

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 #: 0160	NQF Project: Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION	
De.1 Measure Title: Beta-blocker prescribed at discharge for AMI	
De.2 Brief description of measure: Percentage of acute myocardial infarction (AMI) patients who are prescribed a beta-blocker at hospital discharge	
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 N/A	
De.4 National Priority Partners Priority Area: Population health	
De.5 IOM Quality Domain: Effectiveness	
De.6 Consumer Care Need: Living with illness	

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. <i>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.</i>	A Y <input type="checkbox"/> N <input type="checkbox"/>
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:	
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	B Y <input type="checkbox"/>

every 3 years. Yes, information provided in contact section	N <input type="checkbox"/>
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ► Purpose: Public reporting, Internal quality improvement Accountability, Payment incentive	C Y <input type="checkbox"/> N <input type="checkbox"/>
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 <input type="checkbox"/> N <input type="checkbox"/>
(for NQF staff use) Have all conditions for consideration been met? Staff Notes to Steward (if submission returned):	Met Y <input type="checkbox"/> N <input type="checkbox"/>
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. <i>Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)</i> 1a. High Impact	Eval Ratin g
(for NQF staff use) Specific NPP goal :	
1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, Leading cause of morbidity/mortality, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 Summary of Evidence of High Impact: In 2010, an estimated 785,000 Americans will have a new coronary event, and approximately 470,000 will have a recurrent event. An estimated additional 195,000 silent first myocardial infarctions occur each year. Approximately every 25 seconds, an American will have a coronary event, and approximately every minute, one will die. In 2004, AMI resulted in 695,000 hospital stays and \$31 billion in health expenditures. The risk of further cardiovascular complications, including recurrent MI, sudden cardiac death, heart failure, stroke, and angina pectoris, among AMI survivors is substantial. 1a.4 Citations for Evidence of High Impact: · Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G, Ferguson TB, Ford E, Furie K, Gillespie C, Go A, Greenlund K, Haase N, Hailpern S, Ho PM, Howard V, Kissela B, Kittner S, Lackland D, Lisabeth L, Marelli A, McDermott MM, Meigs J, Mozaffarian D, Mussolino M, Nichol G, Roger VL, Rosamond W, Sacco R, Sorlie P, Stafford R, Thom T, Wasserthiel-Smoller S, Wong ND, Wylie-Rosett J; on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2010 update: a report from the American Heart Association. <i>Circulation</i> . 2010;121:e46-e215.	1a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
1b. Opportunity for Improvement	1b C <input type="checkbox"/>

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).

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).

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Beta-blockers reduce morbidity and mortality. Hospital performance rates have gradually increased over the years this measure has been reported to the public. Providers understand the importance of sending their patients home on beta-blockers. Ongoing use of this measure will help ensure that high performing providers maintain high performance and the relatively lower performing providers have an impetus to improve.

P
M
N

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:

National performance rates:

2Q09: 98.1%

3Q09: 98.2%

4Q09: 98.3%

1Q10: 98.2%

1b.3 Citations for data on performance gap:

Clinical warehouse data:

2Q09: 101,277 AMI patients, 3,068 hospitals

3Q09: 97,272 AMI patients, 3,040 hospitals

4Q09: 103,296 AMI patients, 3,063 hospitals

1Q10: 105,436 AMI patients, 3,111 hospitals

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

At the univariate analysis level (unadjusted odds ratios), rates ranged from 96.3% for Hispanic/Latinos, to 97.8% for Native-Americans and African-Americans, 98.2% for Asians/Pacific Islanders, and 98.3% for White/Caucasians. The difference from the lowest to the highest rates was 2.0 percentage points. The rate for Caucasians was higher than the rates for all minority groups.

1b.5 Citations for data on Disparities:

2009 Clinical warehouse data (Total 382,023 patients with race not missing): 304,013 Caucasian patients, 40,008 African-American patients, 28,382 Hispanic patients, 7,738 Asian/Pacific Islander patients, and 1,882 Native American patients.

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): Long-term use of beta-blockers for patients who have suffered an acute myocardial infarction reduces mortality and morbidity. Studies have identified a 20% reduction in this risk. Further, there is evidence of effectiveness in broad populations of patients with AMI. National guidelines strongly recommend long-term beta-blocker therapy for the secondary prevention of subsequent cardiovascular events in patients discharged after AMI. The initiation and indefinite continuation of beta-blockers is considered a Class I recommendation in ACC/AHA UA/NSTEMI and STEMI guidelines.

1c.2-3. Type of Evidence: Evidence-based guideline, Randomized controlled trial, Systematic synthesis of research, Meta-analysis

1c.4 Summary of Evidence (as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome):

Beta-blockers lower systolic blood pressure and heart rate (which in turn reduce myocardial oxygen consumption or MVO₂). The benefits of beta-blocker therapy for secondary prevention (i.e. among patients who have experienced a myocardial infarction) are well established. Beta-blockers reduce mortality and morbidity. Data from large trials suggest that therapy should be continued for at least 2 to 3 years. Among patients with ST-segment elevation myocardial infarction (STEMI), the greatest mortality benefit accrues to patients with the greatest baseline risk: those with impaired ventricular function or ventricular arrhythmias and those who have not undergone reperfusion. However, long-term beta-blocker therapy is recommended for all AMI survivors, including those who have undergone revascularization because of evidence of a mortality benefit in such patients.

1c.5 Rating of strength/quality of evidence (also provide narrative description of the rating and by whom):

1c
C
P
M
N

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 th... [1]

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... [2]

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.

ACCF/AHA Task Force on Practice Guidelines: [UA/NSTEMI] Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies; Limited populations evaluated; [STEMI] Level of Evidence A: Data derived from multiple randomized clinical trials or meta-analyses; Multiple populations evaluated.

1c.6 Method for rating evidence: The methodology used by the ACCF/AHA Task Force on Practice Guidelines is fully documented in their publication "Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines" (http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf). The guidelines are based upon a comprehensive assessment, both electronic and manual, of the English-language medical literature. This search focuses on high-quality randomized controlled trials, meta-analyses and systematic reviews, and when applicable observational studies. In some cases where higher quality data is not available, observational studies and case series are also considered. The quality of the design and execution of these studies is determined. When appropriate, data tables are generated from the available literature. After a review of the available literature, the writing committee rates the evidence according to the schemes outlined in their publication.

1c.7 Summary of Controversy/Contradictory Evidence: Aside from avoiding use in patients with clear contraindications to beta-blocker therapy, there is substantial support in existing guidelines for the use of chronic beta-blocker therapy for secondary prevention in patients surviving AMI.

1c.8 Citations for Evidence (other than guidelines):

- Chae CU, Hennekens CH. Beta blockers. In: Hennekens CH, ed. Clinical trials in cardiovascular disease: a companion to Braunwald's Heart Disease. Philadelphia, PA: WB Saunders Co Ltd; 1999:79-94.
- Antman E, Braunwald E. Acute myocardial infarction. In: Braunwald E, Zipes DP, Libby P, eds. Heart disease: a textbook of cardiovascular medicine, 6th ed. Philadelphia, PA: WB Saunders Co Ltd; 2001:1114-1251.
- Brand DA, Newcomer LN, Freiburger A, Tian H. Cardiologists' practices compared with practice guidelines: use of beta-blockade after acute myocardial infarction. J Am Coll Cardiol 1995;26:1432-6.
- Chen J, Radford MJ, Wang Y, Marciniak TA, Krumholz HM. Are beta-blockers effective in elderly patients who undergo coronary revascularization after acute myocardial infarction? Arch Intern Med 2000;160:947-52.
- The Beta-Blocker Pooling Project Research Group. The Beta-Blocker Pooling Project (BBPP): subgroup findings from randomized trials in post infarction patients. Eur Heart J 1988;9:8-16.
- Barron HV, Viskin S, Lundstrom RJ, Wong CC, Swain BE, Truman AF, Selby JV. Effect of beta-adrenergic blocking agents on mortality rate in patients not revascularized after myocardial infarction: data from a large HMO. Am Heart J 1997;134:608-13.
- Yusuf S, Wittes J, Friedman L. Overview of results of randomized clinical trials in heart disease. I. Treatments following myocardial infarction. JAMA. 1988; 260(14):2088:2093.

1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number): [UA/NSTEMI]

5.2.2. Beta Blockers (p. e91)

1. Beta blockers are indicated for all patients recovering from UA/NSTEMI unless contraindicated. (For those at low risk, see Class IIa recommendation below). Treatment should begin within a few days of the event, if not initiated acutely, and should be continued indefinitely.

It is reasonable to prescribe beta blockers to low-risk patients (i.e., normal LV function, revascularized, no high-risk features) recovering from UA/NSTEMI in the absence of absolute contraindications.

[STEMI]

Beta Blockers (p. 236)

It is beneficial to start and continue beta-blocker therapy indefinitely in all patients who have had MI, acute coronary syndrome, or LV dysfunction with or without HF symptoms, unless contraindicated.

1c.10 Clinical Practice Guideline Citation:

- Anderson JL, Adams CD, Antman EM, Bridges CR, Califf RM, Casey DE Jr, et al. ACC/AHA 2007 guidelines for the management of patients with unstable angina/non-ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines for the Management of Patients With Unstable Angina/Non-ST-Elevation Myocardial Infarction): developed in collaboration with the American College of Emergency Physicians, American College of Physicians, Society for Academic Emergency

<p>Medicine, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol. 2007;50:e1-157.</p> <p>Antman EM, Hand M, Armstrong PW, Bates ER, Green LA, Halasyamani LK, et al. 2007 focused update of the ACC/AHA 2004 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 (Writing Group to Review New Evidence and Update the ACC/AHA 2004 Guidelines for the Management of Patients With ST-Elevation Myocardial Infarction). J Am Coll Cardiol. 2008;51:210-47.</p> <p>1c.11 National Guideline Clearinghouse or other URL: http://content.onlinejacc.org/cgi/reprint/51/2/210.pdf, http://content.onlinejacc.org/cgi/reprint/50/7/e1.pdf</p> <p>1c.12 Rating of strength of recommendation (<i>also provide narrative description of the rating and by whom</i>): Rating made by ACCF/AHA Task Force on Practice Guidelines: [UA/NSTEMI and STEMI] Class I recommendation - Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective. Benefit >>> Risk. Procedure/treatment should be performed/administered; [UA/NSTEMI] Class IIa recommendation - Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment. Weight of evidence/opinion is in favor of usefulness/efficacy. Benefit >> Risk. Additional studies with focused objectives needed. It is reasonable to perform procedure/treatment.</p> <p>1c.13 Method for rating strength of recommendation (<i>If different from USPSTF system, also describe rating and how it relates to USPSTF</i>): [UA/NSTEMI and STEMI] The methodology used by the ACCF/AHA Task Force on Practice Guidelines is fully documented in their publication "Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines" (http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf). Recommendations are assigned strength by the Task Force based upon evidence, benefit vs. risk vs. harm, and patient preference. Both the ACCF/AHA Guidelines and the USPSTF assess evidence with respect to two parameters: 1) the magnitude of the benefit, and 2) the certainty of this benefit. However, they use different coding systems. In ascertaining magnitude of the benefit, the ACCF/AHA uses a Class I-III scale and the USPSTF uses a high-moderate-low scale. In determining the certainty of this benefit, the ACCF/AHA uses levels of evidence A-C and USPSTF uses a high-moderate-low scale.</p> <p>1c.14 Rationale for using this guideline over others: The ACCF/AHA guidelines are widely accepted national guidelines that address the therapy of patients with AMI; they use an explicit and transparent methodology; and have thus served as the foundation of national quality measures.</p>	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Importance to Measure and Report</i>?	1
Steering Committee: Was the threshold criterion, <i>Importance to Measure and Report</i>, met? Rationale:	1 Y <input type="checkbox"/> N <input type="checkbox"/>
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, <u>as specified</u> , 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-specs C <input type="checkbox"/>
2a. Precisely Specified	

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.

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).

