]</b>  <b>Overall Reliability:</b> H-11; M-3; L-0; I-0  <b>Overall Validity:</b> H-4; M-10; L-0; I-0  <b>Committee Discussion:</b> 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.</p>
<p><b>3. Usability:</b>  <b>3a. Measure performance results are publicly reported:</b> H-9; M-0; L-0; I-0  <b>TAP Discussion:</b> The TAP believes this subcriterion has been met as NCQA does extensive audits of their material on a regular basis.</p> <p><b>3b. Measure results are meaningful/useful for public reporting and quality improvement:</b> H-5; M-4; L-0; I-0  <b>TAP Discussion:</b> The TAP feels the results are usable and understandable.</p> <p><b>3c. Data and results can be decomposed for transparency and understanding:</b> H-6; M-3; L-0; I-0  <b>TAP Discussion:</b> The TAP feels this subcriterion has been met as NCQA does extensive audits of their material on a regular basis, and the measure can be deconstructed to facilitate transparency.</p> <p><b>3d. Harmonized or justification for differences:</b> N/A</p>
<p><b>Overall Usability:</b> H-7; M-7; L-0; I-0  <b>Committee Discussion:</b> 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.</p>
<p><b>4. Feasibility:</b>  <b>4a. Data elements routinely generated during care process:</b> H-9; M-0; L-0; I-0  <b>TAP Discussion:</b> The TAP believes this subcriterion has been met as data is a byproduct of care.</p> <p><b>4b. Data elements available electronically:</b> H-9; M-0; L-0; I-0  <b>TAP Discussion:</b> The TAP believes this subcriterion has been met; all data is available electronically.</p> <p><b>4c. Susceptibility to inaccuracies/ unintended consequences identified:</b> H-6; M-3; L-0; I-0  <b>TAP Discussion:</b> The TAP believes this subcriterion has been met.</p> <p><b>4d. Data collection strategy can be implemented:</b> H-8; M-1; L-0; I-0  <b>TAP Discussion:</b> The TAP believes this subcriterion has been met.</p>
<p><b>Overall Feasibility:</b> H-10; M-4; L-0; I-0  <b>Committee Discussion:</b> There were no new additional comments from the Steering Committee relating to feasibility of NCQA measures.</p>

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

**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. 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).

**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/OptumInsight, 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.

**1.Importance to Measure and Report**

**1a.High Impact:** H-8; M-0; L-0; I-0

**TAP Discussion:** The TAP agreed that pneumonia is a high impact and high cost area.

**1b. Resource use/cost problems:** H-8; M-0; L-0; I-0

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

**1c. Purpose clearly described:** H-8; M-0; L-0; I-0

**TAP Discussion:** The TAP feel the purpose and objective are clear.

**1d. Resource use service categories consistent and representative:** H-7; M-1; L-0; I-0

**TAP Discussion:** The TAP agrees the service categories are consistent and representative.

**Overall Importance:** Y-14, N-1

**Committee Discussion:** The Steering Committee deemed the measure to be important.

**2.Scientific Acceptability of Measure Properties:**

**2a. Overall Reliability:** H-3; M-3; L-0; I-1

**2a1.Measure well defined and precisely specified:** H-3; M-4; L-0; I-0

**TAP Discussion:** Several TAP members were uncomfortable with the lack of transparency in the risk adjustment specifications and felt that the severity weights, particularly for the elderly, were unclear. The panel also had a hard time identifying clean periods. There was a strong feeling that there should be some separation between community-acquired and healthcare-acquired pneumonia, as they represent very different clinical conditions.

**2a2. The results are repeatable:** H-6; M-1; L-0; I-0

**TAP Discussion:** The TAP had concerns regarding the fact that there is no way to ascertain how Ingenix/OptumInsight came up with the specific weights assigned to comorbidities.

**2b. Overall Validity:** H-0; M-7; L-0; I-0

**2b1. Evidence is consistent with intent:** H-4; M-3; L-0; I-0

**TAP Discussion:** The panel again asked for clarification regarding why the measure has different weighted scores for the elderly.

**2b2.Score/Analysis:** H-0; M-5; L-2; I-0

**TAP Discussion:** The TAP was concerned that they weren't provided with enough information to understand how Ingenix/OptumInsight assigned risk scores. Questions regarding how diagnostic descriptions leads to increased utilization were raised. The TAP remained doubtful as to whether all claims included in this measure should be considered attributable to one distinct population.

<p><b>2b3. Exclusions:</b> H-2; M-4; L-1; I-0  <b>TAP Discussion:</b> The TAP felt that more data around the impact of exclusions (e.g. sensitivity analysis) would be helpful. Ingenix/OptumInsight confirmed that there are no clinical exclusions from the measure, only cost exclusions.</p> <p><b>2b4. Risk Adjustment:</b> H-1; M-3; L-2; I-1  <b>TAP Discussion:</b> The TAP believed that the risk-adjustment methodology is not readily transparent. More information on how risk scores are assigned was requested from the developers.</p> <p><b>2b5. Identification of statistically significant/meaningful differences:</b> H-0; M-7; L-0; I-0  <b>TAP Discussion:</b> Data submitted does demonstrate variation in resource use. However, there was a general feeling that meaningfulness is questionable since types of pneumonia cannot be separated out.</p> <p><b>2b6. Multiple data sources:</b> N/A (using all administrative data)</p> <p><b>2c. Stratification for disparities:</b> H-2; M-5; L-0; I-0  <b>TAP Discussion:</b> Gender and age can be stratified, but race data is not available in administrative claims.</p>
<p><b>Overall Scientifically Acceptable:</b> Yes [Y-13; N-3 (Committee Vote)]  <b>Overall Reliability:</b> H-3; M-11; L-2; I-0  <b>Overall Validity:</b> H-1; M-13; L-2; I-0  <b>Committee Discussion:</b> The Steering Committee agreed that this measure would not be clinically relevant at the physician level 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. The Committee reiterated the TAP's concern that Ingenix/OptumInsight specified the measure for use in patients over 65 using commercial data to calibrate the model. Commercial patients over 65 are not representative of the general over 65 population.</p>
<p><b>Usability:</b></p> <p><b>3a. Measure performance results are publicly reported:</b> H-0; M-6; L-1; I-0  <b>TAP Discussion:</b> The TAP agrees that despite the fact that multiple care organizations are currently using this measure, the inability to distinguishing between types of pneumonia severely limits the usability of the measure. They concurred that for individual organizations this limitation might be acceptable, but the measure wouldn't be useful as a national consensus standard.</p> <p><b>3b. Measure results are meaningful/useful for public reporting and quality improvement:</b> H-1; M-5; L-1; I-0  <b>TAP Discussion:</b> The TAP agrees that this subcriterion has been met.</p> <p><b>3c. Data and results can be decomposed for transparency and understanding:</b> H-1; M-5; L-1; I-0  <b>TAP Discussion:</b> The TAP feels the measure would be more transparent if more user-friendly detail were provided.</p> <p><b>3d. Harmonized or justification for differences:</b> N/A</p>
<p><b>Overall Usability:</b> H-3; M-11; L-1; I-1  <b>Committee Discussion:</b> There were no additional concerns identified by the Steering Committee for this criterion.</p>
<p><b>4. Feasibility:</b></p> <p><b>4a. Data elements routinely generated during care process:</b> H-7; M-0; L-0; I-0  <b>TAP Discussion:</b> The TAP believes this subcriterion has been met; data is a byproduct of care.</p> <p><b>4b. Data elements available electronically:</b> H-7; M-0; L-0; I-0  <b>TAP Discussion:</b> The TAP believes this subcriterion has been met; data available electronically.</p> <p><b>4c. Susceptibility to inaccuracies/ unintended consequences identified:</b> H-1; M-5; L-0; I-1  <b>TAP Discussion:</b> The TAP concluded there was a lack of information in the submission regarding data cleaning and missing data to sufficiently understand those areas.</p> <p><b>4d. Data collection strategy can be implemented:</b> H-5; M-2; L-0; I-0  <b>TAP Discussion:</b> The TAP agrees this subcriterion has been met.</p>
<p><b>Overall Feasibility:</b> H-1; M-8; L-7; I-0  <b>Committee Discussion:</b> See Ingenix/OptumInsight feasibility discussion in the discussion of recurring measure evaluation themes below.</p>

### 1598 Total Resource Use Population-based PMPM Index

**Description:** Total cost of care reflects a mix of complicated factors such as patient illness burden, service utilization, and negotiated prices. Separating out and reporting the resource use index along with the total cost of care index provides a more complete picture of population-based drivers of health care costs. Total Cost Index (TCI) is a measure of a primary care provider's risk-adjusted cost effectiveness at managing the population for which they care for. TCI includes all costs associated with treating members, including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary, and behavioral health services. The Resource Use Index (RUI) is an underlying risk-adjusted measure of the frequency and intensity of services utilized to manage a provider group's patients. Resource use includes all resources associated with treating members, including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary, and behavioral health services.

**Resource Use Type:** Per capita (population- or patient-based)

**Data Type:** Administrative claims, other

**Resource Use Service Category:** 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; Inpatient services: Labor (hours, FTE, etc.); 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; Ambulatory services: Labor (hours, FTE, etc.); Durable Medical Equipment (DME)

**Care Setting:** Ambulatory Care: Ambulatory Surgery Center (ASC); Ambulatory Care: Clinic/Urgent Care; Ambulatory Care: Clinician Office; Behavioral Health/Psychiatric: Inpatient; Behavioral Health/Psychiatric: Outpatient; Dialysis Facility

**Level of Analysis:** Clinician: Group/Practice; Population: Community

**Measure Developer:** HealthPartners

**Committee Recommendation for Endorsement:** Recommended for Endorsement: Y-11 ; N-6

#### **Committee Questions for Developer:**

1. The measure's resource use index relies on total care relative resource use categories, which are constructed so they are additive across various sites of care and then add in pharmacy data. How was this done?
2. Are the data distorted due to billed charges?
3. What is the attributable population in this measure?
4. How are variables in geographic location accounted for?
5. This measure is restricted to commercial, under -65 -years -of -age population. Is there anything that prohibits its use in the Medicare population?
6. Do users have to use the ACG software for risk-adjustment?

#### **Developer Responses:**

1. Health Partners relies on sector- specific relative value units, the billed charges across the sectors of care are used to build relativity. The payments are then appropriately adjusted. Final quality checks for thresholds are then performed. This method will eventually be patented and shared with the community.
2. The measure uses billed charges controls for confounding variables. The measure uses the billed amount to allow for the claims (the most standard piece of information), then goes across the different components and applies the discount rate. The adjustment factor is for the paid/billed ratios.
3. The attributable populations (which are scalable to different units of analysis) are PPO and HMO. Look at practice specialty of physician and claims history and attribute patients to the clinic with the majority of visits.
4. Depending on the application and the user, the measure can be flexible and usable across different locations. In the market there are multiple hospitals with different price points. Cost points may be consistent; however, the price they charge may be different. Actual paid (allowable inclusive liability) amount is used in the measure; the billed amount is used only to gauge the relativity (e.g., inpatient to outpatient services).
5. HealthPartners is a largely commercial- based health plan, so they do not have access to Medicare data. Theoretically, if these claims were available in the database, one would be able to use it.
6. Users are not required to use the ACG software for risk-adjustment. Any risk- adjustment methodology may be used, as long as all methods are comparable (see Society of Actuaries report). Health Partners has a history of working with ACG software and have tested the measure using the ACG risk adjuster. They have specified the measure to be used at the group level with the risk- adjustment methodology developed by Johns Hopkins, and if it is NQF-endorsed, it would be endorsed only at the group level for use with this specific software.

#### **Committee Conditions:**

1. The Committee determined there were actually two measures of cost described within the measure submission as presented: resource use index and a total cost index. There was some discussion about which should be evaluated for the purposes of this project or whether the measures should be considered as a pair. Because this project is not accepting paired measures, the Committee has agreed to evaluate the resource use index, which appears to be most

applicable to the goals of this project at this time.

