NATIONAL QUALITY FORUM

Measure Evaluation 4.1
December 2009

This form contains the measure information submitted by stewards. Blank fields indicate no information was provided. Attachments also may have been submitted and are provided to reviewers. The subcriteria and most of the footnotes from the evaluation criteria are provided in Word comments within the form and will appear if your cursor is over the highlighted area. Hyperlinks to the evaluation criteria and ratings are provided in each section.

TAP/Workgroup (if utilized)**:** Complete all yellow highlighted areas of the form. Evaluate the extent to which each subcriterion is met. Based on your evaluation, summarize the strengths and weaknesses in each section.

Note: *If there is no TAP or workgroup, the SC also evaluates the subcriteria (yellow highlighted areas).*

Steering Committee: Complete all pink highlighted areas of the form. Review the workgroup/TAP assessment of the subcriteria, noting any areas of disagreement; then evaluate the extent to which each major criterion is met; and finally, indicate your recommendation for the endorsement. Provide the rationale for your ratings.

Evaluation ratings of the extent to which the criteria are met
C = Completely (unquestionably demonstrated to meet the criterion)
P = Partially (demonstrated to partially meet the criterion)
M = Minimally (addressed BUT demonstrated to only minimally meet the criterion)
N = Not at all (NOT addressed; OR incorrectly addressed; OR demonstrated to NOT meet the criterion)
NA = Not applicable (only an option for a few subcriteria as indicated)

(for NQF staff use) **NQF Review #:** 0290 **NQF Project:** Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Median Time to Transfer to Another Facility for Acute Coronary Intervention
De.2 **Brief description of measure**: Median time from emergency department arrival to time of transfer to another facility for acute coronary intervention.
1.1-2 **Type of Measure:** Process De.3 **If included in a composite or paired with another measure, please identify composite or paired measure**
De.4 **National Priority Partners Priority Area:** Safety De.5 **IOM Quality Domain**: Timeliness De.6 **Consumer Care Need:** Getting better

CONDITIONS FOR CONSIDERATION BY NQF	
Four conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards:	**NQF Staff**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	**A** **Y**☐ **N**☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** **Y**☐ **N**☐

C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Payment incentive	**C** Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☐ N☐
(for NQF staff use**)** **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	**Eval Rating**
(for NQF staff use) **Specific NPP goal**:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers 1a.2 1a.3 **Summary of Evidence of High Impact:** The early use of primary angioplasty in patients with ST-segment myocardial infarction (STEMI) results in a significant reduction in mortality and morbidity. The earlier primary coronary intervention is provided, the more effective it is (Brodie, 1998 and DeLuca, 2004). National guidelines recommend the prompt initiation of percutaneous coronary intervention (PCI) in patients presenting with ST-segment elevation myocardial infarction (Antman, 2004). Patients transferred for primary PCI rarely meet recommended guidelines for door-to-balloon time (Nallamothu, 2005). Times to treatment in transfer patients undergoing primary PCI may influence the use of PCI as an intervention (Nallamothu, 2005). Current recommendations support a door-to-balloon time of 90 minutes or less (Krumholz, 2008). 1a.4 **Citations for Evidence of High Impact:** • Antman EM, Anbe DT, Armstrong PW, Bates ER, Green LA, Hand M, Hochman JS, Krumholz HM, Kushner FG, Lamas GA, Mullany CJ, Ornato JP, Pearle DL, Sloan MA, Smith SC Jr. ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1999 Guidelines for the Management of Patients With Acute Myocardial Infarction). 2004. • Krumholz HM, Anderson JL, Bachelder BL, Fesmire FM, Fihn SD, Foody JM, et al. ACC/AHA 2008 performance measures for adults with ST-elevation and non-ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures (Writing Committee to Develop Performance Measures for ST-Elevation and Non-ST-Elevation Myocardial	**1a** C☐ P☐ M☐ N☐

Comment [KP1]: 1a. The measure focus addresses:
•a specific national health goal/priority identified by NQF's National Priorities Partners; OR
•a demonstrated high impact aspect of healthcare (e.g., affects large numbers, leading cause of morbidity/mortality, high resource use (current and/or future), severity of illness, and patient/societal consequences of poor quality).