<p>2a.1 Numerator Statement (<i>Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome</i>): AMI patients who are prescribed a beta-blocker at hospital discharge</p> <p>2a.2 Numerator Time Window (<i>The time period in which cases are eligible for inclusion in the numerator</i>): From hospital arrival to time of hospital discharge.</p> <p>2a.3 Numerator Details (<i>All information required to collect/calculate the numerator, including all codes, logic, and definitions</i>): Refer to http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036: · Section 1 - Data Dictionary Alphabetical Data Dictionary - pages 1-88 through 1-89. · Appendices Appendix C - Medication Tables - pages Appendix C-7 through Appendix C-9. · Section 2 - Measurement Information Section 2.1 - Acute Myocardial Infarction (AMI) - pages AMI-5-1 through AMI-5-5.</p>	P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
<p>2a.4 Denominator Statement (<i>Brief, text description of the denominator - target population being measured</i>): AMI patients (International Classification of Diseases, 9th revision, Clinical Modification [ICD-9-CM] principal diagnosis code of AMI: 410.00, 410.01, 410.10, 410.11, 410.20, 410.21, 410.30, 410.31, 410.40, 410.41, 410.50, 410.51, 410.60, 410.61, 410.70, 410.71, 410.80, 410.81, 410.90, 410.91)</p> <p>2a.5 Target population gender: Female, Male 2a.6 Target population age range: Greater than or equal to 18 years old</p> <p>2a.7 Denominator Time Window (<i>The time period in which cases are eligible for inclusion in the denominator</i>): From hospital arrival to time of hospital discharge.</p> <p>2a.8 Denominator Details (<i>All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions</i>): ICD-9-CM Principal Diagnosis codes: 410.00: Anterolateral wall, acute myocardial infarction-episode of care unspecified 410.01: Anterolateral wall, acute myocardial infarction-initial episode 410.10: Other anterior wall, acute myocardial infarction-episode of care unspecified 410.11: Other anterior wall, acute myocardial infarction-initial episode 410.20: Inferolateral wall, acute myocardial infarction-episode of care unspecified 410.21: Inferolateral wall, acute myocardial infarction-initial episode 410.30: Inferoposterior wall, acute myocardial infarction-episode of care unspecified 410.31: Inferoposterior wall, acute myocardial infarction-initial episode 410.40: Other inferior wall, acute myocardial infarction-episode of care unspecified 410.41: Other inferior wall, acute myocardial infarction-initial episode 410.50: Other lateral wall, acute myocardial infarction-episode of care unspecified 410.51: Other lateral wall, acute myocardial infarction-initial episode 410.60: True posterior wall, acute myocardial infarction-episode of care unspecified 410.61: True posterior wall, acute myocardial infarction-initial episode 410.70: Subendocardial, acute myocardial infarction-episode of care unspecified 410.71: Subendocardial, acute myocardial infarction-initial episode 410.80: Other specified sites, acute myocardial infarction-episode of care unspecified 410.81: Other specified sites, acute myocardial infarction-initial episode 410.90: Unspecified site, acute myocardial infarction-episode of care unspecified 410.91: Unspecified site, acute myocardial infarction-initial episode</p>	
<p>2a.9 Denominator Exclusions (<i>Brief text description of exclusions from the target population</i>): Exclusions</p> <ul style="list-style-type: none"> • <18 years of age • Patients who have a length of stay greater than 120 days • Patients enrolled in clinical trials • Discharged to another hospital 	

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.

- Expired
- Left against medical advice
- Discharged to home for hospice care
- Discharged to a health care facility for hospice care
- Patients with comfort measures only documented
- Patients with a documented reason for no beta-blocker at discharge

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

Refer to

<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036>:

- Section 1 - Data Dictionary | Alphabetical Data Dictionary - pages 1-20 through 1-21, 1-90, 1-98 through 1-104, 1-117, 1-118 through 1-120, 1-204, and 1-327 through 1-330.
- Appendices | Appendix C - Medication Tables PDF - pages Appendix C-7 through Appendix C-9, and Appendix H - Miscellaneous Tables - page Appendix H-5.
- Section 2 - Measurement Information | Section 2.1 - Acute Myocardial Infarction (AMI) - pages AMI-5 plus AMI-5-1 through AMI-5-5.

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions*):

N/A

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: Rate/proportion

2a.20 Interpretation of Score: Better quality = Higher score

2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps*):

Refer to

<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036>: Section 2 - Measurement Information | Section 2.1 - Acute Myocardial Infarction (AMI) - pages AMI-5 plus AMI-5-1 through AMI-5-5.

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

Benchmarks are established using the ABC methodology, based on the actual performance of the top facilities. ABC benchmarks identify superior performance and encourage poorer performers to improve. The methodology is a data-driven, peer-group performance feedback used to positively affect outcomes.

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):*

Patients admitted to the hospital for inpatient acute care with an ICD-9-CM Principal Diagnosis Code for AMI as defined in section 2a.8, a patient age greater than or equal to 18 years, and a length of stay less than or equal to 120 days would be included in the initial patient population and eligible to be sampled.

Monthly Sample Size Based on Population Size (Average monthly initial patient population size: Minimum required sample size):

>= 516: 104

131-515: 20% of Initial Patient Population size

26-130: 26

< 26: 100%

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested*)

Paper medical record/flow-sheet, 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.*):

Centers for Medicare & Medicaid Services (CMS) Abstraction & Reporting Tool (CART). Vendor tools also available.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL
<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1135267770141>

2a.29-31 Data dictionary/code table web page URL or attachment: URL Refer to
<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036>: Section 1 - Data Dictionary | Alphabetical Data Dictionary.

2a.32-35 Level of Measurement/Analysis (Check the level(s) for which the measure is specified and tested)
 Facility/Agency, Population: national, Program: QIO

2a.36-37 Care Settings (Check the setting(s) for which the measure is specified and tested)
 Hospital

2a.38-41 Clinical Services (Healthcare services being measured, check all that apply)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample (description of data/sample and size): CDAC (Clinical Data Abstraction Center) validation sample: 3Q09.

2b.2 Analytic Method (type of reliability & rationale, method for testing):
 CDAC validation sampling involves SDPS selection of sample of 5 cases/quarter across all topics (AMI, HF, Pneumonia, etc.) from each hospital with a minimum of 6 discharges (across all topics) in the Clinical Data Warehouse within 4 months + 15 days following 3Q09. Hospital-abstracted data is compared to CDAC-adjudicated data.

2b.3 Testing Results (reliability statistics, assessment of adequacy in the context of norms for the test conducted):
 Beta-Blocker Prescribed at Discharge - 97.8%
 Clinical Trial - 98.9%
 Comfort Measures Only - 94.3%
 Reason for No Beta-Blocker at Discharge - 77.7%

2b
 C
 P
 M
 N

2c. Validity testing

2c.1 Data/sample (description of data/sample and size): Face validity is regularly assessed with the Technical Expert Panel responsible for reviewing and supporting the measure topic.

2c.2 Analytic Method (type of validity & rationale, method for testing):
 Face validity

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):
 The exclusions of age < 18 years, length of stay > 120 days, and enrollment in a clinical trial are common to the other measures in the AMI measure set, and to the inpatient Hospital Inpatient Quality Reporting Program measure set in general. Patients with documented comfort measures only or those discharged to hospice are appropriate exclusions, as the goal in these cases is palliative care - Therefore, the non-use of beta-blockers is often clinically appropriate. Patients who leave against medical advice or who expire are appropriately

2d
 C
 P
 M
 N
 NA

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

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.

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) 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.

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
 •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 ... [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.

<p>excluded, and it is sensible for those who are discharged to another hospital (where the patient goes on to continue acute care treatment) to be omitted as well. Lastly, there are clinically important contraindications to the use of beta-blockers. Reasons vary, from patient refusal, beta-blocker allergies, and 2nd/3rd degree heart block on ECG, to clinical conditions such as hypotension. In these types of cases, the non-use of beta-blockers should not count against the provider if the clinical reason for not prescribing beta-blockers is documented. All exclusions in this measure (with the exception of the age, length of stay, and clinical trial) are concordant with the current ACC/AHA Clinical Performance Measures for Adults With ST-Elevation and Non-ST-Elevation Myocardial Infarction.</p> <p>2d.2 Citations for Evidence:</p> <ul style="list-style-type: none"> · 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 Infarction). J Am Coll Cardiol. 2008;52:2046 -99. <p>2d.3 Data/sample (description of data/sample and size): Clinical warehouse data: 144,251 AMI patients, 3,503 hospitals, 1Q10.</p> <p>2d.4 Analytic Method (type analysis & rationale): A frequency count was conducted to calculate the percentages outlined in section 2d.5. Frequency counts are a simple, efficient way to determine the occurrence of specific values of a data element in a given data set.</p> <p>2d.5 Testing Results (e.g., frequency, variability, sensitivity analyses): Rates of Exclusion:</p> <ul style="list-style-type: none"> · Patients with comfort measures only documented: 5.8% · Patients enrolled in clinical trials: .5% · Discharged/transferred to another hospital for inpatient care, discharged/transferred to a federal health care facility, discharged/transferred to hospice, expired, or left against medical advice or discontinued care: 14.7% · Patients with a documented reason for no beta-blocker at discharge: 5.9% 	
<p>2e. Risk Adjustment for Outcomes/ Resource Use Measures</p> <p>2e.1 Data/sample (description of data/sample and size): N/A</p> <p>2e.2 Analytic Method (type of risk adjustment, analysis, & rationale): N/A</p> <p>2e.3 Testing Results (risk model performance metrics): N/A</p> <p>2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: N/A</p>	<p>2e</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>2f. Identification of Meaningful Differences in Performance</p> <p>2f.1 Data/sample from Testing or Current Use (description of data/sample and size): Clinical warehouse data: 2Q09: 101,277 AMI patients, 3,068 hospitals 3Q09: 97,272 AMI patients, 3,040 hospitals 4Q09: 103,296 AMI patients, 3,063 hospitals 1Q10: 105,436 AMI patients, 3,111 hospitals</p> <p>2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (type of analysis & rationale): Analysts review quarterly benchmarks established (using the ABC methodology) and trends to identify differences in performance scores and investigate the possible causes. ABC benchmarks identify superior performance and encourage poorer performers to improve. The methodology is a data-driven, peer-group</p>	<p>2f</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>

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 outcome (but not disparities in care) and are present at start of care; ^{Error! Bookmark not defined.} OR rationale/data support no risk adjustment.

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 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.

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 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.

<p>performance feedback used to positively affect outcomes. If measure specifications (algorithms, data elements) are found to cause the difference in performance, they are reviewed for possible updates.</p> <p>2f.3 Provide Measure Scores from Testing or Current Use (<i>description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance</i>): National performance rates: 2Q09: 98.1% (benchmark 99.9%) 3Q09: 98.2% (benchmark 100.0%) 4Q09: 98.3% (benchmark 99.9%) 1Q10: 98.2% (benchmark 100.0%)</p>	
<p>2g. Comparability of Multiple Data Sources/Methods</p> <p>2g.1 Data/sample (<i>description of data/sample and size</i>): Both paper records and electronic health records can be used to collect data. Some allowances have been made as facilities incorporate EHRs in their facilities because vendors do not utilize identical data fields, but customize products according to facility need and preferences.</p> <p>2g.2 Analytic Method (<i>type of analysis & rationale</i>): No tests have been performed on this measure to determine comparability of sources (paper medical record vs. EHR).</p> <p>2g.3 Testing Results (<i>e.g., correlation statistics, comparison of rankings</i>): N/A</p>	<p>2g C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/> <input type="checkbox"/></p>
<p>2h. Disparities in Care</p> <p>2h.1 If measure is stratified, provide stratified results (<i>scores by stratified categories/cohorts</i>): Not stratified, but results according to race, sex, etc can be determined.</p> <p>2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: Since the preliminary univariate analyses suggest potential disparities (the largest difference is greater than or equal to 2.0 percentage points as described in 1b.4), further analyses are needed to control for the simultaneous effect of other potential factors such as age, gender, comorbidity, and hospital characteristics and to take into account the correlation/cluster effect of patients discharged from the same hospitals.</p>	<p>2h C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/> <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Scientific Acceptability of Measure Properties</i>?</p>	<p>2</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Scientific Acceptability of Measure Properties</i>, met? Rationale:</p>	<p>2 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>3. USABILITY</p>	
<p>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)</p>	<p>Eval Ratin g</p>
<p>3a. Meaningful, Understandable, and Useful Information</p> <p>3a.1 Current Use: In use</p> <p>3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large) (<i>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</i>): Hospital Inpatient Quality Reporting Program: . http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2</p>	<p>3a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

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

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.

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.