**Developer Response:**

1. The measure calculations for costing within the measure may be used independently; however, they are better used in partnership with one another. The developer agreed to separate the specifications and resubmit a separate measure for total cost (#1604).

**1.Importance to Measure and Report**

**1a.High Impact:** H-15, M-2, L-1, I-0, N/A-0

*Committee Discussion:* This measure is considered highly important and relates to NPP/national goals.

**1b. Resource use/cost problems:** H-13, M-3, L-0, I-1, N/A-1

*Committee Discussion:* This measure does not explain much as an isolated measure. However, it does inform providers of areas where there is overuse or underuse; given the fact of that overuse and waste is an issue, there is a place for this in the resource use project.

**1c. Purpose clearly described:** H-12, M-5, L-1, I-0, N/A-0

*Committee Discussion:* This criterion has been met because the measure is targeting an area known to have variation, and relevant service categories, and the objective has been clearly described.

**1d. Resource use service categories consistent and representative:** H-12, M-6, L-0, I-0, N/A-0

*Committee Discussion:* This criterion has been met. The supporting information provided by the measure developer also helps to demonstrate this.

**1.Overall Importance:** Y- 18, N-0

**2.Scientific Acceptability of Measure Properties:**

**2a. Reliability:**

**2a1. Well- defined/precise specifications:** H-5, M-8, L-4, I-0, N/A-0

*Committee Discussion:* HealthPartners (HP) uses regional and national data; there is a great deal of actionable data at this level. It may be difficult to be implemented in other systems. Since this is a population measure, it is missing whether or not people are described on an individual basis and then tied to a region, making it difficult to determine whether or not it was appropriately specified. The total eligible individuals may only have pharmacy claims or are not using any services; however, this may vary across systems. This measure is intended for a commercial population; non-users would not be attributed. The patient has to be a user of primary care services to be included; attribution (prospective and retrospective) is at the physician group level (with 2 or more physicians). The peer groups are based on the group to which the physician belongs to. The measure has been tested on groups that have at least 600 patients at the group practice level. High claims data are included and truncated after a certain threshold, resulting in roughly 5-8% excluded. These individuals are excluded based on the published guidelines by Society of Actuaries. The pharmacy relative values come from using the average billed amount, and the paid amount is defined as the paid-to-billed ratio.

**2a2. Reliability testing:** H-10, M-6, L-0, I-1, N/A-0

*Committee Discussion:* Assumption that clinical and administrative claims data is accurate from a coding perspective, which is true for the majority of resource measures. For claims data, the hospital-based claims take more time to process than professional claims, so time frames need to be taken into account when applying them to this measure. The measure developer informed the Committee that the timeline of 3 months is specified; all claims are electronic and therefore arrive quickly into the system. The Committee believes the reliability matrix is acceptable. Health Partners did a very good job examining the reliability of the data using its commercial database. They performed two types of sampling; the first was a 90% sample of the actual values. It selected one patient at a time until they it reached 90%; this gives an idea of the influence of extreme values. Health Partners selected 90% of the data 500 times and compared the results obtained from the averages to the entire sample; the results showed there is represent very small change. The difference between the samples is only 0.9%, so that demonstrating reliability and that the potential influence of these extreme values is small. The other approach used was a boot strapping technique, which is similar; but instead of a 90% sample, however, the developers selected a sample with replacement, this simulates the reliability and is a very common methodology. The developers found a very small range of change in the sample population; this has some variability in respect to the sample. It's important to note that NQF does not require developers use a certain type of methodology. The analysis has been done at the provider level and depicts the measure to be reliable.

**2b. Validity:**

**2b1. Specifications consistent with resource use/cost problem:**

*Committee Discussion:* This section appeared to be sufficient and meets the criterion. This measure excludes patients who have not had a primary care visit; however, within the system this may be giving all the information needed to feed back to providers on how they are using services.

**2b2. Validity testing:** H-5, M-8, L-2, I-1, N/A-0

*Committee Discussion:* Adequate sample size, large area, 19 providers across approximately 200 hospitals. Health Partners (HP) has nearly 7,000 members who are Medicare/Medicaid recipients. HP has about 700,000 total members within the marketplace area (including CMS data/commercial data), and the non-user rate is around 9%. Roughly 50% of the data presented in the validity sample comes from commercial data. Because this measure has only been tested only in a commercial population, it will be NQF endorsed only

in a commercial population. Peer group averaging can serve as a benchmark, if that is a sufficient measure in all markets. Within a commercial network and scheme, it may work; however, how these will be used it is not clear how these will be utilized. The validity was obtained in terms of the risk -adjusted and the non-risk adjusted values. One would anticipate the values between expected and observed would be close - values of 0.98 for non-risk adjusted to actual money spent. After the measure risk-adjustment was applied, this correlation went down to 0.215. When the correlation is restricted to different places, they look at the correlation between total resource use to the risk adjusted methods. There were a number of test performed and they show the direction of the correlation, which was high in this case.

**2b3. Exclusions:** H-6, M-8, L-1, I-2, N/A-0

**Committee Discussion:** This measure excludes sub-populations that haven't had primary care visits. The measure also excludes "never users" and "super users" by truncating them out. The group-oriented market may exclude those outside the group. HP has not seen this as a problem, as there is a low non-user rate. The bulk of members are attributed in this model through primary care, a smaller percentage only see a specialist. Those who are over the age of 65 are excluded.

**2b4. Risk-adjustment:** H-5, M-9, L-0, I-3, N/A-0

**Committee Discussion:** Health Partners uses the 9.0 version of the ACG risk-adjustment method developed by Johns Hopkins, the most recent 9.0 version and they HP has a long-standing market history of using this product. HP relied heavily on a study conducted by the Society of Actuaries that concluded a number of commercially available risk-adjustment methodologies are satisfactory for this purpose. The risk-adjustment was tested and demonstrated to be effective. It is significant to note for consumers that a user ID and password is necessary to access the site. The Johns Hopkins software is proprietary; however, Hopkins has recently announced the software to be free of charge to health insurance exchanges. For the ordinary user, the software is available for a fee based on a scale from large to small organizations, non-profits, etc. CMS offers an open -source risk -adjustment tool, the Hierarchical Condition Categories (HCCs).

**2b5. Identification of statistically significant/meaningful differences:** H-7, M-6, L-1, I-3, N/A-0

**Committee Discussion:** The Committee believes that this sub criterion has been met.

**2b6. Multiple data sources:** N/A

**Committee Discussion:** N/A

**2c. Stratification for disparities:** N/A

**Committee Discussion:** N/A

**2. Overall Scientifically Acceptable: Yes [Y-13, N-4 (Committee Vote)]**

**3. Usability:**

**3a. Measure performance results are publicly reported:** H-5, M-5, L-2, I-2, N/A-2

**Committee Discussion:** The data is publicly reported, but it's difficult to find on the Health Partners website. Currently the measure is used for benefit design and transparency; there are plans for community collaborations in the future.

**3b. Measure results are meaningful/useful for accountability and quality improvement:** H-3, M-9, L-3, I-1, N/A-2

**Committee Discussion:** The Committee discussed the issue that publicly reported measures may not have the same value for quality improvement. This measure is being reported out to the public at large, as well as to members of Health Partners, and has been for quite some time. During the three-year NQF maintenance review this criterion would be looked at even further to see how the measure has progressed. This is a fairly complicated measure for the public, in that the methodology may not be fully understandable to the average person. It must be communicated that more resource use does not necessarily mean *better* service. For resource use, it may be up to those producing the consumer reports on may need to be the ones determining how to present it to the public in the most understandable way.

**3c. Data and results can be decomposed for transparency and understanding:** H-3, M-8, L-4, I-2

**Committee Discussion:** On the Health Partners website, they have converted the results to dollar signs. This calculation is available to the public at large. There have also been focus groups conducted in order to gauge the clarity of the information available online. It may be difficult to decipher differences in providers and resource use; at some point there is the issue of hierarchical modeling and how to devise low -volume providers by evaluating the measure itself occurs. To some extent, the issue is raised are whether the measure is useful to the public because it does not explain the quality of care or outcome relating to resource use.

**3d. Harmonized or justification for differences:** N/A

**Committee Discussion:** N/A

**3. Overall Usability: H-3, M-11, L-2, I-1, N/A-0**

**4. Feasibility:**

**4a. Data elements routinely generated during care process:** H-9, M-1, L-3, I-0, N/A-4

**Committee Discussion:** This measure is based on data that is generated as a byproduct of care. The Committee believes this criterion has been met.

**4b. Data elements available electronically:** H-13, M-0, L-1, I-3, N/A-0

**Committee Discussion:** These measures are all available via electronic sources. The Committee believes this criterion has been met.

**4c. Susceptibility to inaccuracies/ unintended consequences identified:** H-4, M-8, L-3, I-2, N/A-0

**Committee Discussion:** This measure has met the criteria for inaccuracies and unintended consequences. Third-party administrators can work together to match up their coding; this would not be a barrier for these measures. There is a great deal of regulatory variation that can be applied to self-insured entities, and runs the risk of measuring smaller percentages of practices.

**4d. Data collection strategy can be implemented:** H-5, M-5, L-4, I-2, N/A-0

**Committee Discussion:** The Committee believes this sub criterion has been met.

**4. Overall Feasibility:** H-7, M-7, L-1, I-1, N/A-0

### 1604 Total Cost of Care Population-Based PMPM Index

**Description:** Total Cost of Care reflects a mix of complicated factors such as patient illness burden, service utilization, and negotiated prices. Total Cost Index (TCI) is a measure of a primary care provider's risk-adjusted cost effectiveness at managing the population they care for. TCI includes all costs associated with treating members, including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary, and behavioral health services. A Total Cost of Care Index when viewed together with a resource use measure provides a more complete picture of population-based drivers of healthcare costs.

**Resource Use Type:** Cost/resource use

**Data Type:** Administrative claims

**Resource Use Service Category:** 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; Inpatient services: Labor (hours, FTE, etc.); 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; Ambulatory services: Labor (hours, FTE, etc.); Durable Medical Equipment (DME)

**Care Setting:** Ambulatory Care: Ambulatory Surgery Center (ASC); Ambulatory Care: Clinic/Urgent Care; Ambulatory Care: Clinician Office; Behavioral Health/Psychiatric: Inpatient; Behavioral Health/Psychiatric: Outpatient; Dialysis Facility; Emergency Medical Services/Ambulance; Home Health; Hospice; Hospital/Acute Care Facility; Imaging Facility; Laboratory; Pharmacy; Post-Acute/Long Term Care Facility; Nursing Home/Skilled Nursing Facility; Post-Acute/Long Term Care Facility; Rehabilitation

**Level of Analysis:** Clinician: Group/Practice; Population: Community

**Measure Developer:** HealthPartners

**Committee Recommendation for Endorsement:** Recommended for Endorsement: Y-9 ; N-8, Abstain- 1

#### Committee Conditions/Questions for Developer:

1. What tools are used to collect patient satisfaction information?
2. In this measure it appears the total cost measure is reduced to an index and then compared to a peer group. Is it correct that any variations in input costs should be factored into that peer group comparison?
3. How are regional comparisons made between regions with very different cost/payment structures?
4. Are the actual prices based on what the plan has paid or what has been billed?
5. Have you tested this measure within a system that uses behavioral or pharmacy carve-outs?
6. When the costs per member per month (PMPM) are calculated, is this the average premium they are paying for the carve-out for every member in the group specific, or is it adjusted to reflect it?
7. What is the numerator for this measure?
8. Is this measure only valid only for comparing costs within the same well-defined population?
9. How does the use of the attribution guideline impact the calculation of the total cost index?