Infarction). J Am Coll Cardiol. 2008;52:2046-99.
• Peacock WF, Hollander JE, Smalling RW, and Bresler MJ. Reperfusion Strategies in the emergency treatment of ST-segment elevation myocardial infarction. Am J Emerg Med 2007; 25: 353-66.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Target is to transport patients to facility for acute coronary intervention to receive intervention within 90 minutes of presentation at original facility.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
Q1 2010 Analysis Provider Level
1,050 hospitals submitted 3,164 eligible cases.
Median 68.75 minutes
Min 0 minutes
Max 540 minutes *capped
5th percentile 269 minutes
10th percentile 189 minutes
25th percentile 115 minutes
75th percentile 49 minutes
90th percentile 36 minutes
95th percentile 30 minutes

1b.3 Citations for data on performance gap:
1,050 hospitals submitted 3,164 eligible cases. Median patient time was 62 minutes. Median provider time was 68.75 minutes.

1b.4 Summary of Data on disparities by population group:
N/A

1b.5 Citations for data on Disparities:
N/A

1b C☐ P☐ M☐ N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** The early use of primary angioplasty in patients with ST-segment myocardial infarction (STEMI) results in a significant reduction in mortality and morbidity. The earlier primary coronary intervention is provided, the more effective it is (Brodie, 1998 and DeLuca, 2004). National guidelines recommend the prompt initiation of percutaneous coronary intervention (PCI) in patients presenting with ST-segment elevation myocardial infarction (Antman, 2004). Patients transferred for primary PCI rarely meet recommended guidelines for door-to-balloon time (Nallamothu, 2005). Times to treatment in transfer patients undergoing primary PCI may influence the use of PCI as an intervention (Nallamothu, 2005). Current recommendations support a door-to-balloon time of 90 minutes or less (Krumholz, 2008).

1c.2-3. Type of Evidence: Evidence-based guideline

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
The early use of primary angioplasty in patients with ST-segment myocardial infarction (STEMI) results in a significant reduction in mortality and morbidity. The earlier primary coronary intervention is provided, the more effective it is (Brodie, 1998 and DeLuca, 2004). National guidelines recommend the prompt initiation of percutaneous coronary intervention (PCI) in patients presenting with ST-segment elevation myocardial infarction (Antman, 2004). Patients transferred for primary PCI rarely meet recommended guidelines for door-to-balloon time (Nallamothu, 2005). Times to treatment in transfer patients undergoing primary PCI may influence the use of PCI as an intervention (Nallamothu, 2005). Current recommendations support a door-to-balloon time of 90 minutes or less (Krumholz, 2008).

1c C☐ P☐ M☐ N☐

Comment [KP2]: 1b. Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall poor performance, in the quality of care across providers and/or population groups (disparities in care).

Comment [k3]: 1 Examples of data on opportunity for improvement include, but are not limited to: prior studies, epidemiologic data, measure data from pilot testing or implementation. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem.

Comment [k4]: 1c. The measure focus is:
•an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;
OR
•if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
oIntermediate outcome – evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
oProcess – evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
oStructure – evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
oPatient experience – evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
oAccess – evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
oEfficiency – demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.

Comment [k5]: 4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., ... [1]

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
A -ABC Scale ACC/AHA Strength of Evidence and Meta Analysis

1c.6 Method for rating evidence: ACC/AHA Strength of Evidence and Meta Analysis

1c.7 Summary of Controversy/Contradictory Evidence: N/A

1c.8 Citations for Evidence (*other than guidelines***):** N/A

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
"if PCI is chosen, the delay from patient contact with the healthcare system (typically, arrival at the ED or contact with paramedics) to balloon inflation should be less than 90 minutes." Page 593

1c.10 Clinical Practice Guideline Citation: Antman EM, Anbe DT, Armstrong PW, Bates ER, Green LA, Hand M, Hochman JS, Krumholz HM, Kushner FG, Lamas GA, Mullany CJ, Ornato JP, Pearle DL, Sloan MA, Smith SC Jr. ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1999 Guidelines for the Management of Patients With Acute Myocardial Infarction). 2004.

1c.11 National Guideline Clearinghouse or other URL: N/A

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
B ABC Scale

1c.13 Method for rating strength of recommendation (*If different from USPSTF system, also describe rating and how it relates to USPSTF)***:**
ABC Scale
- Level A (randomized controlled trial/ meta-analysis):

High quality randomized controlled trial that considers all important outcomes. High-quality meta-analysis (quantitative systematic review) using comprehensive search strategies.
- Level B (other evidence):

A well-designed, nonrandomized clinical trial. A nonquantitative systematic review with appropriate search strategies and well-substantiated conclusions. Includes lower quality randomized controlled trials, clinical cohort studies, and case-controlled studies with nonbiased selection of study participants and consistent findings. Other evidence, such as high-quality, historical, uncontrolled studies, or well-designed epidemiologic studies with compelling findings, is also included.
- Level C (consensus/expert opinion):

Consensus viewpoint or expert opinion. Expert opinion is sometimes the best evidence available.