<p>&cid=1138115987129 http://www.hospitalcompare.hhs.gov/</p> <p>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): Hospital Inpatient Quality Reporting Program (Measures can be used by individual hospitals for internal quality improvement):</p> <p>http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1138115987129 http://www.hospitalcompare.hhs.gov/</p> <p>Additionally, the Joint Commission also uses this measure for accreditation.</p> <p>Testing of Interpretability (Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement)</p> <p>3a.4 Data/sample (description of data/sample and size): Unknown. [Feedback on the Hospital Compare website (used for public reporting) is collected through another contractor.]</p> <p>3a.5 Methods (e.g., focus group, survey, QI project): Voluntary electronic survey by visitors to website.</p> <p>3a.6 Results (qualitative and/or quantitative results and conclusions): Not available.</p>	
<p>3b/3c. Relation to other NQF-endorsed measures</p> <p>3b.1 NQF # and Title of similar or related measures: NQF #0613: MI - Use of Beta Blocker Therapy</p>	
<p>(for NQF staff use) Notes on similar/related endorsed or submitted measures:</p>	
<p>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 <u>or</u> different topic but same target population): 3b.2 Are the measure specifications harmonized? If not, why? No, this measure's specifications are not harmonized with NQF #0613 measure specifications, as the latter's measure population uses the outpatient setting and includes patients diagnosed with MI at anytime in the past. This measure is concentrated on care of the AMI patient who is admitted for inpatient care; a completely different focus in terms of setting and care. NQF #0613 does provide for the exclusion of patients with an allergy to beta-blockers in the past or those with documentation of heart block, similar to this measure, but it also automatically excludes patients with asthma, COPD, bradycardia, hypotension, aortic stenosis, evidence of metastatic disease or active treatment of malignancy (chemotherapy or radiation therapy) in the last 6 months, patients who have been in a skilled nursing facility in the last 3 months, patients on peripheral artery disease medications, and heart transplant patients - Conditions which our team believes are relative contraindications which require that the physician specifically document a linkage to the non-use of beta-blockers (vs. automatic exclusion).</p>	<p>3b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>3c. Distinctive or Additive Value 3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures: No NQF-endorsed measures with same topic and target population.</p> <p>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: No NQF-endorsed measures with same topic and target population.</p>	<p>3c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Usability?</p>	<p>3</p>
<p>Steering Committee: Overall, to what extent was the criterion, Usability, met? Rationale:</p>	<p>3 C <input type="checkbox"/></p>

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).

	<input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> 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 Ratin g
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? Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), 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 <input type="checkbox"/> C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N
4b. Electronic Sources	
4b.1 Are all the data elements available electronically? (<i>elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims</i>) No	4b <input type="checkbox"/> C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N
4b.2 If not, specify the near-term path to achieve electronic capture by most providers. Retooling work with HHS is expected to be completed in 2011.	
4c. Exclusions	
4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications? No	4c <input type="checkbox"/> C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA
4c.2 If yes, provide justification.	
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences	
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. 1. Since the time of last NQF endorsement (May 2007), the HeartCare measures team met with other topic teams within the Hospital Inpatient Quality Reporting Program (namely, children's asthma and surgical care) to examine the medication constructs being used. The measure designs at that time automatically excluded patients with a documented contraindication or reason to a medication from the measure, regardless of whether the medication ended up being prescribed. That type of design was resulting in a substantial amount of "false exclusions" from the measure. The decision was made to rearrange the measure such that patients who were prescribed the medication would remain in the measure (i.e., be included in the numerator) when a reason for not prescribing the medication was documented, effective with April 1, 2009 discharges. It is believed that the number of false exclusions has significantly decreased as a result. 2. Because the denominator exclusion "Patients with a documented reason for no beta-blocker at discharge" allows for any physician/advance practice nurse/physician assistant/pharmacist-documented "other reason" for not prescribing a beta-blocker at discharge to count as an exclusion, overuse of this exclusion has the potential for distorting performance rates. However, overall trends in measure numerator and denominator counts do not suggest obvious gaming of the measure. There has been no increasing trend in the use of this reason data element. Nevertheless, exclusion rates for this measure will continue to be monitored for consistency, from quarter to quarter. 3. The data elements used in this measure are closely tracked. Questions submitted by abstractors are recorded, and trends related to published abstraction guidelines and disagreements over measure inclusions and exclusions in general are discussed in-depth every 6 months. Revisions in measure specifications, including data element definitions, are made as issues surface (e.g., how to handle documentation of a hold on a beta-blocker at discharge or a planned delay to start a beta-blocker after discharge, what constitutes acceptable physician documentation of a reason for not prescribing beta-blockers). The frequency of	4d <input type="checkbox"/> C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N

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.

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

<p>questions pertaining to each data element are tracked by the Hospital Inpatient Quality Reporting Program QIOSC. Clearly the number of questions a data element receives is another indication of how difficult the specifications for the measure might be. Frequency reports are reviewed regularly, to help identify where issues in data element definitions may exist. Of note, in an August 2010 report run by the Hospital Inpatient Quality Reporting Program QIOSC, the number of questions about the abstraction of the two data elements unique to this measure, Beta-Blocker Prescribed at Discharge and Reason for No Beta-Blocker at Discharge, amounted to 28, only 6.1% of the total 458 Quest questions received for AMI for that month. Lastly, CDAC validation reports (which compare hospital data to CDAC data) and internal CDAC abstractor accuracy reports are monitored, to ensure good quality data. In sum, issues which may surface in questions submitted by users and CDAC validation/accuracy reports will continue to be closely monitored to identify any additional problems, and revisions will be made if warranted.</p>	
<p>4e. Data Collection Strategy/Implementation</p> <p>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: The reordering of the "medication prescribed" and "reason for no medication" specifications done for April 1, 2009+ discharges (as described in section 4d.1) reduces abstraction burden. Abstractors no longer have to do an exhaustive search for acceptable reasons for not prescribing beta-blockers at discharge in cases where the patient was prescribed a beta-blocker, saving valuable abstraction time.</p> <p>4e.2 Costs to implement the measure (<i>costs of data collection, fees associated with proprietary measures</i>): Varies according to data collection method (use of vendor) and type of abstractor used to collect clinical data. We have not received feedback that this measure has caused undue burden to the facilities collecting data.</p> <p>4e.3 Evidence for costs: N/A</p> <p>4e.4 Business case documentation: N/A</p>	<p>4e C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Feasibility</i>?</p>	<p>4</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Feasibility</i>, met? Rationale:</p>	<p>4 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
RECOMMENDATION	
<p>(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.</p>	<p>Time-limited <input type="checkbox"/></p>
<p>Steering Committee: Do you recommend for endorsement? Comments:</p>	<p>Y <input type="checkbox"/> N <input type="checkbox"/> A <input type="checkbox"/></p>
CONTACT INFORMATION	
<p>Co.1 Measure Steward (Intellectual Property Owner) Co.1 <u>Organization</u> Centers for Medicare & Medicaid Services, 7500 Security Boulevard , Baltimore, Maryland, 21244-1850</p> <p>Co.2 <u>Point of Contact</u> Kristie, Baus, RN, MS, kristie.baus@cms.hhs.gov, 410-786-8161-</p>	
<p>Measure Developer If different from Measure Steward</p>	

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).

<p>Co.3 Organization Centers for Medicare & Medicaid Services, 7500 Security Boulevard, Baltimore, Maryland, 21244-1850</p>
<p>Co.4 Point of Contact Kristie, Baus, RN, MS, kristie.baus@cms.hhs.gov, 410-786-8161-</p>
<p>Co.5 Submitter If different from Measure Steward POC Jo, DeBuhr, RN, BSN, broncosrule@att.net, 303-457-3195-, OFMQ</p>
<p>Co.6 Additional organizations that sponsored/participated in measure development The Joint Commission</p>

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.
 This measure is reviewed and maintained by the Heart Care Technical Expert Panel. Quarterly teleconferences are held to discuss issues pertinent to this measure (and its specifications) and potential revisions. Current members:
 Frederick Masoudi, MD, MSPH Workgroup Chair: Denver Health Medical Center, University of Colorado at Denver and Health Sciences Center
 Don Casey, MD, MPH, MBA: VP Quality and Chief Medical Officer, Atlantic Health, Rep. of the American College of Physicians
 Elizabeth Delong, PhD: Professor and Chair, Duke University, Biostatistics and Bioinformatics, Co-Director, Outcomes Research and Assessment
 Joseph Drozda, MD: Clinical Investigator, Mercy Health Research, Executive Committee Member, PCPI, Rep. of American Medical Association
 John P. Erwin, III: Professor of Medicine, Co-Director, Cardiovascular Fellowship Program, Hospital Champion, Acute Myocardial Infarction Quality Improvement, Scott and White Hospital and Clinic
 Kerri Fei: Senior Policy Analyst, Measure Development Operations, American Medical Association
 Susan Fitzgerald, RN, MS: Associate Director, Science and Quality, American College of Cardiology
 Gary Francis, MD: Professor of Medicine, University of Minnesota, Rep. of Heart Failure Society of America
 David C. Goff, MD, PhD: Professor and Chair, Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine
 Kathleen Grady, CNS: Administrative Director, Center for Heart Failure, Bluhm Cardiovascular Institute Division of Cardiothoracic Surgery, Northwestern Memorial Hospital
 Darryl Gray, MD: Medical Officer, Agency for Healthcare Research and Quality
 Lee Green, MD: Professor, University of Michigan Medical School
 Ed Havranek, MD: Professor of Medicine, Denver Health Medical Center, University of Colorado School of Medicine
 Paul A. Heidenreich: Assistant Professor of Medicine, Associate Professor by courtesy of Health Research and Policy at the VA Palo Alto Health Care System and CHP/PCOR Fellow
 Alice C. Jacobs, MD: Professor of Medicine, Director, Cardiac Cath Lab, Boston University Medical Center
 Marvin Konstam, MD: Director, Cardiovascular Center, Tufts Medical Center, Rep. of Heart Failure Society of America
 Harlan Krumholz, MD: Harold H. Hines, Jr. Professor of Medicine and Epidemiology and Public Health, Yale University School of Medicine
 Jerod Loeb, PhD: Executive Vice President, Quality Measurement & Research, The Joint Commission
 Ann [Hiniker] Loth, RN, MS, CNS: Certified Clinical Nurse Specialist, Mayo Foundation
 Joseph Messer, MD, MACC: Professor of Medicine, Rush University Medical Center, Rep. of American Medical Association
 Eric Peterson, MD, MPH: Professor of Medicine, Director Cardiovascular Research, Duke Clinical Research Institute, Duke University Medical Center
 Martha Radford, MD: Chief Quality Officer, Professor of Medicine, New York University School of Medicine
 Rose Marie Robertson, MD: Chief Science Officer, American Heart Association
 John Rumsfeld, MD, PhD, FACC, FAHA: Staff Cardiologist, Cardiovascular Outcomes Researcher, Denver Veterans Affairs Medical Center
 David Shahian, MD: Research Director, Center for Quality and Safety, Massachusetts General Hospital
 Melanie Shahriary, RN, BSN: Associate Director, Performance Measures and Data Standards, American College of Cardiology
 John Spertus, MD, MPH, FACC: Director of Cardiovascular Education and Outcomes Research, Mid America Heart

<p>Institute, University of Missouri Samantha Tierney: Senior Policy Analyst I, American Medical Association Gayle Whitman, PhD, RN, FAAN, FAHA: Sr Vice President, Office of Science Operations, American Heart Association Janet Wright, MD, FACC: Senior Vice President for Science and Quality, American College of Cardiology Contractor Staff: Dale Bratzler, DO, MPH: CEO, Principal Clinical Coordinator, Oklahoma Foundation for Medical Quality Jo DeBuhr, RN: Project Specialist, AMI/HF Inpatient Measures, Oklahoma Foundation for Medical Quality/Colorado Foundation for Medical Care Chris Leber, RN: Project Specialist, AMI/HF Inpatient Measures, Oklahoma Foundation for Medical Quality/Colorado Foundation for Medical Care CMS Staff: Kristie Baus, MS, RN: Government Task Leader, Centers for Medicare and Medicaid Services David Nilasena, MD: Chief Medical Officer, Region VI, Centers for Medicare and Medicaid</p>
<p>Ad.2 If adapted, provide name of original measure: N/A Ad.3-5 If adapted, provide original specifications URL or attachment</p>
<p>Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 Year the measure was first released: 1999 Ad.7 Month and Year of most recent revision: 10, 2010 Ad.8 What is your frequency for review/update of this measure? Every 6 months Ad.9 When is the next scheduled review/update for this measure? 07, 2011</p>
<p>Ad.10 Copyright statement/disclaimers:</p>
<p>Ad.11 -13 Additional Information web page URL or attachment:</p>
<p>Date of Submission (MM/DD/YY): 12/27/2010</p>

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:
 - o Intermediate outcome - evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
 - o Process - 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).
 - o Structure - 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.
 - o Patient 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.
 - o Access - evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
 - o Efficiency - 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.