#### Developer Response:

1. HealthPartners historically used a health plan -specific survey, but in the Minnesota community, they use Minnesota Community Measurement in the Minnesota community.
2. Benchmarking is done based on the plan average, so the variation for a health plan, for example, would be among the groups within the plan.
3. Comparisons between regions would be based on the ability to access an adequate data set, the type of attribution model that has been used employed by the measure user of the measure, and the business application of the measure (e.g., use by consumers or internal benchmarking).
4. The measure counts what the plan is paying, plus the member liability (i.e., member co-pay).
5. Medical and pharmacy PMPM costs are calculated separately and then added together. However, in the HealthPartners system there are no carve-outs for behavioral health. For systems that do have behavioral and pharmacy carve-outs, it is recommended that the user is consistent in how these data are cleaned and used in the measure.
6. For pharmacy costs, for example, the numerator would be the plan and plus the member co-pay, with the denominator being only those with the pharmacy benefit, thus accounting for the carve-out. Members impacted affected by the carve-out are not left out of the measure, but are examined separately with medical and behavioral together. They are accounted for at the aggregate level.
7. Total costs for patients in the group (100% of services), regardless of attribution rules.
8. It is possible to compare across these groups, but the measure would be used to show a cost differential. The user would then have the option of using a geographic adjuster to account for these differences in business applications.
9. Attribution does not impact the calculation of the index.

#### 1.Importance to Measure and Report

**1a.High Impact:** H-15, M-2, L-1, I-0, N/A-0

**Committee Discussion:** The Committee agrees this criterion was adequately met.



**1b. Resource use/cost problems:** H-14, M-4, L-0, I-0, N/A-0

*Committee Discussion:* The Committee agrees this criterion was adequately met.

**1c. Purpose clearly described:** H-11, M-7, L-0, I-0, N/A-0

*Committee Discussion:* The Committee agrees this criterion was adequately met.

**1d. Resource use service categories consistent and representative:** H-11, M-7, L-0, I-0, N/A-0

*Committee Discussion:* The Committee agrees this criterion was adequately met.

**1. Overall Importance:** Y- 18, N-0

**2. Scientific Acceptability of Measure Properties:**

**2a. Reliability:**

**2a1. Well- defined/precise specifications:** H-5, M-8, L-4, I-1, N/A-0

*Committee Discussion:* There was concern that whether the total cost PMPM measurement for a health plan is useful, because it does not use standardized prices, it does not seem to be generalizable to different populations outside of the geographical region in which it is used. While geographic adjusters are available for helping to address regional differences, it should not be up to the user to figure this out along with the many other factors that contribute to the PMPM resource use/costs in a community. There was disagreement among the Committee on whether the lack of nationally comparability and potential limited use for this measure conflicts with the intent of endorsement. While some believe endorsed measures should be generalizable for various regions and markets, others believe it is useful and acceptable to have a measure endorsed for use within the context of a region for comparisons. There are some systems, health plans, and consumers that are interested in knowing actual costs. For example, there are many health systems are looking for this type of measure; particularly in California, for Medicare and commercial population ACO's, actual costs for total cost of care are of great interest. This measure provides real economic information that resource use measures that use standardized prices do not give information that will guide people's choices. If, for example, from an ACO's perspective, adjusting is undesirable, the actual total cost to the system is of interest for accountability purposes. The Committee and developers also acknowledged that all endorsed measures are not useful for every region and population.

**2a2. Reliability testing:** H-10, M-4, L-4, I-0, N/A-0

*Committee Discussion:* An analysis of the reliability testing was conducted by the NOF statistical consultant and shared with the Committee. His analysis was based on bootstrapping simulations restricted to each provider group; this was done three times in each year of data for each provider group. They used a variation simulation and compared its results to the observed variability to measure the signal-to-noise ratio. In addition, they compared how the ratios changed over time by provider, demonstrating insignificant differences. The reliability testing was deemed accepted and demonstrated a high level of measure score reliability.

**Overall Reliability:** H-8, M-6, L-4, I-0, N/A-0

**2b. Validity:**

**2b1. Specifications consistent with resource use/cost problem:** H-4, M-5, L-9, I-0, N/A-0

*Committee Discussion:* Committee members expressed a great deal of concern about the primary care attribution guideline submitted for this measure. Attribution instructions could be submitted as well thought- out guidelines, allowing for user flexibility to use the method outlined, or another method that suits the user's specific application while still enabling the use of the core measure specifications that have been validated. Developers also had the option of submitting attribution instructions as specifications, which require the user to apply the method specified in order to fully implement the measure fully. The attribution approach for this measure was submitted as guidelines. Within the context of these attribution guidelines, there were concerns with the inclusion of inpatient costs to the total cost, but the attribution model attributes based on outpatient resource use. For example, a doctor could be held responsible for a patient's inpatient stay before ever seeing the patient in an outpatient visit. There were concerns about how the use of this type of model might affect practice and potentially incentivize providers not to take on new patients who haven't have not seen a PCP. Another concern with the attribution guideline is accounting for care managed primarily by a specialist, since the guideline attributes to primary care providers (PCPs). Within the HealthPartners system 75% of its users use PCPs; this is not the case for many other areas in the country. Finally, within this attribution approach, non-users of the system are not attributed. This measure can be used in conjunction with measure 1598, which is specified in the exact same manner but uses standardized pricing. When used together the difference between the actual and standardized prices can be used to reflect differences in regional pricing.

Secondary to the concerns around the attribution guideline, is the level of analysis, which includes the physician group level. A physician group is defined by the developer as 2 or more physicians, with a recommended minimum of 600 patients in the sample.

The Committee voted on this criterion with the understanding that the attribution approach was submitted as a guideline.

**2b2. Validity testing:** H-7, M-5, L-5, I-0, N/A-0

*Committee Discussion:* The NOF Statistical Consultant conducted an analysis of the validity testing and shared it with the Committee. The validity testing sought to demonstrate face validity. Testing was conducted on provider groups, not for individual providers. As previously mentioned, the recommended minimum sample size is 600 patients. The Committee There expressed concern expressed about how would this measure operate for groups with only 2 -3 physicians.

**2b3. Exclusions:** H-3, M-6, L-9; I-0

*Committee Discussion:* Patients that who do not have a PCP are excluded from the denominator. The Committee expressed concern

with this exclusion, as members who seek care from a specialist may be using resources within the system, but those resources are not counted in the total cost. This brings concerns that there may be potential for “gaming the system” using this measure—a system’s total cost may appear lower if most of its care is provided by specialist. The issue of pharmacy carve-outs and how they are handled in this measure were also discussed relevant to this criterion.

**2b4. Risk-adjustment:** H-7, M-7, L-2, I-2, N/A-0

**Committee Discussion:** This measure uses ACG’s to risk adjust. It is a widely known and accepted methodology developed and maintained by a John’s Hopkins group. The use of the ACG risk adjuster is open to the public for a fee based on the type of user. Fees associated with the using of the adjuster are discussed below in Feasibility criterion 4d. Adjustment in the underlying populations also has also been applied. The NQF Statistical Consultant conducted an analysis of the risk -adjustment model was conducted by the NQF statistical consultant and shared it with the Committee. The risk- adjustment model was included in a correlation analysis with the physician total cost index (TCI) and the observed actual costs, and which demonstrated that the risk -adjustment model adequately accounts for variation, lowering the correlation between the TCI and actual costs.

**2b5. Identification of statistically significant/meaningful differences:** H-7, M-5, L -2, I-4

**Committee Discussion:** Most Committee members agreed the measure adequately demonstrated this criterion. Others believed that given the concerns with the exclusions, focus on primary care encounters, validity testing at the group level only, and comparisons across regions, the ability to determine statistically significant differences is unclear.

**2b6. Multiple data sources:** N/A

**Committee Discussion:** N/A

**Overall Validity:** H-4, M-6, L-7, I-0, N/A-0

**2c. Stratification for disparities:** H-1, M-8, L-3, I-7, N/A-0

**Committee Discussion:** Due to the limitations in the administrative claims data to capture race and ethnicity, it is difficult to assess how they might be accounted for in the measure. However, if the data were available, the Committee agrees the measure is constructed such that it would be able to report stratified data. The HealthPartners system does collect race and language information and is working on eliminating disparities in its system; however, this measure has not been stratified to report on disparities at this time.

**2.Overall Scientifically Acceptable: Yes [Y-9, N-10 (Committee Vote)]**

**3. Usability:**

**3a. Measure performance results are publicly reported:** H-9, M-7, L-0, I-0, N/A-0

**Committee Discussion:** Measure is currently in use in the Minnesota region.

**3b. Measure results are meaningful/useful for accountability and quality improvement:** H-4, M-8, L-4, I-0, N/A-0

**Committee Discussion:** The Committee’s discussion of the generalizability and comparability of the measure geographically and across varied patient populations also applies to the utility of this type of data for accountability and for the intended audiences. See discussion in 2a1, 2b1, and 2b2.

**3c. Data and results can be decomposed for transparency and understanding:** H-7, M-6, L-3, I-0, N/A-0

**Committee Discussion:** Behavioral health and pharmacy carve- outs are a concern. Comparisons should not be made between entities with carve- outs and those without.

**3d. Harmonized or justification for differences:** N/A

**Discussion:** N/A

**3.Overall Usability: H-6, M-7, L-2, I-0, N/A-0**

**4. Feasibility:**

**4a. Data elements routinely generated during care process:** H-11, M-7, L-0, I-0, N/A-0

**Committee Discussion:** The Committee agreed this criterion has been adequately demonstrated as this measure uses administrative claims data, which are generated as a byproduct of care delivery.

**4b. Data elements available electronically:** H-11, M-6, L-1, I-0, N/A-0

**Committee Discussion:** The Committee agreed this criterion has been adequately demonstrated, as this measure uses administrative claims data, which are available electronically. Due to the issue of carve-outs, however, not all data are available electronically (i.e; pharmacy data).

**4c. Susceptibility to inaccuracies/ unintended consequences identified:** H-4, M-6, L-8, I-0, N/A-0

**Committee Discussion:** The committee suggested a title change to indicate this measure should only be used for measuring costs in the in primary care setting. Setting the threshold of a visit with the PCP should be more than 1 visit (HP responded nonusers can be brought into play at the health plan level).

**4d. Data collection strategy can be implemented:** H-0, M-13, L-3, I-0, N/A-0

**Committee Discussion:** Consideration of pricing table. Carve- outs an issue

**4. Overall Feasibility: H-3, M-8, L-7, I-0, N/A-0**

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

**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/OptumInsight

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

**1.Importance to Measure and Report –**

**1a.High Impact:** H-6; M-1; L-0; I-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

**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

**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-0

**TAP Discussion:** The TAP felt that the resource use service categories were appropriate.

**Overall Importance:** Y-17, N-0

**Committee Discussion:** The Steering Committee deemed this measure to be important.

**2.Scientific Acceptability of Measure Properties:**

**2a. Reliability:**

**2a1.Measure well defined and precisely specified:** H-0; M-3; L-4; I-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. 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