1c.14 Rationale for using this guideline over others:
ACC/AHA Strength of Evidence and Meta Analysis

Comment [k6]: 3 The strength of the body of evidence for the specific measure focus should be systematically assessed and rated (e.g., USPSTF grading system http://www.ahrq.gov/clinic/uspstf07/methods/benefit.htm). If the USPSTF grading system was not used, the grading system is explained including how it relates to the USPSTF grades or why it does not. However, evidence is not limited to quantitative studies and the best type of evidence depends upon the question being studied (e.g., randomized controlled trials appropriate for studying drug efficacy are not well suited for complex system changes). When qualitative studies are used, appropriate qualitative research criteria are used to judge the strength of the evidence.

Comment [k7]: USPSTF grading system **http://www.ahrq.gov/clinic/uspstf/grades.htm**: **A -** The USPSTF recommends the service. There is high certainty that the net benefit is substantial. **B -** The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial. **C -** The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small. Offer or provide this service only if other considerations support the offering or providing the service in an individual patient. **D -** The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits. **I -** The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Eval Rating**
2a. MEASURE SPECIFICATIONS	

S.1 Do you have a web page where current detailed measure specifications can be obtained?
S.2 If yes, provide web page URL:

2a. Precisely Specified

2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):**
Continuous Variable Statement:
Time (in minutes) from emergency department arrival to transfer to another facility for acute coronary intervention

Included Populations:
- ICD-9-CM Principal Diagnosis Code for AMI as defined in Appendix A, OP Table 6.1, and
- E/M Code for emergency department encounter as defined in Appendix A, OP Table 1.0a, and
- Patients discharged/transferred to a short-term general hospital for inpatient care, to a Federal healthcare facility, or to a Critical Access Hospital, and
- Patients not receiving Fibrinolytic Administration as defined in the Data Dictionary, and
- Patients with Transfer for Acute Coronary Intervention as defined in the Data Dictionary

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator***):**
During the measurement period

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**
Patients with:
- An E/M Code for emergency department encounter as defined in Appendix A, OP Table 1.0, and
- Patients discharged/transferred to a short-term general hospital for inpatient care, or to a Federal healthcare facility, and
- An ICD-9-CM Principal Diagnosis Code for AMI as defined in Appendix A, OP Table 1.1, and
- ST-segment elevation or LBBB on the ECG performed closest to ED arrival, and
- Patients with Transfer for Acute Coronary Intervention as defined in the Data Dictionary

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
Time (in minutes) from emergency department arrival to transfer to another facility for acute coronary intervention.

2a.5 Target population gender: Female, Male
2a.6 Target population age range: 18 years of age and older

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
During the measurement period

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Patients with:
- An E/M Code for emergency department encounter as defined in Appendix A, OP Table 1.0, and
- Patients discharged/transferred to a short-term general hospital for inpatient care, or to a Federal healthcare facility, and
- An ICD-9-CM Principal Diagnosis Code for AMI as defined in Appendix A, OP Table 1.1, and
- ST-segment elevation or LBBB on the ECG performed closest to ED arrival, and
- Patients with Transfer for Acute Coronary Intervention as defined in the Data Dictionary

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):**
- Patients less than 18 years of age
- Patients receiving Fibrinolytic Administration as defined in the Data Dictionary

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**

2a-specs
C☐ P☐ M☐ N☐

Comment [KP8]: 2a. The measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. The required data elements are of high quality as defined by NQF's Health Information Technology Expert Panel (HITEP) .

Comment [k9]: 11 Risk factors that influence outcomes should not be specified as exclusions.
12 Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Specifications available at http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1196289981244

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
Specifications available at http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1196289981244

2a.12-13 Risk Adjustment Type: No risk adjustment necessary

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method***):**
N/A

2a.15-17 Detailed risk model available Web page URL or attachment:

2a.18-19 Type of Score: Continuous variable
2a.20 Interpretation of Score: Better quality = Lower score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps***):**
Specifications available at http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1196289981244

2a.22 Describe the method for discriminating performance *(e.g., significance testing)***:**
N/A

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
Sampling Approaches
As previously stated in this section, hospitals have the option to sample from their population, or submit their entire population. Hospitals that choose to sample must ensure that the sampled data represent their outpatient population by using either the simple random sampling or systematic random sampling method and that the sampling techniques are applied consistently within a quarter. For example, quarterly samples for a sampling population must use consistent sampling techniques across the quarterly submission period.