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.

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).

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 #: 0142	NQF Project: Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION	
De.1 Measure Title: Aspirin prescribed at discharge for AMI	
De.2 Brief description of measure: Percentage of acute myocardial infarction (AMI) patients who are prescribed aspirin at hospital discharge	
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 N/A	
De.4 National Priority Partners Priority Area: Population health	
De.5 IOM Quality Domain: Effectiveness	
De.6 Consumer Care Need: Living with illness	

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
<p>A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. <i>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.</i></p> <p>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</p> <p>A.2 Indicate if Proprietary Measure (as defined in measure steward agreement):</p> <p>A.3 Measure Steward Agreement: Government entity and in the public domain - no agreement necessary</p> <p>A.4 Measure Steward Agreement attached:</p>	<p>A</p> <p>Y <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
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	<p>B</p> <p>Y <input type="checkbox"/></p>

every 3 years. Yes, information provided in contact section	N <input type="checkbox"/>
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ► Purpose: Public reporting, Internal quality improvement Accountability, Payment incentive	C Y <input type="checkbox"/> N <input type="checkbox"/>
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 <input type="checkbox"/> N <input type="checkbox"/>
(for NQF staff use) Have all conditions for consideration been met? Staff Notes to Steward (if submission returned):	Met Y <input type="checkbox"/> N <input type="checkbox"/>
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. <i>Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)</i> 1a. High Impact	Eval Ratin g
(for NQF staff use) Specific NPP goal :	
1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, Leading cause of morbidity/mortality, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 Summary of Evidence of High Impact: In 2010, an estimated 785,000 Americans will have a new coronary event, and approximately 470,000 will have a recurrent event. An estimated additional 195,000 silent first myocardial infarctions occur each year. Approximately every 25 seconds, an American will have a coronary event, and approximately every minute, one will die. In 2004, AMI resulted in 695,000 hospital stays and \$31 billion in health expenditures. The risk of further cardiovascular complications, including recurrent MI, sudden cardiac death, heart failure, stroke, and angina pectoris, among AMI survivors is substantial. 1a.4 Citations for Evidence of High Impact: · Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G, Ferguson TB, Ford E, Furie K, Gillespie C, Go A, Greenlund K, Haase N, Hailpern S, Ho PM, Howard V, Kissela B, Kittner S, Lackland D, Lisabeth L, Marelli A, McDermott MM, Meigs J, Mozaffarian D, Mussolino M, Nichol G, Roger VL, Rosamond W, Sacco R, Sorlie P, Stafford R, Thom T, Wasserthiel-Smoller S, Wong ND, Wylie-Rosett J; on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2010 update: a report from the American Heart Association. <i>Circulation</i> . 2010;121:e46-e215.	1a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
1b. Opportunity for Improvement	1b C <input type="checkbox"/>

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).

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).

<p>1b.1 Benefits (improvements in quality) envisioned by use of this measure: Aspirin use reduces the risk of death. Hospital performance rates have gradually increased over the years this measure has been reported to the public. Providers understand the importance of sending their patients home on aspirin. Ongoing use of this measure will help ensure that high performing providers maintain high performance and the relatively lower performing providers have an impetus to improve.</p> <p>1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers: National performance rates: 2Q09: 98.3% 3Q09: 98.4% 4Q09: 98.5% 1Q10: 98.5%</p> <p>1b.3 Citations for data on performance gap: Clinical warehouse data: 2Q09: 103,335 AMI patients, 3,057 hospitals 3Q09: 99,874 AMI patients, 3,019 hospitals 4Q09: 105,659 AMI patients, 3,062 hospitals 1Q10: 107,852 AMI patients, 3,096 hospitals</p> <p>1b.4 Summary of Data on disparities by population group: At the univariate analysis level (unadjusted odds ratios), rates ranged from 96.5% for Hispanic/Latinos, to 97.4% for African-Americans, 98.0 for Asians/Pacific Islanders, 98.5 for White/Caucasians, and 98.6% for Native Americans. The difference from the lowest to the highest rates was 2.1 percentage points. The rate for Caucasians was higher than the rates for minority groups except Native-Americans.</p> <p>1b.5 Citations for data on Disparities: 2009 Clinical warehouse data (Total 389,674 patients with race not missing): 310,489 Caucasian patients, 40,591 African-American patients, 28,805 Hispanic patients, 7,854 Asian/Pacific Islander patients, and 1,935 Native American patients.</p>	<p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>1c. Outcome or Evidence to Support Measure Focus</p> <p>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): Aspirin therapy in patients who have suffered an acute myocardial infarction reduces the risk of adverse events and mortality. Studies have demonstrated that aspirin can reduce this risk by 20%. National guidelines strongly recommend long-term aspirin for the secondary prevention of subsequent cardiovascular events in eligible older patients discharged after AMI. The initiation and indefinite continuation of aspirin is considered a Class I recommendation in ACC/AHA UA/NSTEMI and STEMI guidelines.</p> <p>1c.2-3. Type of Evidence: Evidence-based guideline, Randomized controlled trial, Systematic synthesis of research, Meta-analysis</p> <p>1c.4 Summary of Evidence (as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome): Some of the strongest evidence available about the long-term benefits of therapy in patients with acute coronary events pertains to ASA. By irreversibly inhibiting COX-1 within platelets, ASA prevents the formation of thromboxane A2, thereby diminishing platelet aggregation. This platelet inhibition is the plausible mechanism for the clinical benefit of ASA, both because it is fully present with low doses of ASA and because platelets represent one of the principal participants in thrombus formation after plaque disruption. Among clinical investigations with ASA, trials in STEMI and NSTEMI have consistently documented a striking benefit of ASA compared with placebo independent of the differences in study design, such as time of entry after the acute phase, duration of follow-up, and dose used. The protective effect of ASA has been sustained for at least 1 to 2 years in clinical trials in UA/NSTEMI. Studies in patients with prior MI, stroke, or transient ischemic attack also suggest significant benefit during the first 2 years of therapy.</p> <p>1c.5 Rating of strength/quality of evidence (also provide narrative description of the rating and by whom):</p>	<p>1c <input type="checkbox"/></p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>

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 th... [1]

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 ... [2]

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.

ACCF/AHA Task Force on Practice Guidelines, Level of Evidence A: [UA/NSTEMI] Data derived from multiple randomized trials or meta-analyses, Multiple populations evaluated; [STEMI] Data derived from multiple randomized clinical trials or meta-analyses.

1c.6 Method for rating evidence: [UA/NSTEMI] The methodology used by the ACCF/AHA Task Force on Practice Guidelines is fully documented in their publication "Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines" (http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf). The guidelines are based upon a comprehensive assessment, both electronic and manual, of the English-language medical literature. This search focuses on high-quality randomized controlled trials, meta-analyses and systematic reviews, and when applicable observational studies. In some cases where higher quality data is not available, observational studies and case series are also considered. The quality of the design and execution of these studies is determined. When appropriate, data tables are generated from the available literature. After a review of the available literature, the writing committee rates the evidence according to the schemes outlined in their publication.

[STEMI] The method of rating evidence used by the Writing Committee on the Management of Patients with ST-Elevation Myocardial Infarction in 2004 is not as well documented, but is implicitly consistent with the approach described in the ACCF/AHA methodology manual. Following comprehensive searching of the scientific and medical literature on AMI, with special emphasis on STEMI, the writing committee weighed the strength of evidence for or against a particular treatment or procedure. A level of evidence rating of "A" was given when multiple (3-5) population risk strata were evaluated (data available from clinical trials or registries about the usefulness/efficacy in different sub-populations, such as gender, age, history of diabetes, history of prior MI, history of heart failure, and prior aspirin use.) and there was general consistency of direction and magnitude of effect.

1c.7 Summary of Controversy/Contradictory Evidence: Aside from avoiding use in patients with clear contraindications to aspirin therapy, there is substantial support in existing guidelines for the use of chronic aspirin therapy for secondary prevention in patients surviving AMI.

1c.8 Citations for Evidence (other than guidelines):

- Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy—I: prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients (published erratum appears in BMJ 1994;308:1540). BMJ 1994;308:81-106.
- Théroux P, Ouimet H, McCans J, et al. Aspirin, heparin, or both to treat acute unstable angina. N Engl J Med 1988;319:1105-11.
- The RISC Group. Risk of myocardial infarction and death during treatment with low dose aspirin and intravenous heparin in men with unstable coronary artery disease. Lancet 1990;336:827-30.
- Randomized trial of intravenous streptokinase, oral aspirin, both or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Lancet. 1988 Aug 13;2(8607):349-60.

1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number):

3.2.1. Antiplatelet Therapy Recommendations (p. e45)

1. Aspirin should be administered to UA/NSTEMI patients as soon as possible after hospital presentation and continued indefinitely in patients not known to be intolerant of that medication.

6.3.1.6.8.2.1. Aspirin (p. e73)

A daily dose of aspirin (initial dose of 162 to 325 mg orally; maintenance dose of 75 to 162 mg) should be given indefinitely after STEMI to all patients without a true aspirin allergy.

1c.10 Clinical Practice Guideline Citation: [3.2.1.] Anderson JL, Adams CD, Antman EM, Bridges CR, Califf RM, Casey DE Jr, et al. ACC/AHA 2007 guidelines for the management of patients with unstable angina/non-ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines for the Management of Patients With Unstable Angina/Non-ST-Elevation Myocardial Infarction): developed in collaboration with the American College of Emergency Physicians, American College of Physicians, Society for Academic Emergency Medicine, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol. 2007;50:e1-157.

[6.3.1.6.8.2.1] Antman EM, Anbe DT, Armstrong PW, Bates ER, Green LA, Hand M, et al. ACC/AHA guidelines

<p>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.</p> <p>1c.11 National Guideline Clearinghouse or other URL: [3.2.1.] http://content.onlinejacc.org/cgi/reprint/50/7/e1.pdf, [6.3.1.6.8.2.1] http://assets.cardiosource.com/STEMI_2004.pdf</p> <p>1c.12 Rating of strength of recommendation (<i>also provide narrative description of the rating and by whom</i>): Rating made by ACCF/AHA Task Force on Practice Guidelines: [UA/NSTEMI] Class I recommendation - Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective. Benefit >>> Risk. Procedure/treatment should be performed/administered; [STEMI] Class I recommendation - Conditions for which there is evidence and/or general agreement that a given procedure or treatment is beneficial, useful, and effective.</p> <p>1c.13 Method for rating strength of recommendation (<i>If different from USPSTF system, also describe rating and how it relates to USPSTF</i>): [UA/NSTEMI] The methodology used by the ACCF/AHA Task Force on Practice Guidelines is fully documented in their publication "Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines" (http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf). Recommendations are assigned strength by the Task Force based upon evidence, benefit vs. risk vs. harm, and patient preference. [STEMI] The method of rating the strength of a recommendation used by the Writing Committee on the Management of Patients with ST-Elevation Myocardial Infarction in 2004 is not as well documented but is implicitly consistent with the approach described in the ACCF/AHA methodology manual. In sum, strength is assigned based on examination of evidence and careful assessment of benefit vs. risk. Both the ACCF/AHA Guidelines and the USPSTF assess evidence with respect to two parameters: 1) the magnitude of the benefit, and 2) the certainty of this benefit. However, they use different coding systems. In ascertaining magnitude of the benefit, the ACCF/AHA uses a Class I-III scale and the USPSTF uses a high-moderate-low scale. In determining the certainty of this benefit, the ACCF/AHA uses levels of evidence A-C and USPSTF uses a high-moderate-low scale.</p> <p>1c.14 Rationale for using this guideline over others: The ACCF/AHA guidelines are widely accepted national guidelines that address the therapy of patients with AMI; they use an explicit and transparent methodology; and have thus served as the foundation of national quality measures.</p>	
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 <input type="checkbox"/> N <input type="checkbox"/>
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, <u>as specified</u> , produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Ratin g
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-spec
2a. Precisely Specified	C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
2a.1 Numerator Statement (<i>Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome</i>): AMI patients who are prescribed aspirin at hospital discharge	

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.