**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

**TAP Discussion:**

**2b. Validity**

**2b1. Evidence is consistent with intent:** H-2; M-4; L-1; I-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

**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 an individual

<p>provider. These attribution approaches were submitted as guidelines only.</p> <p><b>2b3. Exclusions:</b> H-0; M-2; L-4; I-1  <b>TAP Discussion:</b> The TAP wanted more information on why low cost outliers were excluded and high cost outliers were winsorized; 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 winsorized. Information about the high cost outliers might actually drive targeted interventions.</p> <p><b>2b4. Risk Adjustment:</b> H-0; M-0; L-6; I-1  <b>TAP Discussion:</b> 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.</p> <p><b>2b5. Identification of statistically significant/meaningful differences:</b>  <b>TAP Discussion:</b> There was general agreement that the complexities of the score may make it difficult to discern meaningful differences between providers.</p> <p><b>2b6. Multiple data sources:</b> N/A  <b>Overall Validity:</b> H-0; M-1; L-5; I-0</p> <p><b>2c. Stratification for disparities:</b> H-1; M-0; L-4; I-2  <b>TAP Discussion:</b> Administrative data is limited in its ability to stratify based on race.</p>
<p><b>Overall Scientifically Acceptable:</b> Yes [Y-11; N-5 (Committee Vote)]  <b>Overall Reliability:</b> H-2; M-14; L-0; I-0  <b>Overall Validity:</b> H-1; M-9; L-6; I-0  <b>Committee Discussion:</b> The Steering Committee was concerned with the lack of specification regarding the measure's use of MSDRG's in the risk-adjustment methodology. Ingenix/OptumInsight 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.</p>
<p><b>3. Usability:</b></p> <p><b>3a. Measure performance results are publicly reported:</b> H-0; M-5; L-2; I-0  <b>TAP Discussion:</b> The TAP was concerned that this ETG was not currently being used as a stand-alone measure and it was unclear if it was currently being publicly reported.</p> <p><b>3b. Measure results are meaningful/useful for public reporting and quality improvement:</b> H-0; M-4; L-3; I-0  <b>TAP Discussion:</b> The TAP was concerned that this ETG was not currently being used as a stand-alone measure which may impact the need for public reporting.</p> <p><b>3c. Data and results can be decomposed for transparency and understanding:</b> H-0; M-3; L-4; I-0  <b>TAP Discussion:</b> 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.</p> <p><b>3d. Harmonized or justification for differences:</b> N/A</p>
<p><b>Overall Usability:</b> H-0; M-12; L-4; I-1  <b>Committee Discussion:</b> 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/OptumInsight 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.</p>
<p><b>4. Feasibility:</b></p> <p><b>4a. Data elements routinely generated during care process:</b> H-5; M-2; L-0; I-0  <b>TAP Discussion:</b> The TAP believes this subcriterion has been met; data is a byproduct of care.</p> <p><b>4b. Data elements available electronically:</b> H-6; M-1; L-0; I-0  <b>TAP Discussion:</b> The TAP believes this subcriterion has been met; data elements that are available electronically.</p> <p><b>4c. Susceptibility to inaccuracies/ unintended consequences identified:</b> H-0; M-3; L-4; I-0  <b>TAP Discussion:</b> The TAP agrees that much of this surgery is dependent on patient preferences thus the measure should account for these preferences in inclusion and exclusion criteria of the measure. Additionally, providers who treat their patients conservatively can appear to be high cost users since the only patients who get surgery are those who are more severe.</p> <p><b>4d. Data collection strategy can be implemented:</b> H-1; M-5; L-1; I-0  <b>TAP Discussion:</b> No additional issues were raised by the TAP.</p>
<p><b>Overall Feasibility:</b> H-1; M-8; L-7; I-0  <b>Committee Discussion:</b> See Ingenix/OptumInsight feasibility discussion in the discussion of recurring measure evaluation themes below.</p>

## **CANDIDATE CONSENSUS STANDARDS NOT RECOMMENDED FOR ENDORSEMENT**

Seven candidate consensus standards were not recommended for endorsement. The evaluation summary tables follow the list of measures and summarize the results of the TAP’s and Committee’s evaluation of and voting on the candidate consensus standards not recommended for endorsement. Hyperlinks are provided from each summary table to the detailed measure specifications. To access the meeting transcripts and recordings in which these measures are discussed, refer to the [project web page](#).

### ***Cardiovascular***

(1591) ETG-based congestive heart failure (CHF) cost of care (Ingenix/OptumInsight).....29

(1594) ETG-based coronary artery disease (CAD) cost of care (Ingenix/OptumInsight).....32

### ***Diabetes***

(1595) ETG based diabetes cost of care (Ingenix/OptumInsight).....35

### ***Non-Condition Specific***

(1599) ETG-based non-condition specific cost of care (Ingenix/OptumInsight) .....38

### ***Bone/Joint***

(1603) ETG-based hip fracture cost of care (Ingenix/OptumInsight) .....41

### ***Pulmonary***

(1605) ETG-based asthma cost of care (Ingenix/OptumInsight) .....44

(1608) ETG-based chronic obstructive pulmonary disease (COPD) cost of care  
(Ingenix/OptumInsight) .....47

**1591: ETG Based Congestive Heart Failure (CHF) cost of care measure (Ingenix/OptumInsight)**

**Description:** 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. 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).

**Resource Use Type:** Per episode

**Data Type:** Administrative claims, other

**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

**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 : states

**Measure Developer:** Ingenix/OptumInsight

**Committee Recommendation for Endorsement: Y-6; N-8; Abstain-0 (re-vote) [Y-10; N-8; Abstain-0 (initial vote)]**

**Conditions/Questions for Developer:**

1. Why are some of the codes, typically seen in congestive heart failure measures, excluded?
2. How are hospitalizations that occur during the course of the measure handled?
3. Does the episode include events that occur before and/or after the episode?

**Developer Response:**

1. Ingenix/OptumInsight 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/OptumInsight includes codes that were both systolic and diastolic, and used them as a marker to increase the severity score for the episode.
2. 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.
3. No, this measure is insulated from events that occur before or after the episode.

**1. Importance to Measure and Report**

**1a.High Impact: H -8; M-0; L-0; I-0**

**TAP Discussion:** The TAP believes this is a high impact, high cost area that is important to measure and report.

**1b. Resource use/cost problems: H -8; M-0; L-0; I-0**

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

**1c. Purpose clearly described: H -5; M-3; L-0; I-0**

**TAP Discussion:** The TAP believes the purpose of the measure is clearly described.

**1d. Resource use service categories consistent and representative: H -7; M-1; L-0;I-0**

**TAP Discussion:** The TAP believes this subcriterion has been met; the resource use service categories are consistent and representative of the measure.

**Overall Importance: Yes [Y-17; N-1 (Committee Vote)]**

**Committee Discussion:** The Steering Committee believes this is a high impact, high cost area and that the measure has been clearly described. This criterion has been met.

## 2. Scientific Acceptability of Measure Properties:

### 2a. Reliability:

#### 2a1. Well defined/precise specifications: H -3; M-4; L-0; I-1

**TAP Discussion:** The TAP believed 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 there needs to be more clarification around the use of this term. The TAP agrees that this measure is targeting systolic heart failure and then a mix of systolic/diastolic heart failure. Ingenix/OptumInsight 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 users will be able to implement the measure as described.

#### 2a2. Reliability testing: H -7; M-1;L-0;I-0

**TAP Discussion:** The TAP believes this measure has demonstrated extensive benchmarking and comparisons; however they would have liked to see more external comparisons. The testing data submitted was from nine health care organizations, all large commercial insurers that vary geographically. Ingenix/OptumInsight demonstrated reliability by performing parallel development of the data by using two independent approaches. These two different approaches led to the same results as levels near 99.9%The data was tested primarily on commercial databases, however some Part C plan Medicare patients were also included. It is important to note that this measure was submitted for use in the commercial, less than 65 years old population.

### 2b. Validity:

#### 2b1. Specifications consistent with resource use/cost problem: H -2;M-2; L-0; I-0

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

#### 2b2. Validity testing: H -4; M-4; L-0; I-0

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

#### 2b3. Exclusions: H -4; M-3; L-1; I-0

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

#### 2b4. Risk adjustment: H -4; M-2; L-0; I-1

**TAP 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/OptumInsight 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 variables selected for inclusion in the calibration of the risk model, the risk groups selected in terms of a cutoff for the severity score, and there is no rationale presented for why this cutoff point has been chosen.

#### 2b5. Identification of statistically significant/meaningful differences: H -2; M-1; L-3; I-1

**TAP Discussion:** The TAP believes there is little information to compare statistical versus practical significance for this measure. The measure allows the user to determine what is clinically significant based on 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/OptumInsight has created confidence intervals around the observed to expected ratio The minimum sample size to detect statistically significant differences depends upon the case mix of the providers and the variation in performance across providers..

#### 2b6. Multiple data sources: N/A

### 2c. Stratification for disparities: H-0; M-0; L-0; I-0; N/A-8

**TAP Discussion:** Due to the limitations in the administrative claims data, at this time the measure does not stratify for disparities.

**Overall Reliability:** H-3, M-12, L-2, I-0

**Overall Validity:** H-1, M-13, L-4, I-0

**Overall Scientific Acceptability:** Yes [Y-14; N-4(Committee Vote)]

**Committee Discussion:** The Steering Committee discussion focused on how clearly specified the codes used with the measure are, and how well they capture systolic heart failure. This is a measure of systolic heart failure, a paired measure of diastolic heart failure from Ingenix/OptumInsight exists but they did not submit it to the project. Because the Steering Committee could not take into account the existence of the diastolic measure, there was concern around the completeness and accuracy with which this measure would capture systolic heart failure. The diagnosis codes specified are limited to the 428 codes that used the word “systolic”, they do not use some of the 404s and 402s that the other measures have used to capture the larger heart failure population. The measure specifications have been in use for a significant amount of time; Ingenix/OptumInsight has demonstrated that if this measure is used in the same population, at the same time, then the result will be the same roughly 99.9% of the time. The Steering Committee discussed how there are carve outs for mental health & pharmacy data and therefore comparisons within the health plan are the same or likely to be the same. However, when comparing across health plans or across physician groups validity may become an issue when there are differences in the completeness of the data submitted. The Steering Committee expressed concerns over the reliability, validity and risk adjustment method. Specifically, that the measure may be adjusting for comorbidities identified during the measurement period as opposed to comorbidities identified prior to the episode. There was also concern that the risk adjustment may be “over –adjusting”, or

possibly "adjusting away" significant differences.
<p><b>3. Usability:</b></p> <p><b>3a. Measure performance results are publicly reported: H-1; M-1; L-2;I-2</b>  <i>TAP Discussion:</i> The TAP was concerned with the availability of this data to the public and requested clarification from NOF on what is required for "public reporting". 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: H-3; M-1; L-0; I-2</b>  <i>TAP Discussion:</i> The TAP agrees 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: H-0; M-2; L-3;I-1</b>  <i>TAP Discussion:</i> 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.</p> <p><b>3d. Harmonized or justification for differences: N/A</b></p>
<p><b>Overall Usability: H-0; M-10; L-7; I-0, N/A-0</b>  <b>Committee Discussion:</b> The Steering Committee discussed the fact that more information would be needed to explain the results of this measure to the public and to be used for internal quality improvement. The Steering Committee believes there are challenges for the use of this measure, which include its complexity and lack of clarity in the specifications. The Steering Committee agrees it is difficult to assess the extent of which the measure can be decomposed as it is currently specified.</p>
<p><b>4. Feasibility:</b></p> <p><b>4a. Data elements routinely generated during care process: H-5; M-0; L-0; I-1</b>  <i>TAP 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: H-5;M-0; L-0;I-1</b>  <i>TAP 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: H-0; M-4; L-1; I-1</b>  <i>TAP Discussion:</i> The TAP noted that Ingenix/OptumInsight 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: H-3; M-0; L-1; I-2</b>  <i>TAP Discussion:</i> The majority of the TAP agreed that barriers to use are minimal. (NOF Note: This is prior to the submission of product pricing information shared only with the Steering Committee)</p>
<p><b>Overall Feasibility: H-2; M-8; L-7; I-1</b>  <b>Committee Discussion:</b> See Ingenix/OptumInsight feasibility discussion in the discussion of recurring measure evaluation themes below.</p>



**1594 ETG Based Coronary Artery Disease (CAD) cost of care measure (Ingenix/OptumInsight)**

**Description:** 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).

**Resource Use Type:** Per episode

**Data Type:** Administrative claims, other

**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

**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 : states

**Measure Developer:** Ingenix/OptumInsight

**Committee Recommendation for Endorsement:** Y-5; N-9; Abstain – 0 (re-vote) [Y-8; N-10; Abstain-0 (initial vote)]

**1. Importance to Measure and Report**

**1a. High Impact:** H-5; M-0; L-0; I-0

**TAP Discussion:** The TAP believes this is a high impact, high cost area; this sub criterion has been met.

**1b. Resource use/cost problems:** H-5; M-0; L-0; I-0

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

**1c. Purpose clearly described:** H-5; M-0; L-0; I-0

**TAP Discussion:** The TAP believes this subcriterion has been met; the measure purpose is clearly described.

**1d. Resource use service categories consistent and representative:** H-3; M-2; L-0; I-0

**TAP Discussion:** The TAP believes this subcriterion has been met; the resource use categories are consistent and representative.