- Simple random sampling - selecting a sample size (n) from a population of size (N) in such a way that every case has the same chance of being selected.
- Systematic random sampling - selecting every kth record from a population of size (N) in such a way that a sample size of n is obtained, where k = N/n rounded to the lower digit. The first sample record (i.e., the starting point) must be randomly selected before taking every kth record. This is a two-step process:

a) Randomly select the starting point by choosing a number between one and k using a table of random numbers or a computer-generated random number; and
b) Then select every kth record thereafter until the selection of the sample size is completed.

Each hospital is ultimately responsible that the sampling techniques applied for their hospital adhere to the sampling requirements outlined in this manual. Performance measurement systems are responsible for ensuring that the sampling techniques are applied consistently across their client hospitals.
Monthly Sampling Guidelines
It is important to point out that if a hospital elects to use the monthly sampling guidelines, the hospital is still required to meet the minimum quarterly sampling requirements. A hospital may choose to use a larger sample size than is required. Hospitals whose population size is less than the minimum number of cases per quarter for the measure set cannot sample (i.e., the entire population of cases must be selected). Given the potential for substantial variation in monthly population sizes, the monthly sample sizes should be based on the known or anticipated quarterly population size. When necessary, appropriate oversampling should be employed to ensure that the hospital meets the minimum quarterly sample size requirements. Refer to Table 3 below for guidelines in determining the number of cases that need to be sampled for each population per month per hospital based on the quarterly population size.

Table 3: Sample Size Guidelines per Month per Hospital

Population per Quarter Monthly Sample Size = 80 use all cases 81-100 27 101-125 32 126-150 37 151-175 41 176-200 44 201-225 48 226-250 51 251-275 54 276-300 57 301-325 59 326-350 62 351- 75 64 376-400 66 401-425 68 426-450 70 451-500 73 501-600 79 601-700 83 701-800 87 801-900 90 901-1,000 93 1,001-2,000 108 2,001-3,000 114 3,001-4,000 117 4,001-5,000 119 5,001-10,000 124 10,001-20,000 126	
2a.24 Data Source (*Check the source(s) for which the measure is specified and tested)* Paper medical record/flow-sheet, Electronic administrative data/claims, Electronic Health/Medical Record **2a.25 Data source/data collection instrument (***Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.***):** N/A **2a.26-28 Data source/data collection instrument reference web page URL or attachment:** URL http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1196289981244 **2a.29-31 Data dictionary/code table web page URL or attachment:** URL http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1196289981244 **2a.32-35 Level of Measurement/Analysis** (*Check the level(s) for which the measure is specified and tested)* Facility/Agency, Population: national, Can be measured at all levels **2a.36-37 Care Settings (***Check the setting(s) for which the measure is specified and tested)* Hospital, Ambulatory Care: Emergency Dept, Ambulatory Care: Hospital Outpatient **2a.38-41 Clinical Services** (*Healthcare services being measured, check all that apply*) Clinicians: Nurses, Clinicians: PA/NP/Advanced Practice Nurse, Clinicians: Physicians (MD/DO)	
TESTING/ANALYSIS	
2b. Reliability testing **2b.1 Data/sample** *(description of data/sample and size)***:** Currently under going validation through the CMS	**2b** C☐ P☐

Comment [KP10]: 2b. Reliability testing demonstrates the measure results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

Clinical Data Abstraction Center

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
N/A

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
N/A

M☐ N☐

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: Currently under going validation through the CMS Clinical Data Abstraction Center

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
N/A

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
N/A

2c C☐ P☐ M☐ N☐

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
N/A

2d.2 Citations for Evidence:
N/A

2d.3 Data/sample *(description of data/sample and size)*: N/A

2d.4 Analytic Method *(type analysis & rationale)*:
N/A

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
N/A

2d C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: N/A

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
N/A

2e.3 Testing Results *(risk model performance metrics)*:
N/A

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: N/A

2e C☐ P☐ M☐ N☐ NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: N/A

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
N/A

2f.3 Provide Measure Scores from Testing or Current Use *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

2f C☐ P☐ M☐ N☐

Comment [k11]: 8 Examples of reliability testing include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing may address the data items or final measure score.