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).

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator*):
 From hospital arrival to time of hospital discharge

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions*):

Refer to

<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036>:

- Section 1 - Data Dictionary | Alphabetical Data Dictionary - pages 1-75 through 1-76.
- Appendices | Appendix C - Medication Tables - pages Appendix C-3 through Appendix C-6.
- Section 2 - Measurement Information | Section 2.1 - Acute Myocardial Infarction (AMI) - pages AMI-2-1 through AMI-2-5.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured*):

AMI patients (International Classification of Diseases, 9th revision, Clinical Modification [ICD-9-CM] principal diagnosis code of AMI: 410.00, 410.01, 410.10, 410.11, 410.20, 410.21, 410.30, 410.31, 410.40, 410.41, 410.50, 410.51, 410.60, 410.61, 410.70, 410.71, 410.80, 410.81, 410.90, 410.91)

2a.5 Target population gender: Female, Male

2a.6 Target population age range: Greater than or equal to 18 years old

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator*):
 From hospital arrival to time of hospital discharge

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions*):

ICD-9-CM Principal Diagnosis codes:

- 410.00: Anterolateral wall, acute myocardial infarction-episode of care unspecified
- 410.01: Anterolateral wall, acute myocardial infarction-initial episode
- 410.10: Other anterior wall, acute myocardial infarction-episode of care unspecified
- 410.11: Other anterior wall, acute myocardial infarction-initial episode
- 410.20: Inferolateral wall, acute myocardial infarction-episode of care unspecified
- 410.21: Inferolateral wall, acute myocardial infarction-initial episode
- 410.30: Inferoposterior wall, acute myocardial infarction-episode of care unspecified
- 410.31: Inferoposterior wall, acute myocardial infarction-initial episode
- 410.40: Other inferior wall, acute myocardial infarction-episode of care unspecified
- 410.41: Other inferior wall, acute myocardial infarction-initial episode
- 410.50: Other lateral wall, acute myocardial infarction-episode of care unspecified
- 410.51: Other lateral wall, acute myocardial infarction-initial episode
- 410.60: True posterior wall, acute myocardial infarction-episode of care unspecified
- 410.61: True posterior wall, acute myocardial infarction-initial episode
- 410.70: Subendocardial, acute myocardial infarction-episode of care unspecified
- 410.71: Subendocardial, acute myocardial infarction-initial episode
- 410.80: Other specified sites, acute myocardial infarction-episode of care unspecified
- 410.81: Other specified sites, acute myocardial infarction-initial episode
- 410.90: Unspecified site, acute myocardial infarction-episode of care unspecified
- 410.91: Unspecified site, acute myocardial infarction-initial episode

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population*): Exclusions:

- <18 years of age
- Patients who have a length of stay greater than 120 days
- Patients enrolled in clinical trials
- Discharged to another hospital
- Expired
- Left against medical advice
- Discharged to home for hospice care

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.

<ul style="list-style-type: none"> • Discharged to a health care facility for hospice care • Patients with comfort measures only documented • Patients with a documented reason for no aspirin at discharge <p>2a.10 Denominator Exclusion Details (<i>All information required to collect exclusions to the denominator, including all codes, logic, and definitions</i>): Refer to http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036: <ul style="list-style-type: none"> · Section 1 - Data Dictionary Alphabetical Data Dictionary - pages 1-20 through 1-21, 1-69 through 1-71, 1-90, 1-98 through 1-104, 1-117, 1-118 through 1-120, 1-204, and 1-321 through 1-323. · Appendices Appendix C - Medication Tables PDF - pages Appendix C-3 through Appendix C-6 plus Appendix C-9, and Appendix H - Miscellaneous Tables - page Appendix H-5. · Section 2 - Measurement Information Section 2.1 - Acute Myocardial Infarction (AMI) - pages AMI-5 plus AMI-2-1 through AMI-2-5. </p>
<p>2a.11 Stratification Details/Variables (<i>All information required to stratify the measure including the stratification variables, all codes, logic, and definitions</i>): N/A</p>
<p>2a.12-13 Risk Adjustment Type: No risk adjustment necessary</p>
<p>2a.14 Risk Adjustment Methodology/Variables (<i>List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method</i>): N/A</p>
<p>2a.15-17 Detailed risk model available Web page URL or attachment:</p>
<p>2a.18-19 Type of Score: Rate/proportion 2a.20 Interpretation of Score: Better quality = Higher score 2a.21 Calculation Algorithm (<i>Describe the calculation of the measure as a flowchart or series of steps</i>): Refer to http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036: Section 2 - Measurement Information Section 2.1 - Acute Myocardial Infarction (AMI) - pages AMI-5 plus AMI-2-1 through AMI-2-5.</p>
<p>2a.22 Describe the method for discriminating performance (<i>e.g., significance testing</i>): Benchmarks are established using the ABC methodology, based on the actual performance of the top facilities. ABC benchmarks identify superior performance and encourage poorer performers to improve. The methodology is a data-driven, peer-group performance feedback used to positively affect outcomes.</p>
<p>2a.23 Sampling (Survey) Methodology <i>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):</i> Patients admitted to the hospital for inpatient acute care with an ICD-9-CM Principal Diagnosis Code for AMI as defined in section 2a.8, a patient age greater than or equal to 18 years, and a length of stay less than or equal to 120 days would be included in the initial patient population and eligible to be sampled. Monthly Sample Size Based on Population Size (Average monthly initial patient population size: Minimum required sample size): >= 516: 104 131-515: 20% of Initial Patient Population size 26-130: 26 < 26: 100%</p>
<p>2a.24 Data Source (<i>Check the source(s) for which the measure is specified and tested</i>) Paper medical record/flow-sheet, Electronic Health/Medical Record</p>
<p>2a.25 Data source/data collection instrument (<i>Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.</i>): Centers for Medicare & Medicaid Services (CMS) Abstraction & Reporting Tool (CART). Vendor tools also available.</p>

<p>2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1135267770141</p>	
<p>2a.29-31 Data dictionary/code table web page URL or attachment: URL Refer to http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036: Section 1 - Data Dictionary Alphabetical Data Dictionary.</p>	
<p>2a.32-35 Level of Measurement/Analysis (Check the level(s) for which the measure is specified and tested) Facility/Agency, Population: national, Program: QIO</p>	
<p>2a.36-37 Care Settings (Check the setting(s) for which the measure is specified and tested) Hospital</p>	
<p>2a.38-41 Clinical Services (Healthcare services being measured, check all that apply)</p>	
TESTING/ANALYSIS	
2b. Reliability testing	
<p>2b.1 Data/sample (description of data/sample and size): CDAC (Clinical Data Abstraction Center) validation sample: 3Q09.</p>	
<p>2b.2 Analytic Method (type of reliability & rationale, method for testing): CDAC validation sampling involves SDPS selection of sample of 5 cases/quarter across all topics (AMI, HF, Pneumonia, etc.) from each hospital with a minimum of 6 discharges (across all topics) in the Clinical Data Warehouse within 4 months + 15 days following 3Q09. Hospital-abstracted data is compared to CDAC- adjudicated data.</p>	
<p>2b.3 Testing Results (reliability statistics, assessment of adequacy in the context of norms for the test conducted): Aspirin Prescribed at Discharge - 97.5% Clinical Trial - 98.9% Comfort Measures Only - 94.3% Reason for No Aspirin at Discharge - 75.5%</p>	2b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
2c. Validity testing	
<p>2c.1 Data/sample (description of data/sample and size): Face validity is regularly assessed with the Technical Expert Panel responsible for reviewing and supporting the measure topic.</p>	
<p>2c.2 Analytic Method (type of validity & rationale, method for testing): Face validity</p>	2c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
<p>2c.3 Testing Results (statistical results, assessment of adequacy in the context of norms for the test conducted): N/A</p>	2c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
2d. Exclusions Justified	
<p>2d.1 Summary of Evidence supporting exclusion(s): The exclusions of age < 18 years, length of stay > 120 days, and enrollment in a clinical trial are common to the other measures in the AMI measure set, and to the inpatient Hospital Inpatient Quality Reporting Program measure set in general. Patients with documented comfort measures only or those discharged to hospice are appropriate exclusions, as the goal in these cases is palliative care - Therefore, the non-use of aspirin is often clinically appropriate. Patients who leave against medical advice or who expire are appropriately excluded, and it is sensible for those who are discharged to another hospital (where the patient goes on to continue acute care treatment) to be omitted as well. Lastly, there are clinically important contraindications to the use of aspirin. Reasons vary, from patient refusal, aspirin allergies, and current Coumadin therapy (Coumadin</p>	2d C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/>

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.

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) 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.

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
 •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 ... [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.

<p>3Q09: 99,874 AMI patients, 3,019 hospitals 4Q09: 105,659 AMI patients, 3,062 hospitals 1Q10: 107,852 AMI patients, 3,096 hospitals</p> <p>2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (type of analysis & rationale): Analysts review quarterly benchmarks established (using the ABC methodology) and trends to identify differences in performance scores and investigate the possible causes. ABC benchmarks identify superior performance and encourage poorer performers to improve. The methodology is a data-driven, peer-group performance feedback used to positively affect outcomes. If measure specifications (algorithms, data elements) are found to cause the difference in performance, they are reviewed for possible updates.</p> <p>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): National performance rates: 2Q09: 98.3% (benchmark 100.0%) 3Q09: 98.4% (benchmark 100.0%) 4Q09: 98.5% (benchmark 100.0%) 1Q10: 98.5% (benchmark 100.0%)</p>	
<p>2g. Comparability of Multiple Data Sources/Methods</p> <p>2g.1 Data/sample (description of data/sample and size): Both paper records and electronic health records can be used to collect data. Some allowances have been made as facilities incorporate EHRs in their facilities because vendors do not utilize identical data fields, but customize products according to facility need and preferences.</p> <p>2g.2 Analytic Method (type of analysis & rationale): No tests have been performed on this measure to determine comparability of sources (paper medical record vs. EHR).</p> <p>2g.3 Testing Results (e.g., correlation statistics, comparison of rankings): N/A</p>	<p>2g C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>2h. Disparities in Care</p> <p>2h.1 If measure is stratified, provide stratified results (scores by stratified categories/cohorts): Not stratified, but results according to race, sex, etc can be determined.</p> <p>2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: Since the preliminary univariate analyses suggest potential disparities (the largest difference is greater than or equal to 2.0 percentage points as described in 1b.4), further analyses are needed to control for the simultaneous effect of other potential factors such as age, gender, comorbidity, and hospital characteristics and to take into account the correlation/cluster effect of patients discharged from the same hospitals.</p>	<p>2h C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Scientific Acceptability of Measure Properties?</p>	<p>2</p>
<p>Steering Committee: Overall, to what extent was the criterion, Scientific Acceptability of Measure Properties, met? Rationale:</p>	<p>2 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>3. USABILITY</p>	
<p>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)</p>	<p>Eval Rating</p>
<p>3a. Meaningful, Understandable, and Useful Information</p>	<p>3a</p>

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 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.

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

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.

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.