**Overall Importance:** Y-16, N-1 (Committee Vote)

**Committee Discussion:** The Steering Committee believes this is a high impact, high cost area and that the measure has been clearly described. This criterion has been met.

## 2. Scientific Acceptability of Measure Properties:

### 2a. Reliability:

#### 2a1. Well defined/precise specifications: H-3; M-1; L-0; I-0

**TAP Discussion:** The diagnoses 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, the Steering Committee raised 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.

#### 2a2. Reliability testing: H-3; M-1; L-0; I-0

**TAP Discussion:** The measure is specified in a way that it has been used over a long period of time, Ingenix/OptumInsight 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.

### 2b. Validity:

#### 2b1. Specifications consistent with resource use/cost problem: H-3; M-1; L-0; I-0

**TAP Discussion:** The TAP believes this subcriterion has been met; a specific population is defined and measured.

#### 2b2. Validity testing: H-3; M-0; L-0; I-0

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

#### 2b3. Exclusions: H-2; M-1; L-0; I-0

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

#### 2b4. Risk adjustment: H-2; M-1; L-0; I-0

**TAP Discussion:** 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/OptumInsight to supply this information to the Steering Committee.

#### 2b5. Identification of statistically significant/meaningful differences: H-1; M-0; L-1; I-1

**TAP Discussion:** 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 is not clinically significant.

#### 2b6. Multiple data sources: N/A

**TAP Discussion:** N/A

### 2c. Stratification for disparities: H-0; M-0; L-0; I-0; N/A-8

**TAP Discussion:** Due to the limitations in the administrative claims data, at this time the measure does not stratify for disparities.

**Overall Reliability :** H-5; M-11; L-2; I-0

**Overall Validity:** H-2; M-10; L-6; I-0

**Overall Scientifically Acceptable:** Yes [Y-12; N- 5 (Committee vote)]

**Committee Discussion:** The Steering Committee agreed that the measure accurately identified the primary incurring diagnoses codes as 410s through 414s. Within those strata there is a range of conditions – ranging from chronic, stable coronary artery disease to patients with cardiogenic shock complicated by a flail mitral posterior leaflet. The Steering Committee discussed how there is a large spectrum of risk adverse outcomes within this population. Furthermore, this carries the risk of different resource use for each specific condition included in the measure. The measure was submitted for implementation across various levels of analysis, however for individual clinicians there is not a sample size guideline. Regarding specific reliability testing, the measure is specified in a way that it has been used over a long period of time. The Steering Committee discussed how there are carve outs for mental health & pharmacy data and therefore comparisons within the health plan are the same or likely to be the same. However, when comparing across health plans or across physician groups validity may become an issue. There were concerns around the risk adjustment method. Specifically, the Committee was concerned that the measure may be adjusting for comorbidities identified during the measurement episode as opposed to comorbidities identified prior to the episode. There was also concern that the risk adjustment may be "over –adjusting", or possibly "adjusting away" significant differences.

## 3. Usability:

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

**TAP 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 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: H-0; M-2; L-1; I-0

**TAP Discussion:** The TAP agrees 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: H-0; M-3; L-0; I-0

**TAP Discussion:** 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 deconstructed for understanding as it is currently specified.

**3d. Harmonized or justification for differences: N/A**

**Overall Usability: H-1; M-11; L-4; I-1**

**Committee Discussion:** The Steering Committee agrees that more information would be needed to explain the results of this measure to the public and to be used for internal quality improvement. The Steering Committee discussed the challenges for the use of this measure, which include its complexity and lack of clarity in the specifications. The Steering Committee agrees it is difficult to assess the extent of which the measure can be decomposed as it is currently specified.

**4. Feasibility:**

**4a. Data elements routinely generated during care process: H-3; M-0; L-0; I-0**

**TAP Discussion:** The TAP believes that this sub criterion has been met; all of the data elements are generated during the care process.

**4b. Data elements available electronically: H-3; M-0; L-0; I-0**

**TAP Discussion:** The TAP believes that this sub criterion has been met; all of the data is available electronically.

**4c. Susceptibility to inaccuracies/ unintended consequences identified: H-2; M-1; L-0; I-0**

**TAP Discussion:** The TAP noted that Ingenix/OptumInsight does not have a formal audit system to ensure that all of the numbers are included & correct. In general, when dealing with any measure that uses administrative data there are various inaccuracies, pertaining particularly to coding inaccuracies and variation.

**4d. Data collection strategy can be implemented: H-2; M-0; L-1; I-0**

**TAP Discussion:** The majority of the TAP agreed that barriers to use are minimal. (NQF Note: This is prior to the submission of product pricing information shared only with the Steering Committee)

**Overall Feasibility: H-3; M-8; L-6; I-1**

**Committee Discussion:** See Ingenix/OptumInsight feasibility discussion in the discussion of recurring measure evaluation themes below.

**1595: ETG Based Diabetes Cost of Care Measure (Ingenix/OptumInsight)**

**Description:** The measure focuses on resources used to deliver episodes of care for patients with Diabetes. Diabetes 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 diabetes. A number of resource use measures are defined for diabetes 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. The focus of this submission is for Diabetes episodes and will cover both measures at the Diabetes base and severity level and also a Diabetes composite measure where Diabetes episode results are combined across Diabetes severity levels. At the most detailed level, the measure is defined as the base condition of diabetes and an assigned level of severity (e.g., resources per episode for diabetes, severity level 1 episodes).

**Resource Use Measure Type:** Per episode

**Data Source:** Administrative claims, Other

**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, Home Health, Hospice, Hospital/Acute Care Facility, Imaging Facility, Laboratory

**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

**Measure Developer:** Ingenix/OptumInsight

**Committee Recommendation for Endorsement:** Y-7; N-7; Abstain -0 (re-vote) [Y-11; N-7; Abstain-0 (initial vote)]

**1. Importance to Measure and Report:**

**1a. High Impact:** H-9 ; M-0 ; L-0 ; I-0

**TAP Discussion:** The TAP believes this is a high cost, impact aspect of healthcare; this subcriterion has been met.

**1b. Resource use/cost problems:** H- 3 ; M-6 ; L-0 ; I-0

**TAP Discussion:** The TAP would have liked to see more evidence of provider variation and other types of variation in treating diabetes in addition to the regional variation.

**1c. Purpose clearly described:** H- 4 ; M-5 ; L-0 ; I-0

**TAP Discussion:** The TAP believes that the intent provided not specific to this diabetes measure, it is a very general statement.

**1d. Resource use service categories consistent and representative:** H- 9 ; M-0 ; L-0 ; I-0

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

**Overall Importance:** Y-18, N-0

**Committee Discussion:** The Steering Committee believes this is a high impact area that should be measured; this subcriterion has been met.

**2. Scientific Acceptability of Measure Properties:**

**2a1. Well defined/precise specifications:** H- 5 ; M-3 ; L-1 ; I-0

**TAP Discussion:** Specifications for co-morbidities, severity levels, etc. are not clear. It is unclear if severity ratings are weighted based on services of comparable cost. Only costs that are mapped back to the diabetes code are counted in the episode. The measure is stratified by severity level not clinical condition. Concerns about how patients with pharmacy benefit (or who run out of pharmacy benefit) are compared to those with full pharmacy benefit.

**2a2. Reliability testing:** H- 7 ; M-1; L-0 ; I-0

**TAP Discussion:** Demonstration of internal consistency was presented to demonstrate reliability. The Committee requested additional reliability tests in during maintenance. Additional detail in terms of the r2 of the risk adjustment model and calibration results was requested.

**2b1. Specifications consistent with resource use/cost problem:** H- 1 ; M-6 ; L-1 ; I-0

**TAP Discussion:** TAP was unclear on whether diabetes education codes were included in the specifications?

**2b2. Validity testing:** H- 4 ; M-3; L-0 ; I-1

**TAP Discussion:** The TAP believes adequate validity testing information provided. More robust methods should be considered in future evaluations.

**2b3. Exclusions:** H-0; M-7 ; L1 ; I-0

**TAP Discussion:** TAP was unclear on how exclusions were identified.

**2b4. Risk adjustment :** H-0 ; M-4 ; L-4 ; I-0

**TAP Discussion:** The TAP was concerned about the inability to distinguish between complications and comorbidities.

**2b5. Identification of statistically significant/meaningful differences:** H- 0 ; M-4 ; L-4 ; I-0

**TAP Discussion:** Insufficient evidence that the sample size threshold and analysis at the physician level is meaningful at that level.

<p><a href="#">1595: ETG Based Diabetes Cost of Care Measure (Ingenix/OptumInsight)</a></p> <p>Unclear how the 30 sample size was selected.</p> <p><b>2b6. Multiple data sources:</b> N/A</p> <p><b>2c. Stratification for disparities:</b> H- 0 ; M-0 ; L-0 ; I-0; N/A-9</p> <p><b>TAP Discussion:</b> Due to the limitations in the administrative claims data, at this time the measure does not stratify for disparities.</p>
<p><b>Overall Scientifically Acceptable:</b> Yes [Y-10; N-8 (Committee Vote)]</p> <p><b>Committee Discussion:</b> As an introduction to the measure, the developer summarized their responses to the TAP concerns including that the diabetes education codes have been confirmed and are included in the specifications. Similar to the TAP, the Committee expressed concern about the minimum sample size guideline suggesting 30 cases per physician; the Committee questioned how this number was identified and if any statistical analysis was performed to support this guideline. In response to this concern, the developer explained that this sample size was borrowed from previous work done by NCOA on resource utilization and stated that from their perspective, while sample size can be important, ensuring results are statistically significant is more important. The Committee also requested explanation of the attribution model, finding that it was very complex, and questioned of the total sample from their analysis, what percent of physicians have a minimum sample size of 30. The developer explained that the attribution model seeks to identify the highest number of contacts between the physician and the patient related to diabetes; in case of a tie, the provider with the highest actual cost gets attributed the episode. Another concern identified by the Committee relates to how the measure captures costs related to the sequelae of diabetes (e.g., renal disease, eye disease, CHF); the measure as presented does not currently account for these costs as they trigger alternate episodes. There was also discussion on how this measure (or measures like it) might be paired with quality (process) measures, as it measures resource use and adjusts for conditions <i>before</i> care is provided. The Committee also spent some time discussing and trying to understand the episode trigger mechanisms, such as when a patient enters the episode in the middle of the 12-episode; in this case the episode is marked incomplete. There was a question to the developer about what percentage of the claims was higher or lower than expected. The developer was unable to answer the question off hand but will get back to the Committee with this information. The issue of mental health and pharmacy carve outs was a prevalent issue throughout the discussion of these measures. For this measure mental health is not stratified for when it is carved out.</p>
<p><b>3. Usability:</b></p> <p><b>3a. Measure performance results are publicly reported:</b> H- 0; M-1 ; L-1; I-6</p> <p><b>TAP Discussion:</b> The usability information submitted is not specific to diabetes, but for all Ingenix/OptumInsight measures. TAP expressed concerns with the availability of this data to the public and requested clarification from NOF on what is required for "public reporting". The NOF CSAC and BOD continue to discuss this issue; NOF staff will continue to filter any new information on the refining of this policy to the TAP to facilitate final ratings of this usability criterion.</p> <p><b>3b. Measure results are meaningful/useful for public reporting and quality improvement:</b> H- 0 ; M-4 ; L-2 ; I-2</p> <p><b>TAP Discussion:</b> The usability information submitted is not specific to diabetes, but for all Ingenix/OptumInsight measures.</p> <p><b>3c. Data and results can be decomposed for transparency and understanding:</b> H- 1 ; M-2 ; L-5 ; I-0</p> <p><b>TAP Discussion:</b> The usability information submitted is not specific to diabetes, but for all Ingenix/OptumInsight measures. Challenges for the use of this measure include, complexity, lack of specificity in specifications. The TAP agrees it is difficult to assess the extent of which the measure can be decomposed as currently specified.</p> <p><b>3d. Harmonized or justification for differences:</b> H-0 ; M-0 ; L-0 ; I-0; N-9</p> <p><b>TAP Discussion:</b> The usability information submitted is not specific to diabetes, but for all Ingenix/OptumInsight measures.</p>
<p><b>Overall Usability:</b> H-0; M-9; L-6; I-3</p> <p><b>Committee Discussion:</b> While there is a transparency website for physicians to go to in order determine what a score means, it may take a lot of time to do this. The Steering Committee questioned whether this is a reasonable expectation and adequately demonstrates transparency. Other concerns raised by the Steering Committee were related to the attribution model and how the complexity of the methodology might impact how understandable the measure construction and results are. Because this measure is part of an episode grouper and is not used in isolation as an individual measure, the information the developer was able to present on its current use is not specific to the diabetes episode, but the product as a whole.</p>
<p><b>4. Feasibility:</b></p> <p><b>4a. Data elements routinely generated during care process:</b> H- 8 ; M-0 ; L-0 ; I-0</p> <p><b>TAP Discussion:</b> The TAP agrees this subcriterion has been met; measures rely on administrative data.</p> <p><b>4b. Data elements available electronically:</b> H-8 ; M-0 ; L-0; I-0</p> <p><b>TAP Discussion:</b> The TAP agrees this subcriterion has been met; administrative data are in electronic format.</p> <p><b>4c. Susceptibility to inaccuracies/ unintended consequences identified:</b> H-2 ; M-2 ; L-4 ; I-0</p> <p><b>TAP Discussion:</b> The TAP does not feel this subcriterion was adequately met; there are current issues identified with specifications could result in inaccuracies and errors.</p> <p><b>4d. Data collection strategy can be implemented:</b> H- 5 ; M-2 ; L-1 ; I-0</p> <p><b>TAP Discussion:</b> The TAP agrees that barriers to use are minimal. (NOF Note: This is prior to the submission of product pricing information reviewed only by the Steering Committee).</p>