Comment [KP12]: 2c. Validity testing demonstrates that the measure reflects the quality of care provided, adequately distinguishing good and poor quality. If face validity is the only validity addressed, it is systematically assessed.

Comment [k13]: 9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) a ... [2]

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND ... [3]

Comment [k15]: 10 Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, sensitivity analyses with and without the exclusion, and variability of exclusions across providers.

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
•an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outc ... [4]

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in trea ... [5]

Comment [KP18]: 2f. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage ... [6]

Q1 2010 Analysis Provider Level
1,050 hospitals submitted 3,164 eligible cases.
Median 68.75 minutes
Min 0 minutes
Max 540 minutes *capped
5th percentile 269 minutes
10th percentile 189 minutes
25th percentile 115 minutes
75th percentile 49 minutes
90th percentile 36 minutes
95th percentile 30 minutes

2g. Comparability of Multiple Data Sources/Methods

2g.1 **Data/sample** *(description of data/sample and size)*: N/A

2g.2 **Analytic Method** *(type of analysis & rationale)*:
N/A

2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*:
N/A

2g
C☐
P☐
M☐
N☐
NA☐

Comment [KP20]: 2g. If multiple data sources/methods are allowed, there is demonstration they produce comparable results.

2h. Disparities in Care

2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: N/A

2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**
N/A

2h
C☐
P☐
M☐
N☐
NA☐

Comment [KP21]: 2h. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);OR rationale/data justifies why stratification is not necessary or not feasible.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*

2

Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?**
Rationale:

2
C☐
P☐
M☐
N☐

3. USABILITY

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)

Eval Rating

3a. Meaningful, Understandable, and Useful Information

3a.1 **Current Use:** In use

3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*):
CMS Hospital Outpatient Department Quality Data Reporting Program
http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1191255879384

3a.3 **If used in other programs/initiatives (***If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years***):**
N/A

Testing of Interpretability (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*)

3a
C☐
P☐
M☐
N☐

Comment [KP22]: 3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audience(s) for both public reporting (e.g., focus group, cognitive testing) and informing quality improvement (e.g., quality improvement initiatives). An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

3a.4 Data/sample *(description of data/sample and size)*: N/A **3a.5 Methods** *(e.g., focus group, survey, QI project)*: N/A **3a.6 Results** *(qualitative and/or quantitative results and conclusions)*: N/A	
3b/3c. Relation to other NQF-endorsed measures **3b.1 NQF # and Title of similar or related measures:**	
(for NQF staff use) **Notes on similar/related endorsed or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?**	**3b** C☐ P☐ M☐ N☐ NA☐
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** **5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** N/A	**3c** C☐ P☐ M☐ N☐ NA☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability?*	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	**Eval Rating**
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.** Pending funding, e-specifications could be developed.	**4b** C☐ P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?**	**4c** C☐ P☐ M☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable

Comment [KP23]: 3b. The measure specifications are harmonized with other measures, and are applicable to multiple levels and settings.

Comment [k24]: 16 Measure harmonization refers to the standardization of specifications for similar measures on the same topic (e.g., *influenza immunization* of patients in hospitals or nursing homes), or related measures for the same target population (e.g., eye exam and HbA1c for *patients with diabetes*), or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are dictated by the evidence. The dimensions of harmonization can include numerator, denominator, exclusions, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.

Comment [KP25]: 3c. Review of existing endorsed measures and measure sets demonstrates that the measure provides a distinctive or additive value to existing NQF-endorsed measures (e.g., provides a more complete picture of quality for a particular condition or aspect of healthcare, is a more valid or efficient way to measure).

Comment [KP26]: 4a. For clinical measures, required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery. (e.g., BP recorded in the electronic record, not abstracted from the record later by other personnel; patient self-assessment tools, e.g., depression scale; lab values, meds, etc.)

Comment [KP27]: 4b. The required data elements are available in electronic sources. If the required data are not in existing electronic sources, a credible, near-term path to electronic collection by most providers is specified and clinical data elements are specified for transition to the electronic health record.

Comment [KP28]: 4c. Exclusions should not require additional data sources beyond what is required for scoring the measure (e.g., numerator and denominator) unless justified as supporting measure validity.