<p>3a.1 Current Use: In use</p> <p>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). <u>If not publicly reported</u>, state the plans to achieve public reporting within 3 years): Hospital Inpatient Quality Reporting Program: . http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1138115987129 . http://www.hospitalcompare.hhs.gov/</p> <p>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). <u>If not used for QI</u>, state the plans to achieve use for QI within 3 years): Hospital Inpatient Quality Reporting Program (Measures can be used by individual hospitals for internal quality improvement): . http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1138115987129 . http://www.hospitalcompare.hhs.gov/ Additionally, the Joint Commission also uses this measure for accreditation.</p> <p>Testing of Interpretability (Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement)</p> <p>3a.4 Data/sample (description of data/sample and size): Unknown. [Feedback on the Hospital Compare website (used for public reporting) is collected through another contractor.]</p> <p>3a.5 Methods (e.g., focus group, survey, QI project): Voluntary electronic survey by visitors to website.</p> <p>3a.6 Results (qualitative and/or quantitative results and conclusions): Not available.</p>	C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
<p>3b/3c. Relation to other NQF-endorsed measures</p>	
<p>3b.1 NQF # and Title of similar or related measures: NQF #0631: Secondary Prevention of Cardiovascular Events - Use of Aspirin or Antiplatelet Therapy</p>	
<p>(for NQF staff use) Notes on similar/related endorsed or submitted measures:</p>	
<p>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? No, this measure's specifications are not harmonized with NQF #0631 measure specifications, as the latter's measure population uses the outpatient setting, includes patients ages 21 and older, diagnosed with IVD as defined by coronary artery disease, peripheral vascular disease or cerebrovascular disease, who are asked about aspirin use, and assesses the proportion of patients with ischemic vascular disease that are taking aspirin or an antiplatelet agent. This measure is concentrated on care of the AMI patient who is admitted for inpatient care; a completely different focus in terms of setting and care. NQF #0631 does provide for the exclusion of patients with an allergy to aspirin (or antiplatelet drugs) in the past or those with documentation of aspirin (or antiplatelet drug) contraindications, similar to this measure, but it also automatically excludes patients with evidence of metastatic disease or active treatment of malignancy (chemotherapy or radiation therapy) in the last 6 months and patients who have been in a skilled nursing facility in the last 3 months - Conditions which our team believes are relative contraindications which require that the physician specifically document a linkage to the non-use of aspirin (vs. automatic exclusion).</p>	3b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/>
<p>3c. Distinctive or Additive Value 3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-</p>	3c C <input type="checkbox"/>

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).

<p>endorsed measures: No NQF-endorsed measures with same topic and target population.</p> <p>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: No NQF-endorsed measures with same topic and target population.</p>	<p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Usability</i>?</p>	<p>3</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Usability</i>, met? Rationale:</p>	<p>3</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>4. FEASIBILITY</p>	
<p>Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)</p>	<p>Eval Ratin g</p>
<p>4a. Data Generated as a Byproduct of Care Processes</p> <p>4a.1-2 How are the data elements that are needed to compute measure scores generated? Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), 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)</p>	<p>4a</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>4b. Electronic Sources</p> <p>4b.1 Are all the data elements available electronically? (<i>elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims</i>) No</p> <p>4b.2 If not, specify the near-term path to achieve electronic capture by most providers. Retooling work with HHS is expected to be completed in 2011.</p>	<p>4b</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>4c. Exclusions</p> <p>4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications? No</p> <p>4c.2 If yes, provide justification.</p>	<p>4c</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences</p> <p>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. 1. Since the time of last NQF endorsement (May 2007), the HeartCare measures team met with other topic teams within the Hospital Inpatient Quality Reporting Program (namely, children's asthma and surgical care) to examine the medication constructs being used. The measure designs at that time automatically excluded patients with a documented contraindication or reason to a medication from the measure, regardless of whether the medication ended up being prescribed. That type of design was resulting in a substantial amount of "false exclusions" from the measure. The decision was made to rearrange the measure such that patients who were prescribed the medication would remain in the measure (i.e., be included in the numerator) when a reason for not prescribing the medication was documented, effective with April 1, 2009 discharges. It is believed that the number of false exclusions has significantly decreased as a result. 2. Because the denominator exclusion "Patients with a documented reason for no aspirin at discharge" allows for any physician/advance practice nurse/physician assistant/pharmacist-documented "other reason"</p>	<p>4d</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>

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.

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

<p>for not prescribing aspirin at discharge to count as an exclusion, overuse of this exclusion has the potential for distorting performance rates. However, overall trends in measure numerator and denominator counts do not suggest obvious gaming of the measure. There has been no increasing trend in the use of this reason data element. Nevertheless, exclusion rates for this measure will continue to be monitored for consistency, from quarter to quarter.</p> <p>3. The data elements used in this measure are closely tracked. Questions submitted by abstractors are recorded, and trends related to published abstraction guidelines and disagreements over measure inclusions and exclusions in general are discussed in-depth every 6 months. Revisions in measure specifications, including data element definitions, are made as issues surface (e.g., how to handle documentation of a hold on aspirin at discharge or a planned delay to start aspirin after discharge, what constitutes acceptable physician documentation of a reason for not prescribing aspirin). The frequency of questions pertaining to each data element are tracked by the Hospital Inpatient Quality Reporting Program QIOSC. Clearly the number of questions a data element receives is another indication of how difficult the specifications for the measure might be. Frequency reports are reviewed regularly, to help identify where issues in data element definitions may exist. Of note, in an August 2010 report run by the Hospital Inpatient Quality Reporting Program QIOSC, the number of questions about the abstraction of the two data elements unique to this measure, Aspirin Prescribed at Discharge and Reason for No Aspirin at Discharge, amounted to 15, only 3.3% of the total 458 Quest questions received for AMI for that month. Lastly, CDAC validation reports (which compare hospital data to CDAC data) and internal CDAC abstractor accuracy reports are monitored, to ensure good quality data. In sum, issues which may surface in questions submitted by users and CDAC validation/accuracy reports will continue to be closely monitored to identify any additional problems, and revisions will be made if warranted.</p>	
<p>4e. Data Collection Strategy/Implementation</p> <p>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: The reordering of the “medication prescribed” and “reason for no medication” specifications done for April 1, 2009+ discharges (as described in section 4d.1) reduces abstraction burden. Abstractors no longer have to do an exhaustive search for acceptable reasons for not prescribing aspirin at discharge in cases where the patient was prescribed the aspirin, saving valuable abstraction time.</p> <p>4e.2 Costs to implement the measure (<i>costs of data collection, fees associated with proprietary measures</i>): Varies according to data collection method (use of vendor) and type of abstractor used to collect clinical data. We have not received feedback that this measure has caused undue burden to the facilities collecting data.</p> <p>4e.3 Evidence for costs: N/A</p> <p>4e.4 Business case documentation: N/A</p>	<p>4e C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Feasibility</i>?</p>	<p>4</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Feasibility</i>, met? Rationale:</p>	<p>4 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>RECOMMENDATION</p>	
<p>(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.</p>	<p>Time-limited <input type="checkbox"/></p>
<p>Steering Committee: Do you recommend for endorsement? Comments:</p>	<p>Y <input type="checkbox"/> N <input type="checkbox"/></p>

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).

	A <input type="checkbox"/>
CONTACT INFORMATION	
Co.1 Measure Steward (Intellectual Property Owner)	
Co.1 Organization Centers for Medicare & Medicaid Services, 7500 Security Boulevard , Baltimore, Maryland, 21244-1850	
Co.2 Point of Contact Kristie , Baus, RN, MS, kristie.baus@cms.hhs.gov, 410-786-8161-	
Measure Developer If different from Measure Steward	
Co.3 Organization Centers for Medicare & Medicaid Services, 7500 Security Boulevard, Baltimore, Maryland, 21244-1850	
Co.4 Point of Contact Kristie , Baus, RN, MS, kristie.baus@cms.hhs.gov, 410-786-8161-	
Co.5 Submitter If different from Measure Steward POC Jo, DeBuhr, RN, BSN, broncosrule@att.net, 303-457-3195-, OFMQ	
Co.6 Additional organizations that sponsored/participated in measure development The Joint Commission	
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. This measure is reviewed and maintained by the Heart Care Technical Expert Panel. Quarterly teleconferences are held to discuss issues pertinent to this measure (and its specifications) and potential revisions. Current members: Frederick Masoudi, MD, MSPH Workgroup Chair: Denver Health Medical Center, University of Colorado at Denver and Health Sciences Center Don Casey, MD, MPH, MBA: VP Quality and Chief Medical Officer, Atlantic Health, Rep. of the American College of Physicians Elizabeth Delong, PhD: Professor and Chair, Duke University, Biostatistics and Bioinformatics, Co-Director, Outcomes Research and Assessment Joseph Drozda, MD: Clinical Investigator, Mercy Health Research, Executive Committee Member, PCPI, Rep. of American Medical Association John P. Erwin, III: Professor of Medicine, Co-Director, Cardiovascular Fellowship Program, Hospital Champion, Acute Myocardial Infarction Quality Improvement, Scott and White Hospital and Clinic Kerri Fei: Senior Policy Analyst, Measure Development Operations, American Medical Association Susan Fitzgerald, RN, MS: Associate Director, Science and Quality, American College of Cardiology Gary Francis, MD: Professor of Medicine, University of Minnesota, Rep. of Heart Failure Society of America David C. Goff, MD, PhD: Professor and Chair, Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine Kathleen Grady, CNS: Administrative Director, Center for Heart Failure, Bluhm Cardiovascular Institute Division of Cardiothoracic Surgery, Northwestern Memorial Hospital Darryl Gray, MD: Medical Officer, Agency for Healthcare Research and Quality Lee Green, MD: Professor, University of Michigan Medical School Ed Havranek, MD: Professor of Medicine, Denver Health Medical Center, University of Colorado School of Medicine Paul A. Heidenreich: Assistant Professor of Medicine, Associate Professor by courtesy of Health Research and Policy at the VA Palo Alto Health Care System and CHP/PCOR Fellow Alice C. Jacobs, MD: Professor of Medicine, Director, Cardiac Cath Lab, Boston University Medical Center Marvin Konstam, MD: Director, Cardiovascular Center, Tufts Medical Center, Rep. of Heart Failure Society of America Harlan Krumholz, MD: Harold H. Hines, Jr. Professor of Medicine and Epidemiology and Public Health, Yale University School of Medicine Jerod Loeb, PhD: Executive Vice President, Quality Measurement & Research, The Joint Commission Ann [Hiniker] Loth, RN, MS, CNS: Certified Clinical Nurse Specialist, Mayo Foundation Joseph Messer, MD, MACC: Professor of Medicine, Rush University Medical Center, Rep. of American Medical	

<p>Association</p> <p>Eric Peterson, MD, MPH: Professor of Medicine, Director Cardiovascular Research, Duke Clinical Research Institute, Duke University Medical Center</p> <p>Martha Radford, MD: Chief Quality Officer, Professor of Medicine, New York University School of Medicine</p> <p>Rose Marie Robertson, MD: Chief Science Officer, American Heart Association</p> <p>John Rumsfeld, MD, PhD, FACC, FAHA: Staff Cardiologist, Cardiovascular Outcomes Researcher, Denver Veterans Affairs Medical Center</p> <p>David Shahian, MD: Research Director, Center for Quality and Safety, Massachusetts General Hospital</p> <p>Melanie Shahriary, RN, BSN: Associate Director, Performance Measures and Data Standards, American College of Cardiology</p> <p>John Spertus, MD, MPH, FACC: Director of Cardiovascular Education and Outcomes Research, Mid America Heart Institute, University of Missouri</p> <p>Samantha Tierney: Senior Policy Analyst I, American Medical Association</p> <p>Gayle Whitman, PhD, RN, FAAN, FAHA: Sr Vice President, Office of Science Operations, American Heart Association</p> <p>Janet Wright, MD, FACC: Senior Vice President for Science and Quality, American College of Cardiology</p> <p>Contractor Staff:</p> <p>Dale Bratzler, DO, MPH: CEO, Principal Clinical Coordinator, Oklahoma Foundation for Medical Quality</p> <p>Jo DeBuhr, RN: Project Specialist, AMI/HF Inpatient Measures, Oklahoma Foundation for Medical Quality/Colorado Foundation for Medical Care</p> <p>Chris Leber, RN: Project Specialist, AMI/HF Inpatient Measures, Oklahoma Foundation for Medical Quality/Colorado Foundation for Medical Care</p> <p>CMS Staff:</p> <p>Kristie Baus, MS, RN: Government Task Leader, Centers for Medicare and Medicaid Services</p> <p>David Nilasena, MD: Chief Medical Officer, Region VI, Centers for Medicare and Medicaid</p>
<p>Ad.2 If adapted, provide name of original measure: N/A</p> <p>Ad.3-5 If adapted, provide original specifications URL or attachment</p>
<p>Measure Developer/Steward Updates and Ongoing Maintenance</p> <p>Ad.6 Year the measure was first released: 1999</p> <p>Ad.7 Month and Year of most recent revision: 10, 2010</p> <p>Ad.8 What is your frequency for review/update of this measure? Every 6 months</p> <p>Ad.9 When is the next scheduled review/update for this measure? 07, 2011</p>
<p>Ad.10 Copyright statement/disclaimers:</p>
<p>Ad.11 -13 Additional Information web page URL or attachment:</p>
<p>Date of Submission (MM/DD/YY): 12/27/2010</p>

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:
 - o Intermediate outcome - evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
 - o Process - 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).
 - o Structure - 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.
 - o Patient 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.
 - o Access - evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
 - o Efficiency - 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.