[1595: ETG Based Diabetes Cost of Care Measure \(Ingenix/OptumInsight\)](#)

4. Feasibility: H-2; M-8; L-8; I-0

*Committee Discussion:* See Ingenix/OptumInsight feasibility discussion in the discussion of recurring measure evaluation themes below.

**1599: ETG Based Non-Condition Specific cost of care measure (Ingenix/OptumInsight)**

**Description:** The measure focuses on resources used to diagnose, manage and treat a population of patients (non-condition specific) during a defined 12-month period of time. The population included in the measurement can be described generally. Examples include a population of individuals enrolled with a health plan, individuals assigned to a patient-centered medical home or accountable care organization (ACO), or a panel of individuals managed by a primary care physician (PCP). A number of resource use measures are defined for this measure set, 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 member per month and comparisons with internal and external benchmarks are made using risk adjustment to support valid comparisons. Risk adjustment is based on the measure of risk assigned to each individual using the Episode Risk Group (ERG) methodology.

**Resource Use Type:** Per capita (population- or patient-based)

**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

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

**Measure Developer:** Ingenix/OptumInsight

**Committee Recommendation for Endorsement: Y-5; N-9; Abstain-0 (re-vote) [ Y-12; N-6; Abstain-0 (initial vote)]**

**Conditions/Questions for Developer:**

1. How does the risk score correlate with the actual expenditures?
2. What is the distinction between ETGs and ERGs?
3. Can this measure be applied to the Medicare population?
4. Have there been any changes in the underlying risk model used in the ETGs since what has been published on the Ingenix/OptumInsight web site a year ago?
5. How are the carve outs, pharmacy and mental health data handled? How was this data validated?

**Developer Response:**

1. Ingenix/OptumInsight provides options for expenditure thresholds for a patient's annual member costs: \$25,000, \$100,000, and \$250,000. Ingenix/OptumInsight explained that these thresholds would vary depending on the application.
2. ETGs are episode-based measures. For example, an episode of diabetes, congestive heart failure or COPD--the severity models are built separately for each of the conditions which allows for risk adjustment for each separate condition-based episode. The results are then tagged for each episode for a member not only by condition, but also by the level of severity. There are hundreds of ETGs that map into the ERGs. Ingenix/OptumInsight maps to the ERG designed for the population-based risk adjustment; they weight each of the ERG markers to the final ERG score. The ERGs looks at age, in which case they may be applied to the Medicare population, however not all of the ETGs take age into account in the risk adjustment model. During the developer testing they didn't find that age had much explanatory power so they are not included in all of the ERGs. The ERG will point to a different weight depending on the age of the individual. However, since this measure has only been tested in a commercial database, per NQF policy, it can only be endorsed for use in commercial populations.
3. The ETG models and the risk models related to the ETGs have not been updated or recalibrated within the last year; therefore the information on the Ingenix/OptumInsight website is still applicable.
4. Ingenix/OptumInsight works with a population that has pharmacy and medical data. Mental health is excluded because the claims are not often available in addition to lack of coding for mental health services. Pharmacy data hasn't been an issue because it's up to the user whether they want to include and compare populations who have pharmacy data. The methodology can be adjusted, you are able to have a mixed population of both medical and pharmacy benefits, and the user is able to isolate the medical resource use data if they choose to.

**1. Importance to Measure and Report :Y-16; N-0**

**Committee Discussion:** This criterion was also discussed during the June 6 conference call. The [summary of the June 6 call](#) is available online.

**1a.High Impact: H-15; M-1; L-0; I-0**

**Committee Discussion:** The Steering Committee has deemed the measure focus to be high impact.

**1b. Resource use/cost problems: H-13; M-3; L-0; I-0**

**Committee Discussion:** The Steering Committee agrees this criterion has been met.

**1c. Purpose clearly described: H-12; M-4; L-0; I-0**

**Committee Discussion:** The Steering Committee believes the measure has met this sub criterion, as the measure's purpose is clearly

described.

**1d. Resource use service categories consistent and representative: H-8; M-8; L-0; I-0**

*Committee Discussion:* The resource use service categories are representative of the measure intent and focus.

**2. Scientific Acceptability of Measure Properties: Yes [Y-9; N-6 (Committee Vote)]**

**2a. Overall Reliability: H-8; M-7; L-1; I-0**

*Committee Discussion:* The Ingenix/OptumInsight team has a robust system where they double code the data – the steps that lead to the production of the data has a 99.9% match between the two approaches.. The Committee agreed that tables present measure results it is unclear if they actually represent that the measure is reliable.

**2a1. Measure well defined and precisely specified: H-10; M-5; L-1; I-0**

*Committee Discussion:* This measure appears to be well defined and specified. This methodology is used in a number of organizations and appears to work well. This sub criterion has been met.

**2a2. Reliability Testing: H-9; M-7; L-0; I-0**

*Committee Discussion:* The Committee agreed that this sub criterion has been met; the results have shown to be repeatable. The Committee suggested more robust reliability testing methods should be explored.

**2b. Overall Validity: H-2; M-10; L-3; I-0**

*Committee Discussion:* In the submission, Ingenix/OptumInsight states that they apply the methodology to data from several different organizations, but this is not detailed in any of the results. Face validity was tested however there is not any description of the results within the submission. The tables that were submitted to demonstrate validity are not clearly labeled or defined.

**2b1. Specifications consistent with intent: H-7; M-8; L-1; I-0**

*Committee Discussion:* The Committee agrees the specifications are consistent with the intent.

**2b2. Validity Testing: H-0; M-8; L-6; I-0**

*Committee Discussion:* This measure has been demonstrated to meet the requirement for face validity.

**2b3. Exclusions: H-9; M-4; L-2 ; I-0**

*Committee Discussion:* There are no exclusions based on cost or other criteria. The Committee reiterated concerns with comparability for plans that have pharmacy carve outs or do not have pharmacy data to those that do.

**2b4. Risk Adjustment: H-6; M-8; L-1; I-0**

*Committee Discussion:* When looking at the ETG codes, a severity score is assigned; the methodology then takes into account the ETG severity score and the number of comorbidities. A retrospective model contains the observed episodes that may occur during that year, but a user will not be able to observe any markers or costs for people who did not undergo services. The ERG risk level determines the individual's ERG risk score which drives the risk adjustment. The Committee acknowledged this methodology is very complex and not completely understood by all members.

**2b5. Identification of statistically significant/meaningful differences: H-5 ; M-7 ; L-3; I-0**

*Committee Discussion:* There is a way to stratify those with or without pharmacy data. The Committee expressed concern that valid comparisons cannot be made across organizations with different levels of data completeness and consistency.

**2b6. Multiple data sources: N/A**

**2c. Stratification for disparities: H-0; M-4; L-2; I-9**

*Committee Discussion:* This measure does not stratify by race and ethnicity. This may be possible in the future, but at the present time this information is not available.

**3. Overall Usability: H-0; M-10; L-5; I-0**

*Committee Discussion:* The Committee questioned on whether this measure has been featured in peer reviewed articles; the developer was unaware of any that could be shared with the Committee. The developers explained that this measure is currently being used to profile physicians. They are unaware of any efforts to publicly report the results, even within health plans to their covered lives.

**3a. Measure performance results are publicly reported: H-0; M-4; L-6; I-4**

*Committee Discussion:* Ingenix/OptumInsight conducted a survey of their customers, some users are publicly reporting the data and others are sharing information with physicians for incentive based programs. Some users have decided to put the information on a website that goes to their providers, which allows them to access their risk scores and score card. Providers are then able to drill down on the scorecard to the claim base level, the patient level and then the overall claims level.

**3b. Measure results are meaningful/useful for public reporting and performance improvement:**

*Committee Discussion:* H-3; M-6; L-3; I-3

**3c. Data and results can be decomposed for transparency and understanding: H-1; M-8; L-5; I-1**

*Committee Discussion:* While Ingenix/OptumInsight has a transparency website open to the public which explains the methodology and approach to measuring resources, the submission reviewed by the Committee was admittedly complex and at times difficult to identify the relevant information.

**3d. Harmonized or justification for differences: N/A**

**4. Feasibility: H-3; M-8, L-6, I-0**

**4a. Data elements routinely generated during care process: H-13; M-2; L-2; I-0**



**Discussion:** The Steering Committee believes that this sub criterion has been met; all of the data elements are generated during the care process.

**4b. Data elements available electronically: H-14, M-4, L-0, I-0**

**Discussion:** The Steering Committee believes that this sub criterion has been met; all of the data is available electronically.

**4c. Susceptibility to inaccuracies/ unintended consequences identified: H-5, M-9, L-3; I-0**

**Discussion:** Mental health is not available and pharmacy data rarely is, when pharmacy data is included it is stratified.

Ingenix/OptumInsight does not have a formal audit system to ensure that all of the numbers are included & correct. In general, when dealing with any measure that uses administrative data there are various inaccuracies, pertaining particularly to coding inaccuracies and variation. Ingenix/OptumInsight provides guidelines how to use small volumes/ sample sizes, however there is not content available to demonstrate this approach. This measure appears less prone to "gaming", as there is not much a user can do to manipulate the start or end of an episode.

**4d. Data collection strategy can be implemented: H-1, M-10, L-13, I-1**

**Discussion:** See Ingenix/OptumInsight feasibility discussion in the discussion of recurring measure evaluation themes below.

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

**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. 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).

**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/OptumInsight

**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. We exclude cases that are low in cost. We have the data to talk about the number of cases that are excluded by varying a low outlier, yes.
3. The hypothesis that that these low cost episodes – ones under 2.5 percent – are either mistakes or miscodes. They are probably incomplete episodes, so we don't count them.

**1. Importance to Measure and Report**

**1a.High Impact:** H-2; M-1; L-2; I-0

**TAP Discussion:** There was general agreement that hip fracture is a major cause of morbidity, mortality and high resource use. The TAP did, however, question the importance of measuring hip fractures in a predominately under 65 group of patients. Ingenix/OptumInsight acknowledged that this was a significant limitation of using administrative data.

**1b. Resource use/cost problems:** H-2; M-2; L-1; I-0

**TAP Discussion:** No issues were identified.

**1c. Purpose clearly described:** H-1; M-4; L-0; I-0

**TAP Discussion:** No issues were identified.

**1d. Resource use service categories consistent and representative:** H-2; M-2; L-1; I-0

**TAP Discussion:** The TAP were concerned that resource use service categories omit nursing homes and inpatient or outpatient rehab services.