No

N☐
NA☐

4c.2 **If yes, provide justification.**

4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences

4d

4d.1 **Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.**
N/A

C☐
P☐
M☐
N☐

Comment [KP29]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

4e. Data Collection Strategy/Implementation

4e.1 **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:**
Measure is easily collected with limited abstraction burdern.

4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:**
N/A

4e.3 **Evidence for costs:**
N/A

4e.4 **Business case documentation:** N/A

4e

C☐
P☐
M☐
N☐

Comment [KP30]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?

4

Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?**
Rationale:

4

C☐
P☐
M☐
N☐

RECOMMENDATION

(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**

Time-limited ☐

Steering Committee: Do you recommend for endorsement?
Comments:

Y☐
N☐
A☐

CONTACT INFORMATION

Co.1 **Measure Steward (Intellectual Property Owner)**
Co.1 **Organization**
Centers for Medicare & Medicaid Services, 7500 Security Boulevard , Mail Stop S3-01-02, Baltimore, Maryland, 21244-1850

Co.2 **Point of Contact**
Wanda, Govan-Jenkins, MS, MBA, RN, Wanda.Govan-Jenkins@CMS.hhs.gov, 410-786-2699-

Measure Developer If different from Measure Steward
Co.3 **Organization**
Oklahoma Foundation for Medical Quality, 14000 Quail Springs Parkway, Suite 400, Oklahoma City, Oklahoma, 73134-2600

Co.4 **Point of Contact**
Wanda, Govan-Jenkins, MS, MBA, RN, Wanda.Govan-Jenkins@CMS.hhs.gov, 410-786-2699-

Co.5 Submitter If different from Measure Steward POC Rebecca, Jones, MSN, RN, rjones@ofmq.com, 405-840-2891-342, Oklahoma Foundation for Medical Quality
Co.6 Additional organizations that sponsored/participated in measure development
ADDITIONAL INFORMATION
Workgroup/Expert Panel involved in measure development **Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.** N/A
Ad.2 If adapted, provide name of original measure: N/A **Ad.3-5 If adapted, provide original specifications URL or attachment**
Measure Developer/Steward Updates and Ongoing Maintenance **Ad.6 Year the measure was first released:** 2008 **Ad.7 Month and Year of most recent revision:** 07, 2010 **Ad.8 What is your frequency for review/update of this measure?** Bi-annual **Ad.9 When is the next scheduled review/update for this measure?** 01, 2011
Ad.10 Copyright statement/disclaimers: N/A
Ad.11 -13 Additional Information web page URL or attachment: URL http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1196289981244
Date of Submission (*MM/DD/YY*)**:** 12/07/2010

Page 3: [1] Comment [k5] **Karen Pace** **10/5/2009 8:59:00 AM**

4 Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multi-step process, the step with the greatest effect on the desired outcome should be selected as the focus of measurement. For example, although assessment of immunization status and recommending immunization are necessary steps, they are not sufficient to achieve the desired impact on health status – patients must be vaccinated to achieve immunity. This does not preclude consideration of measures of preventive screening interventions where there is a strong link with desired outcomes (e.g., mammography) or measures for multiple care processes that affect a single outcome.

Page 8: [2] Comment [k13] **Karen Pace** **10/5/2009 8:59:00 AM**

9 Examples of validity testing include, but are not limited to: determining if measure scores adequately distinguish between providers known to have good or poor quality assessed by another valid method; correlation of measure scores with another valid indicator of quality for the specific topic; ability of measure scores to predict scores on some other related valid measure; content validity for multi-item scales/tests. Face validity is a subjective assessment by experts of whether the measure reflects the quality of care (e.g., whether the proportion of patients with BP < 140/90 is a marker of quality). If face validity is the only validity addressed, it is systematically assessed (e.g., ratings by relevant stakeholders) and the measure is judged to represent quality care for the specific topic and that the measure focus is the most important aspect of quality for the specific topic.

Page 8: [3] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

2d. Clinically necessary measure exclusions are identified and must be:

- supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;

AND

- a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;

AND

- precisely defined and specified:
  - if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

Page 8: [4] Comment [KP16] **Karen Pace** **10/5/2009 8:59:00 AM**

2e. For outcome measures and other measures (e.g., resource use) when indicated:

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR

rationale/data support no risk adjustment.

Page 8: [5] Comment [k17] **Karen Pace** **10/5/2009 8:59:00 AM**

13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

Page 8: [6] Comment [k19] **Karen Pace** **10/5/2009 8:59:00 AM**

14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.