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.

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).

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 #: 0137 NQF Project: Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 Measure Title: ACEI or ARB for left ventricular systolic dysfunction- Acute Myocardial Infarction (AMI) Patients
De.2 Brief description of measure: Percentage of acute myocardial infarction (AMI) patients with left ventricular systolic dysfunction (LVSD) who are prescribed an ACEI or ARB at hospital discharge. For purposes of this measure, LVSD is defined as chart documentation of a left ventricular ejection fraction (LVEF) less than 40% or a narrative description of left ventricular systolic (LVS) function consistent with moderate or severe systolic dysfunction.
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 N/A
De.4 National Priority Partners Priority Area: Population health
De.5 IOM Quality Domain: Effectiveness
De.6 Consumer Care Need: Living with illness

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. <i>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.</i>	A Y <input checked="" type="checkbox"/> N <input type="checkbox"/>
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:	

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 <input checked="" type="checkbox"/> N <input type="checkbox"/>
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ► Purpose: Public reporting, Internal quality improvement Accountability, Payment incentive	C Y <input checked="" type="checkbox"/> N <input type="checkbox"/>
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 <input checked="" type="checkbox"/> N <input type="checkbox"/>
(for NQF staff use) Have all conditions for consideration been met? Staff Notes to Steward (<i>if submission returned</i>):	Met Y <input checked="" type="checkbox"/> N <input type="checkbox"/>
Staff Notes to Reviewers (<i>issues or questions regarding any criteria</i>):	
Staff Reviewer Name(s): RWinkler	

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. <i>Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)</i> 1a. High Impact	Eval Ratin g
(for NQF staff use) Specific NPP goal :	
1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, Leading cause of morbidity/mortality, Severity of illness, Patient/societal consequences of poor quality 1a.2	
1a.3 Summary of Evidence of High Impact: In 2010, an estimated 785,000 Americans will have a new coronary event, and approximately 470,000 will have a recurrent event. An estimated additional 195,000 silent first myocardial infarctions occur each year. Approximately every 25 seconds, an American will have a coronary event, and approximately every minute, one will die. In 2004, AMI resulted in 695,000 hospital stays and \$31 billion in health expenditures. The risk of further cardiovascular complications, including recurrent MI, sudden cardiac death, heart failure, stroke, and angina pectoris, among AMI survivors is substantial.	
1a.4 Citations for Evidence of High Impact: · Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G, Ferguson TB, Ford E, Furie K, Gillespie C, Go A, Greenlund K, Haase N, Hailpern S, Ho PM, Howard V, Kissela B, Kittner S, Lackland D, Lisabeth L, Marelli A, McDermott MM, Meigs J, Mozaffarian D, Mussolino M, Nichol G, Roger VL, Rosamond W, Sacco R, Sorlie P, Stafford R, Thom T, Wasserthiel-Smoller S, Wong ND, Wylie-Rosett J; on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2010 update: a report from the American Heart Association. <i>Circulation</i> . 2010;121:e46-e215.	1a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>

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).

<p>1b. Opportunity for Improvement</p> <p>1b.1 Benefits (improvements in quality) envisioned by use of this measure: Use of angiotensin converting enzyme inhibitors or angiotensin receptor blockers in patients with left ventricular systolic dysfunction significantly reduces mortality and other adverse outcomes. Hospital performance rates have gradually increased over the years this measure has been reported to the public. Providers understand the importance of prescribing ACEIs and ARBs for their AMI patients with LVSD unless contraindications exist. Ongoing use of this measure will help ensure that high performing providers maintain high performance and the relatively lower performing providers have an impetus to improve.</p> <p>1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers: National performance rates: 2Q09: 95.4% 3Q09: 95.4% 4Q09: 95.9% 1Q10: 96.0%</p> <p>1b.3 Citations for data on performance gap: Clinical warehouse data: 2Q09: 19,935 AMI patients, 2,337 hospitals 3Q09: 18,475 AMI patients, 2,293 hospitals 4Q09: 19,758 AMI patients, 2,320 hospitals 1Q10: 19,997 AMI patients, 2,341 hospitals</p> <p>1b.4 Summary of Data on disparities by population group: At the univariate analysis level (unadjusted odds ratios), rates ranged from 94.4% for Native-Americans, to 94.8% for Hispanic/Latinos, 94.9% for Asians/Pacific Islanders, 95.3% for White/Caucasians, and 95.8% for African-Americans. The difference from the lowest to the highest rates was 1.4 percentage points. The rate for Caucasians was higher than the rates for minority groups except African-Americans.</p> <p>1b.5 Citations for data on Disparities: 2009 Clinical warehouse data (Total 74,167 patients with race not missing): 57,482 Caucasian patients, 9,024 African-American patients, 5,896 Hispanic patients, 1,372 Asian/Pacific Islander patients, and 393 Native American patients.</p>	<p>1b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>1c. Outcome or Evidence to Support Measure Focus</p> <p>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): ACE inhibitors reduce mortality and morbidity in patients with left ventricular systolic dysfunction after AMI. Additional benefits of ACEIs include alleviation of symptoms. Clinical trials have established ARB therapy as an acceptable alternative to ACEI, especially in patients who are ACEI intolerant. National guidelines strongly recommend ACEIs for patients hospitalized with AMI who have either clinical heart failure or LVSD. Guideline committees have also supported the inclusion of ARBs in performance measures for AMI.</p> <p>1c.2-3. Type of Evidence: Evidence-based guideline, Randomized controlled trial, Systematic synthesis of research, Meta-analysis</p> <p>1c.4 Summary of Evidence (as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome): Several trials have demonstrated the beneficial effects of angiotensin-converting enzyme inhibitors in patients with an MI, especially among those with LV systolic dysfunction. In the GISSI-3 study, therapy with the ACE inhibitor lisinopril resulted in significantly lower rates of death 42 days after myocardial infarction. Follow-up of patients with LV dysfunction after MI in the TRACE (TRAndolapril Cardiac Evaluation) trial showed that the beneficial effect of the ACE inhibitor trandolapril on mortality and hospitalization rate persists in the long term. In patients with MI complicated by LV systolic dysfunction, HF, or both, the angiotensin receptor blocker (ARB) valsartan was as effective as captopril in patients at high risk for cardiovascular events after MI (VALIANT). Chronic treatment of patients with chronic HF with the ARB</p>	<p>1c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

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])

candesartan (at least half of whom had an MI) in the CHARM (Candesartan in Heart failure Assessment in Reduction of Mortality)-Overall program showed a reduction in cardiovascular deaths and hospital admissions for HF, independent of ejection fraction or baseline treatment in patients who did not tolerate ACE-inhibitors. While many patients can tolerate ACE inhibitors, some cannot due to cough or other side effects; in general, ARBs are generally well tolerated in randomized trials of patients judged to be intolerant of ACE inhibitors.

1c.5 Rating of strength/quality of evidence (also provide narrative description of the rating and by whom): ACCF/AHA Task Force on Practice Guidelines, Level of Evidence A: [UA/NSTEMI and STEMI] Data derived from multiple randomized trials or meta-analyses, Multiple populations evaluated. References used to determine level of evidence must be provided and cited with the recommendation.

1c.6 Method for rating evidence: The methodology used by the ACCF/AHA Task Force on Practice Guidelines is fully documented in their publication "Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines" (http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf). The guidelines are based upon a comprehensive assessment, both electronic and manual, of the English-language medical literature. This search focuses on high-quality randomized controlled trials, meta-analyses and systematic reviews, and when applicable observational studies. In some cases where higher quality data is not available, observational studies and case series are also considered. The quality of the design and execution of these studies is determined. When appropriate, data tables are generated from the available literature. After a review of the available literature, the writing committee rates the evidence according to the schemes outlined in their publication.

1c.7 Summary of Controversy/Contradictory Evidence: Aside from avoiding use in patients with clear contraindications to ACEI or ARB therapy, there is broad support in existing guidelines for the use of ACEI/ARBs in reducing mortality and morbidity.

1c.8 Citations for Evidence (other than guidelines):

- Rutherford JD, Pfeffer MA, Moye LA, et al. Effects of captopril on ischemic events after myocardial infarction: results of the Survival and Ventricular Enlargement trial. SAVE Investigators. *Circulation* 1994;90:1731- 8.
- ACE Inhibitor Myocardial Infarction Collaborative Group. Indications for ACE inhibitors in the early treatment of acute myocardial infarction: systematic overview of individual data from 100,000 patients in randomized trials. *Circulation* 1998;97:2202-12.
- Buch P, Rasmussen S, Abildstrom SZ, Kober L, Carlsen J, Torp-Pedersen C. The long-term impact of the angiotensin-converting enzyme inhibitor trandolapril on mortality and hospital admissions in patients with left ventricular dysfunction after a myocardial infarction: follow-up to 12 years. *Eur Heart J* 2005;26:145-52.
- Yusuf S, Pepine CJ, Garces C et al. Effect of enalapril on myocardial infarction and unstable angina in patients with low ejection fractions. *Lancet*. 1992;340(8829):1173-1178.
- Pfeffer MA, Swedberg K, Granger CB, et al. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-Overall programme. *Lancet* 2003;362:759-66.
- Granger CB, McMurray JJ, Yusuf S, Held P, Michelson EL, Olofsson B, et al. Effects of candesartan in patients with chronic heart failure and reduced left-ventricular systolic function intolerant to angiotensin-converting-enzyme inhibitors: the CHARM Alternative trial. *Lancet* 2003;362:772-6.
- Stecker EC, Fendrick AM, Knight BP, Aaronson KD. Prophylactic pacemaker use to allow beta-blocker therapy in patients with chronic heart failure with bradycardia. *Am Heart J* 2006;151:820-8.
- Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. *Lancet*. 1994;343:1115-22.

1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number): [STEMI]

Renin-Angiotensin-Aldosterone System Blockers: ACE Inhibitors Recommendations (p. 236)

1. ACE inhibitors should be started and continued indefinitely in all patients recovering from STEMI with LVEF less than or equal to 40% and for those with hypertension, diabetes, or chronic kidney disease, unless contraindicated.

Renin-Angiotensin-Aldosterone System Blockers: Angiotensin Receptor Blockers (p. 236)

1. Use of angiotensin receptor blockers is recommended in patients who are intolerant of ACE inhibitors

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.

and have HF or have had an MI with LVEF less than or equal to 40%.

[UA/NSTEMI]

5.2.3. Inhibition of the Renin-Angiotensin-Aldosterone System Recommendations (p. e91)

1. Angiotensin-converting enzyme inhibitors should be given and continued indefinitely for patients recovering from UA/NSTEMI with HF, LV dysfunction (LVEF less than 0.40), hypertension, or diabetes mellitus, unless contraindicated.

2. An angiotensin receptor blocker should be prescribed at discharge to those UA/NSTEMI patients who are intolerant of an ACE inhibitor and who have either clinical or radiological signs of HF and LVEF less than 0.40.

1c.10 Clinical Practice Guideline Citation: Anderson JL, Adams CD, Antman EM, Bridges CR, Califf RM, Casey DE Jr, et al. ACC/AHA 2007 guidelines for the management of patients with unstable angina/non-ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines for the Management of Patients With Unstable Angina/Non-ST-Elevation Myocardial Infarction): developed in collaboration with the American College of Emergency Physicians, American College of Physicians, Society for Academic Emergency Medicine, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol.* 2007;50:e1-157.

Antman EM, Hand M, Armstrong PW, Bates ER, Green LA, Halasyamani LK, et al. 2007 focused update of the ACC/AHA 2004 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 (Writing Group to Review New Evidence and Update the ACC/AHA 2004 Guidelines for the Management of Patients With ST-Elevation Myocardial Infarction). *J Am Coll Cardiol.* 2008;51:210-47.

1c.11 National Guideline Clearinghouse or other URL:

<http://content.onlinejacc.org/cgi/reprint/51/2/210.pdf>,

<http://content.onlinejacc.org/cgi/reprint/50/7/e1.pdf>

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

Rating made by ACCF/AHA Task Force on Practice Guidelines: [UA/NSTEMI and STEMI] Class I recommendation - Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective. Benefit >>> Risk. Procedure/treatment should be performed/administered.

1c.13 Method for rating strength of recommendation (If different from USPSTF system, also describe rating and how it relates to USPSTF):

[UA/NSTEMI and STEMI] The methodology used by the ACCF/AHA Task Force on Practice Guidelines is fully documented in their publication "Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines"

(http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf).