**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/OptumInsight 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.

**2. Scientific Acceptability of Measure Properties:**

**2a. Overall Reliability:** H-1; M-0; L-4; I-0

**2a1.Measure well defined and precisely specified:** H-1; M-2; L-2; I-0

<p><b>TAP Discussion:</b> The TAP was concerned that the measure didn't capture certain co-morbid conditions such as dementia which are critical to understanding resource use for this clinical condition. There was substantial unease that the data does not examine the Medicare population, where the majority of hip-fractures occur.</p> <p><b>2a2. The results are repeatable:</b> H-1; M-2; L-2; I-0</p> <p><b>TAP Discussion:</b> The panel questioned whether one could infer grouper reliability from the tables submitted by Ingenix/OptumInsight. Ingenix/OptumInsight explained that the tables illustrate expected variability in results and point to a relatively consistent cost across health care organizations.</p> <p><b>2b. Overall Validity:</b> H-0; M-1; L-3; I-0</p> <p><b>2b1. Evidence is consistent with intent:</b> H-0; M-0; L-5; I-0</p> <p><b>TAP Discussion:</b> The TAP reiterated their concern that the measure hasn't captured the patient population most likely to be affected by hip fractures. Therefore, the measure may have limited applicability, due to the limitations of using only commercial data. The panel also felt that hip fractures in younger populations versus older populations represent two very different clinical situations.</p> <p><b>2b2. Score/Analysis:</b> H-0; M-1; L-4; I-0</p> <p><b>TAP Discussion:</b> The TAP was uncomfortable with the fact that all age groups were assigned the same risk coefficients. Ingenix/OptumInsight explained that this also represents a limitation of the data set, where they did not have the numbers over 65 to further stratify. Members of the panel believed that certain clinically relevant co-morbidities and complications such as dementia and post-op delirium should be reported on in a hip-fracture measure.</p> <p><b>2b3. Exclusions:</b> H-0; M-1; L-4; I-0</p> <p><b>TAP Discussion:</b> The TAP felt that the reasoning behind the exclusion criteria was unclear and not based on clinical evidence.</p> <p><b>2b4. Risk Adjustment:</b> H-0; M-0; L-4; I-1</p> <p><b>TAP Discussion:</b> The developer described how the measure contains low dollar exclusions. The assumption is that these claims represent incomplete episodes.</p> <p><b>2b5. Identification of statistically significant/meaningful differences:</b> H-0; M-0; L-4; I-1</p> <p><b>TAP Discussion:</b> There was a discussion regarding the relative cost of care ratio and a question about what numbers represent statistically significant differences. Ingenix/OptumInsight explained that the numbers would depend on the confidence interval, the underlying variance of episode cost and the number of total cases.</p> <p><b>2b6. Multiple data sources:</b> N/A (using all administrative data)</p> <p><b>2c. Stratification for disparities:</b> H-0; M-1; L-1; I-3</p> <p><b>TAP Discussion:</b> Racial disparities were addressed in the submission, but the data limits a further examination into these disparities.</p>
<p><b>Overall Scientifically Acceptable:</b> No [Y-7; N-10 (Committee Vote)]</p> <p><b>Overall Reliability:</b> H-1; M-11; L-3; I-2</p> <p><b>Overall Validity:</b> H-0; M-6; L-10; I-0</p> <p><b>Committee Discussion:</b> 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/OptumInsight for this measure represented a fairly <i>uncommon</i> 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 grouper function was not tested or reported on, and Ingenix/OptumInsight provided no information comparing scoring of attribution over episodes of time</p>
<p><b>Usability:</b></p> <p><b>3a. Measure performance results are publicly reported:</b> H-0; M-2; L-3; I-0</p> <p><b>TAP Discussion:</b> The TAP believes this subcriterion has been met.</p> <p><b>3b. Measure results are meaningful/useful for public reporting and quality improvement:</b> H-0; M-1; L-4; I-0</p> <p><b>TAP Discussion:</b> The TAP acknowledged the impressive amount of work Ingenix/OptumInsight put into this measure, but again articulated concern that the measure would have limited meaningful use as it is not capturing the appropriate population. The panel was uneasy with the grouping of two clinically different age cohorts together into one measure; they felt that the clinical situation, treatment path and mortality for a younger population with hip fractures versus an older population were different enough to warrant two separate measures.</p> <p><b>3c. Data and results can be decomposed for transparency and understanding:</b> H-0; M-2; L-3; I-0</p> <p><b>TAP Discussion:</b> The TAP agrees this subcriterion has been met.</p> <p><b>3d. Harmonized or justification for differences:</b> N/A</p>
<p><b>Overall Usability:</b> This measure did not pass the scientific acceptability criterion. As a result, the Committee did not discuss usability.</p>
<p><b>4. Feasibility:</b></p> <p><b>4a. Data elements routinely generated during care process:</b> H-3; M-1; L-1; I-0</p> <p><b>TAP Discussion:</b> The TAP agrees that this subcriterion has been met; all data is routinely generated through the care process.</p> <p><b>4b. Data elements available electronically:</b> H-4; M-0; L-1; I-0</p>

**TAP Discussion:** The TAP agrees that this subcriterion has been met; all data is available electronically.

**4c. Susceptibility to inaccuracies/ unintended consequences identified:** H-1; M-1; L-3; I-0

**TAP Discussion:** The TAP believe that this subcriterion has been met, however Ingenix/OptumInsight does not have a formal audit system in order to monitor for inaccuracies.

**4d. Data collection strategy can be implemented:** H-0; M-2; L-2; I-1

**TAP Discussion:** The TAP believe that this subcriterion has been met. (NQF Staff Note: this is prior to the submission of product pricing information reviewed by the Steering Committee only.)

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

**1605: ETG Based Asthma Cost of Care (Ingenix/OptumInsight)**

**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).  
**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/OptumInsight

**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/OptumInsight 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 unlimited training involved (i.e. help desk support, etc.). There is another option where Ingenix/OptumInsight takes the data and runs it themselves - or uses their PCQ Connect product that prepared the data into report-ready formats.

**1.Importance to Measure and Report**  
**1a.High Impact:** H-9; M-0; L-0  
**TAP Discussion:** The TAP agrees that asthma is a very important health care area to measure.  
**1b. Resource use/cost problems:** H-8; M-1; L-0 ; I-0  
**TAP Discussion:** The TAP agrees the Measure demonstrates cost problems and opportunity for improvement.  
**1c. Purpose clearly described:** H-7; M-2; L-0; I-0  
**TAP Discussion:** The TAP believes the purpose and objective of the measure are clear.  
**1d. Resource use service categories consistent and representative:** H-7; M-2; L-0; I-0  
**TAP Discussion:** The TAP feel this subcriterion has been met.

**Overall Importance:** Y-16, N-0  
**Committee Discussion:** The Steering Committee agreed that asthma constitutes a high impact healthcare area.

**2. Scientific Acceptability of Measure Properties:**  
**2a. Reliability:** H-0; M-8; L-1; I-0  
**2a1.Measure well defined and precisely specified:** H-2; M-6; L-1; I-0  
**TAP Discussion:** This measure is one that's part of a suite of episodes around diseases and conditions included in Ingenix/OptumInsight's episode treatment grouper. This product identifies claims that should be part of an episode of asthma and divides them into year-long segments, looking at asthma as a chronic disease. The episodes are severity adjusted using clinical markers called

condition status factors. Anchor episodes, or face-to-face encounters, are merged together into one episode (i.e. "asthma").

**2a2. The results are repeatable:** H-3; M-5; L-1; I-0

**TAP Discussion:** The TAP didn't understand why Ingenix/OptumInsight used three different population samples, rather than taking a portion of the larger population and testing it multiple times. They would like better communication on the approach as well as more detailed depiction of the data. Repeatability was generally determined to be demonstrated adequately, but for the above reasons, some did question the reliability of the measure score.

**2b. Overall Validity:** H-0; M-6; L-1; I-2

**2b1. Evidence is consistent with intent:** H-2; M-5; L-1; I-1

**TAP Discussion:** It was unclear to the panel whether Ingenix/OptumInsight is actually measuring asthma costs as intended. The determination of what is an asthma cost and what is not isn't transparent. They also agreed that any results are going to be questioned when potentially over 50% of the costs (the pharmacy costs) are not represented. There were suggestions to stratify those health plans that have pharmacy carve-out arrangements.

**2b2. Score/Analysis:** H-1; M-4; L-2; I-2

**TAP Discussion:** Face validity was determined to be appropriate. The TAP continued to express concern about the exclusion of pharmacy costs, which were agreed to be a significant component of asthma care. Pharmacy data is not a requirement to get into the episode (for all ETGs).

**2b3. Exclusions:** H-1; M-7; L-1; I-0

**TAP Discussion:** The TAP was concerned about the lack of transparency regarding which costs were excluded, and why. Confusion existed around what the grouper identified as outliers or exclusions. Winsorizing very high cost episodes, the top 2%, effectively excludes those kinds of patients that would be important to know about. Additional information such as sensitivity analyses would have helped explain the impact of these high cost cases.

**2b4. Risk Adjustment:** H-1; M-4; L-2; I-2

**TAP Discussion:** The TAP expressed the same concerns regarding the risk-adjustment methodology as they had for previous Ingenix/OptumInsight measures. The TAP was apprehensive that because the measure doesn't require use of standardized costs, the playing field is not level and it can't be implemented consistently across organizations if one is using standard and another actual pricing. To examine how refined the risk-adjustment is, R-squares for different severity levels and how they predict resource utilization should be provided.

**2b5. Identification of statistically significant/meaningful differences:** H-0; M-8; L-0; I-1

**TAP Discussion:** The TAP felt confident in Ingenix/OptumInsight's methodology after it was explained.

**2b6. Multiple data sources:** N/A (using all administrative data)

**2c. Stratification for disparities:** H-2; M-6; L-0; I-1

**TAP Discussion:** Gender and age can be stratified, but race data is not available.

**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/OptumInsight 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.

**Usability:**

**3a. Measure performance results are publicly reported:** H-2; M-4; L-2; I-1

**TAP Discussion:** This product is generally used with a suite of ETG's, usually in combination with the pneumonia and COPD measures. There was uncertainty about the measure's usefulness on its own. Since Ingenix/OptumInsight can't ascertain if this measure is being used individually the concern from the panel is how the individual measure could be used.

**3b. Measure results are meaningful/useful for public reporting and quality improvement:** H-0; M-6; L-2; I-1

**TAP Discussion:** The TAP was concerned about the possibility of misinterpretation of results because of the transparency and usability of the results of this measure.

**3c. Data and results can be decomposed for transparency and understanding:** H-3; M-5; L-1; I-0

**TAP Discussion:** The TAP reiterated their concern of the transparency of the score. Ingenix/OptumInsight clarified that there are ways to drill into different aspects of care to see how they might be driving the score.

**3d. Harmonized or justification for differences:** N/A

**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.

**4. Feasibility:**

**4a. Data elements routinely generated during care process:** H-7; M-2; L-0; I-0

*TAP Discussion:* The TAP believes this subcriterion has been met; data is a byproduct of care.

**4b. Data elements available electronically:** H-7; M-2; L-0; I-0

*TAP Discussion:* The TAP agrees this subcriterion has been met; data is available electronically.

**4c. Susceptibility to inaccuracies/ unintended consequences identified:** H-1; M-8; L-0; I-0

*TAP Discussion:* The TAP was generally comfortable with the error checks built into the product.

**4d. Data collection strategy can be implemented:** H-4; M-4; L-0; I-1

*TAP Discussion:* The TAP expressed some concern about the burden this measure would place on a programmer to implement, particularly at smaller health plans.

**Overall Feasibility:** H-1; M-8; L-7; I-0

**Committee Discussion:** See Ingenix/OptumInsight feasibility discussion in the discussion of recurring measure evaluation themes below.

**1608: ETG Based Chronic Obstructive Pulmonary Disease Cost of Care (COPD) (Ingenix/OptumInsight)**

**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. 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).

**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/OptumInsight

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

**Conditions/Questions for Developer:**

1. What was the clinical logic of using 180 days, particularly since your Asthma measure had used 365 days, and both are similar chronic conditions?

**Developer Response:**

1. We will have to examine that further.

**1. Importance to Measure and Report**

**1a.High Impact:** H-7; M-0; L-0; I-0

**TAP Discussion:** The TAP agreed Ingenix/OptumInsight did well with articulating the high impact of COPD.

**1b. Resource use/cost problems:** H-7; M-0; L-0; I-0

**TAP Discussion:** The TAP believe that COPD represents a resource use issue that can be addressed.

**1c. Purpose clearly described:** H-7; M-0; L-0; I-0

**TAP Discussion:** The TAP feel the purpose and objective are clear.

**1d. Resource use service categories consistent and representative:** H-6; M-1; L-0; I-0

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

**Overall Importance:** Y-16, N-0

**Committee Discussion:** There was unanimous agreement that asthma constitutes a high impact area of healthcare.

**2. Scientific Acceptability of Measure Properties:**

**2a. Overall Reliability:** H-4; M-3; L-0; I-0

**2a1.Measure well defined and precisely specified:** H-4; M-3; L-0; I-0

**TAP Discussion:** The TAP discussion focused around the clinical logic around the timeframes chosen.

**2a2. The results are repeatable:** H-5; M-2; L-0; I-0

**TAP Discussion:** The TAP agrees that reliability for this measure is similar to the previously discussed Ingenix/OptumInsight asthma measure.

**2b. Overall Validity:** H-0; M-7; L-0; I-0

**2b1. Evidence is consistent with intent:** H-2; M-5; L-0; I-0

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

**2b2.Score/Analysis:** H-0; M-7; L-0; I-0

**TAP Discussion:** The TAP remained concerned about Ingenix/OptumInsight's testing method for customization, the inability to compare actual versus standardized prices, and the high level of pharmacy exclusions.

**2b3. Exclusions:** H-1; M-6; L-0; I-0

**TAP Discussion:** There are no clinical exclusions, only administrative ones. The TAP felt it was unclear how tie-breaking logic works and noted that it was not specified in the submission how COPD and asthma ETG's interact.



<p><b>2b4. Risk Adjustment:</b> H-0; M-4; L-3; I-0  <i>TAP Discussion:</i> While Ingenix/OptumInsight had a nice description of how they developed their risk-adjustment approach, the panel would have liked to see more description of the modeling presented in the submission.</p> <p><b>2b5. Identification of statistically significant/meaningful differences:</b> H-0; M-7; L-0; I-0  <i>TAP Discussion:</i> The TAP questioned whether the practical significance of the measure since it is a relative cost ratio.</p> <p><b>2b6. Multiple data sources:</b> N/A (using all administrative data)</p> <p><b>2c. Stratification for disparities:</b> H-2; M-5; L-0; I-0  <i>TAP Discussion:</i> Only gender and age are stratified for. Race data is not available.</p>
<p><b>Overall Reliability:</b> H-3; M-10; L-2; I-0  <b>Overall Validity:</b> H-1; M-5; L-9; I-0  <b>Overall Scientifically Acceptable:</b> Yes [Y-5; N-10 (Committee Vote)]  <b>Committee Discussion:</b> The Steering Committee appreciated the change Ingenix/OptumInsight made to the measure's timeframe at 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 1560 Asthma.</p>
<p><b>Usability:</b></p> <p><b>3a. Measure performance results are publicly reported:</b> H-0; M-7; L-0; I-0  <i>TAP Discussion:</i> The TAP expressed doubts regarding whether the measure could be implemented in a user-friendly manner.</p> <p><b>3b. Measure results are meaningful/useful for public reporting and quality improvement:</b> H-0; M-7; L-0; I-0  <i>TAP Discussion:</i> The panel agreed that measure provides useful information for individual health plans. However, they expressed concern about how useful it would be to compare across health plans, due to the fact that standardized pricing is not required.</p> <p><b>3c. Data and results can be decomposed for transparency and understanding:</b> H-3; M-4; L-0; I-0  <i>TAP Discussion:</i> It was agreed that previous discussions regarding Ingenix/OptumInsight transparency would also apply to this measure.</p> <p><b>3d. Harmonized or justification for differences:</b> N/A</p>
<p><b>Overall Usability:</b> This measure did not pass the scientific acceptability criterion. As a result, the Committee did not discuss usability.</p>
<p><b>4. Feasibility:</b></p> <p><b>4a. Data elements routinely generated during care process:</b> H-5; M-2; L-0; I-0  <i>TAP Discussion:</i> The TAP believes this subcriterion has been met; data is a byproduct of care.</p> <p><b>4b. Data elements available electronically:</b> H-7; M-0; L-0; I-0  <i>TAP Discussion:</i> The TAP believes this subcriterion has been met; data available electronically.</p> <p><b>4c. Susceptibility to inaccuracies/ unintended consequences identified:</b> H-3; M-4; L-0; I-0  <i>TAP Discussion:</i> The TAP is comfortable that Ingenix/OptumInsight can accurately identify inaccuracies and errors.</p> <p><b>4d. Data collection strategy can be implemented:</b> H-6; M-1; L-0; I-0  <i>TAP Discussion:</i> The TAP believes this subcriterion has been met.</p>
<p><b>Overall Feasibility:</b> This measure did not pass the scientific acceptability criterion. As a result, the Committee did not vote on feasibility.</p>

## Recurring Measure Evaluation Themes

### *Costing Approaches*

Early in the evaluation process, the Committee agreed that it was important to distinguish measure results obtained using standardized prices and actual prices paid; dividing the costing approaches into separate measures was determined to be the best approach to ensure this distinction was made for standardized implementation and ensure consistent and accurate comparisons of measure results. While the combination of these approaches in a single measure is typical for use in the commercial sector, for use as a national consensus standard, measure results should unambiguously reflect differences in performance for an accountable entity, not differences in the type of data an entity chooses to submit (actual prices or standardized prices). As such, 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. Recognizing that measure results applying both costing approaches are often used and reported together by current users, splitting the measures for purposes of endorsement does not preclude the use of the two measures as a pair. This was requested of both HealthPartners (in cycle one) and of Ingenix/OptumInsight (in cycles one and two). HealthPartners subsequently resubmitted two separate measures, one applying each costing approach; Ingenix/OptumInsight resubmitted all of their measures applying only actual prices paid.

During the initial evaluation and voting for recommendation of the Ingenix/OptumInsight measures, there was not a shared understanding among the Committee that the measures had been submitted with flexibility in the costing approach. Ingenix/OptumInsight chose to resubmit their measures using actual prices paid. Once the measures were resubmitted to the Committee applying the single costing approach, the Committee was given the opportunity to determine if the selection in the costing approach warranted a re-vote. The Committee requested a revote since there was not a shared understanding on the original costing approach by Ingenix/OptumInsight, thus all Ingenix/OptumInsight measures were subject to a re-vote during the Cycle 2 Committee meeting. The re-vote was for overall recommendation for endorsement only. This is reflected as such in the measure evaluation summaries below.

### *Feasibility*

Each of the individual Ingenix/OptumInsight measures [(Episode Treatment Groups (ETGs))] exists as part of a larger grouper system, and requires the use of the entire grouper to produce results for the individual ETGs. Because, each of the condition-specific ETGs submitted to this project require the use of the Ingenix/OptumInsight grouper product to implement the measures, the Committee's discussion of the feasibility criterion for these measures was done for all of these measures at one time. As a part of feasibility discussion, the Committee was provided with a pricing table for each of the products required for implementation of these condition-specific ETGs.

Because these measures primarily use administrative claims data, all of the data required to implement these measures is generated as a byproduct of care and is available electronically. In terms of barriers to use, the purchase and implementation of this product could be cost prohibitive for some entities. Annually, for physicians the cost to implement this project could range from of the small package \$70,000 (for a group of less than 800 physicians) to \$110,000 (for over 2,000 physicians in the group). For health plans, the annual cost could range from \$90,115 (for less than 400,000 covered lives) to is \$135,000 (for over a million covered lives). The Steering Committee concluded that this cost is comparable to the cost of other proprietary fees associated with other risk adjustment models of its caliber (e.g., ACGs used by HealthPartners). These prices include costs associated with the licensure of the proprietary software and the cost of all of their measures, over 558 ETGs, but not implementation. The Steering Committee acknowledged that while the methodology is very complex, the system may be used without Ingenix/OptumInsight's technical support, if the user spends time thoroughly reviewing the documentation.

## **WITHDRAWN FROM CONSIDERATION**

The 21 measures listed below were withdrawn from the review process by the developers for further refinement and testing.

### ***Cardiovascular***

- (1570) Acute Myocardial Infarction Episode-of-Care for 30 Days Following Onset (ABMS-REF)
- (1571) Acute Myocardial Infarction Episode-of-Care for Post-Acute Period (days 31-365) (ABMS-REF)
- (1572) Episode of Care for Management of Chronic Coronary Artery Disease (ABMS)
- (1573) Episode of Care for Management of coronary Artery Disease Post Re-Vascularization (ABMS-REF)
- (1574) Episode of Care for Management of Chronic Congestive Heart Failure over a 12 month period (ABMS-REF)
- (1575) Episode of Care for Management of Post-Hospitalization Chronic Congestive Heart Failure over a 4 Month Period (ABMS-REF)
- (1576) Episode of Care for Patients with Diabetes over a One Year Period (ABMS-REF)
- (1593) ETG Based Acute Myocardial Infarction (AMI) Resource Use Measure (Ingenix/OptumInsight)

### ***Stroke***

- (1596) ETG Based Stroke Resource Use Measure (Ingenix/OptumInsight)

### ***Pulmonary***

- (1577) Episode of care for patients with asthma over a one year period (ABMS-REF)
- (1581) Episode of care for patients with stable chronic obstructive pulmonary disease over a one year period (ABMS-REF)
- (1582) Episode of care for patients with unstable chronic obstructive pulmonary disease over a one year period (ABMS-REF)
- (1587) Episode of care for ambulatory pneumonia (ABMS-REF)
- (1588) Episode of care for community acquired pneumonia hospitalization (ABMS-REF)

### ***Cancer***

- (1578) Episode of care for 60-day period preceding breast biopsy (ABMS-REF)
- (1579) Episode of care for cases of newly diagnosed breast cancer over a 15 month period (ABMS-REF)
- (1583) Episode of care for 21-day period around a colonoscopy (ABMS-REF)
- (1584) Episode of care for treatment of localized colon cancer (ABMS-REF)

### ***Bone/Joint***

- (1585) Episode of care for simple, non-specific lower back pain (acute and subacute) (ABMS-REF)
- (1586) Episode of care for acute/subacute lumbar radiculopathy with or without lower back pain (ABMS-REF)
- (1610) ETG based low back pain resource use measure (Ingenix/OptumInsight)

## ENDNOTES

1. Catlin A, et al., National Health Spending in 2006: A Year of Change for Prescription Drugs, *Health Affairs*, 2008; 27(1):14–29.
2. Banks J, et al., Disease and disadvantage in the United States and in England, *JAMA*, 2006;295(17):2037–2045.
3. Hoyert DL, et al., Annual summary of vital statistics: 2004, *Pediatrics*, 2006; 117(1):168–183.
4. Weiss JE, Mushinski M, International mortality rates and life expectancy: selected countries, *Statistical Bulletin—Metropolitan Life Insurance Company*, 1999;80(1):13–21.
5. Catlin A, et al.
6. CMS Grouper Episode for Medicare. Federal Business Opportunities. Obtained from: [https://www.fbo.gov/index?s=opportunity&mode=form&id=452b7af0f5752ba0c444853793e8024d&tab=core&\\_cview=1](https://www.fbo.gov/index?s=opportunity&mode=form&id=452b7af0f5752ba0c444853793e8024d&tab=core&_cview=1). Last Accessed August 2011.