Recommendations are assigned strength by the Task Force based upon evidence, benefit vs. risk vs. harm, and patient preference.

Both the ACCF/AHA Guidelines and the USPSTF assess evidence with respect to two parameters: 1) the magnitude of the benefit, and 2) the certainty of this benefit. However, they use different coding systems. In ascertaining magnitude of the benefit, the ACCF/AHA uses a Class I-III scale and the USPSTF uses a high-moderate-low scale. In determining the certainty of this benefit, the ACCF/AHA uses levels of evidence A-C and USPSTF uses a high-moderate-low scale.

1c.14 Rationale for using this guideline over others:

The ACCF/AHA guidelines are widely accepted national guidelines that address the therapy of patients with AMI; they use an explicit and transparent methodology; and have thus served as the foundation of national quality measures.

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

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.

2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, <u>as specified</u> , produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Ratin g
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:	
<u>2a. Precisely Specified</u>	
<p>2a.1 Numerator Statement (<i>Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome</i>): AMI patients who are prescribed an ACEI or ARB at hospital discharge</p> <p>2a.2 Numerator Time Window (<i>The time period in which cases are eligible for inclusion in the numerator</i>): From hospital arrival to time of hospital discharge</p> <p>2a.3 Numerator Details (<i>All information required to collect/calculate the numerator, including all codes, logic, and definitions</i>): Refer to http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036: <ul style="list-style-type: none"> · Section 1 - Data Dictionary Alphabetical Data Dictionary - pages 1-18 through 1-19 plus pages 1-67 through 1-68. · Appendices Appendix C - Medication Tables - pages Appendix C-6 through Appendix C-7 plus pages Appendix C-11 through Appendix C-12. · Section 2 - Measurement Information Section 2.1 - Acute Myocardial Infarction (AMI) - pages AMI-3-1 through AMI-3-6. </p> <p>2a.4 Denominator Statement (<i>Brief, text description of the denominator - target population being measured</i>): AMI patients (International Classification of Diseases, 9th revision, Clinical Modification [ICD-9-CM] principal diagnosis code of AMI: 410.00, 410.01, 410.10, 410.11, 410.20, 410.21, 410.30, 410.31, 410.40, 410.41, 410.50, 410.51, 410.60, 410.61, 410.70, 410.71, 410.80, 410.81, 410.90, 410.91); with chart documentation of a left ventricular ejection fraction (LVEF) < 40% or a narrative description of left ventricular systolic (LVS) function consistent with moderate or severe systolic dysfunction</p> <p>2a.5 Target population gender: Female, Male</p> <p>2a.6 Target population age range: Greater than or equal to 18 years old</p> <p>2a.7 Denominator Time Window (<i>The time period in which cases are eligible for inclusion in the denominator</i>): From hospital arrival to time of hospital discharge</p> <p>2a.8 Denominator Details (<i>All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions</i>): ICD-9-CM Principal Diagnosis codes: 410.00: Anterolateral wall, acute myocardial infarction-episode of care unspecified 410.01: Anterolateral wall, acute myocardial infarction-initial episode 410.10: Other anterior wall, acute myocardial infarction-episode of care unspecified 410.11: Other anterior wall, acute myocardial infarction-initial episode 410.20: Inferolateral wall, acute myocardial infarction-episode of care unspecified 410.21: Inferolateral wall, acute myocardial infarction-initial episode 410.30: Inferoposterior wall, acute myocardial infarction-episode of care unspecified 410.31: Inferoposterior wall, acute myocardial infarction-initial episode 410.40: Other inferior wall, acute myocardial infarction-episode of care unspecified</p>	
	2a- spec s C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>

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) .

410.41: Other inferior wall, acute myocardial infarction-initial episode
 410.50: Other lateral wall, acute myocardial infarction-episode of care unspecified
 410.51: Other lateral wall, acute myocardial infarction-initial episode
 410.60: True posterior wall, acute myocardial infarction-episode of care unspecified
 410.61: True posterior wall, acute myocardial infarction-initial episode
 410.70: Subendocardial, acute myocardial infarction-episode of care unspecified
 410.71: Subendocardial, acute myocardial infarction-initial episode
 410.80: Other specified sites, acute myocardial infarction-episode of care unspecified
 410.81: Other specified sites, acute myocardial infarction-initial episode
 410.90: Unspecified site, acute myocardial infarction-episode of care unspecified
 410.91: Unspecified site, acute myocardial infarction-initial episode
 LVSD - Refer to
<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036>:
 · Section 1 - Data Dictionary | Alphabetical Data Dictionary - pages 1-257 through 1-260.

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population*): Exclusions:
 • <18 years of age
 • Patients who have a length of stay greater than 120 days
 • Discharged to another hospital
 • Expired
 • Left against medical advice
 • Discharged to home for hospice care
 • Discharged to a health care facility for hospice care
 • Patients with comfort measures only documented
 • Patients enrolled in clinical trials
 • Patients with a documented reason for no ACEI and no ARB at discharge

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions*):
 Refer to
<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036>:
 · Section 1 - Data Dictionary | Alphabetical Data Dictionary - pages 1-20 through 1-21, 1-90, 1-98 through 1-104, 1-117 through 1-120, 1-204, 1-257 through 1-260, and 1-315 through 1-320.
 · Appendices | Appendix C - Medication Tables PDF - pages Appendix C-6 through Appendix C-7 plus pages Appendix C-11 through Appendix C-12, and Appendix H - Miscellaneous Tables - page Appendix H-5.
 · Section 2 - Measurement Information | Section 2.1 - Acute Myocardial Infarction (AMI) - pages AMI-5 plus AMI-3-1 through AMI-3-6

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions*):
 N/A

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: Rate/proportion
2a.20 Interpretation of Score: Better quality = Higher score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps*):
 Refer to
<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036>: Section 2 - Measurement Information | Section 2.1 - Acute Myocardial Infarction (AMI) - pages AMI-5 plus AMI-3-1 through AMI-3-6.

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

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.

Benchmarks are established using the ABC methodology, based on the actual performance of the top facilities. ABC benchmarks identify superior performance and encourage poorer performers to improve. The methodology is a data-driven, peer-group performance feedback used to positively affect outcomes.

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):*
 Patients admitted to the hospital for inpatient acute care with an ICD-9-CM Principal Diagnosis Code for AMI as defined in section 2a.8, a patient age greater than or equal to 18 years, and a length of stay less than or equal to 120 days would be included in the initial patient population and eligible to be sampled.
 Monthly Sample Size Based on Population Size (Average monthly initial patient population size: Minimum required sample size):
 >= 516: 104
 131-515: 20% of Initial Patient Population size
 26-130: 26
 < 26: 100%

2a.24 Data Source *(Check the source(s) for which the measure is specified and tested)*
 Paper medical record/flow-sheet, 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.):*
 Centers for Medicare & Medicaid Services (CMS) Abstraction & Reporting Tool (CART). Vendor tools also available.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL
<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1135267770141>

2a.29-31 Data dictionary/code table web page URL or attachment: URL Refer to
<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036>: Section 1 - Data Dictionary | Alphabetical Data Dictionary.

2a.32-35 Level of Measurement/Analysis *(Check the level(s) for which the measure is specified and tested)*
 Facility/Agency, Population: national, Program: QIO

2a.36-37 Care Settings *(Check the setting(s) for which the measure is specified and tested)*
 Hospital

2a.38-41 Clinical Services *(Healthcare services being measured, check all that apply)*

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size):* CDAC (Clinical Data Abstraction Center) validation sample: 3Q09.

2b.2 Analytic Method *(type of reliability & rationale, method for testing):*
 CDAC validation sampling involves SDPS selection of sample of 5 cases/quarter across all topics (AMI, HF, Pneumonia, etc.) from each hospital with a minimum of 6 discharges (across all topics) in the Clinical Data Warehouse within 4 months + 15 days following 3Q09. Hospital-abstracted data is compared to CDAC-adjudicated data.

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted):*
 ACEI Prescribed at Discharge - 91.0%
 ARB Prescribed at Discharge - 86.4%
 Clinical Trial - 98.9%
 Comfort Measures Only - 94.3%

2b
 C
 P
 M
 N

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.

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.

<p>LVSD - 94.7% Reason for No ACEI and No ARB at Discharge - 77.5%</p>	
<p>2c. Validity testing</p> <p>2c.1 Data/sample (description of data/sample and size): Face validity is regularly assessed with the Technical Expert Panel responsible for reviewing and supporting the measure topic.</p> <p>2c.2 Analytic Method (type of validity & rationale, method for testing): Face validity</p> <p>2c.3 Testing Results (statistical results, assessment of adequacy in the context of norms for the test conducted): N/A</p>	<p>2c</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>2d. Exclusions Justified</p> <p>2d.1 Summary of Evidence supporting exclusion(s): The exclusions of age < 18 years, length of stay > 120 days, and enrollment in a clinical trial are common to the other measures in the AMI measure set, and to the inpatient Hospital Inpatient Quality Reporting Program measure set in general. Patients with documented comfort measures only or those discharged to hospice are appropriate exclusions, as the goal in these cases is palliative care - Therefore, the non-use of ACEI/ARB is often clinically appropriate. Patients who leave against medical advice or who expire are appropriately excluded, and it is sensible for those who are discharged to another hospital (where the patient goes on to continue acute care treatment) to be omitted as well. Lastly, there are clinically important contraindications to the use of ACEIs or ARBs. Reasons vary, from patient refusal and ACEI/ARB allergies, to clinical conditions such as moderate or severe aortic stenosis or severe hypotension. In these types of cases, the non-use of ACEI/ARB should not count against the provider if the clinical reason for not prescribing the ACEI/ARB is documented. Exclusions in this measure are concordant with the 2008 ACC/AHA Clinical Performance Measures for Adults With ST-elevation and non-ST-elevation Myocardial Infarction.</p> <p>2d.2 Citations for Evidence: 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 Infarction). J Am Coll Cardiol. 2008;52:2046-99.</p> <p>2d.3 Data/sample (description of data/sample and size): Clinical warehouse data: 144,247 AMI patients, 3,502 hospitals, 1Q10.</p> <p>2d.4 Analytic Method (type analysis & rationale): A frequency count was conducted to calculate the percentages outlined in section 2d.5. Frequency counts are a simple, efficient way to determine the occurrence of specific values of a data element in a given data set.</p> <p>2d.5 Testing Results (e.g., frequency, variability, sensitivity analyses): Rates of Exclusion: <ul style="list-style-type: none"> Patients with comfort measures only documented: 5.8% Patients enrolled in clinical trials: 0.5% Discharged/transferred to another hospital for inpatient care, discharged/transferred to a federal health care facility, discharged/transferred to hospice, expired, or left against medical advice or discontinued care: 14.7% LVSD not documented as either EF < 40% or a narrative description consistent with moderate or severe systolic dysfunction: 61.4% Patients with a documented reason for no ACEI and no ARB at discharge: 3.7% </p>	<p>2d</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>2e. Risk Adjustment for Outcomes/ Resource Use Measures</p> <p>2e.1 Data/sample (description of data/sample and size): N/A</p>	<p>2e</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p>

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) 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.

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
 •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).

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 outcome (but not disparities in care) and are present at start of care; OR
 rationale/data support no risk adjustment.

<p>2e.2 Analytic Method (<i>type of risk adjustment, analysis, & rationale</i>): N/A</p> <p>2e.3 Testing Results (<i>risk model performance metrics</i>): N/A</p> <p>2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: N/A</p>	<p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p> <p><input type="checkbox"/></p>
<p>2f. Identification of Meaningful Differences in Performance</p> <p>2f.1 Data/sample from Testing or Current Use (<i>description of data/sample and size</i>): Clinical warehouse data: 2Q09: 19,935 AMI patients, 2,337 hospitals 3Q09: 18,475 AMI patients, 2,293 hospitals 4Q09: 19,758 AMI patients, 2,320 hospitals 1Q10: 19,997 AMI patients, 2,341 hospitals</p> <p>2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (<i>type of analysis & rationale</i>): Analysts review quarterly benchmarks established (using the ABC methodology) and trends to identify differences in performance scores and investigate the possible causes. ABC benchmarks identify superior performance and encourage poorer performers to improve. The methodology is a data-driven, peer-group performance feedback used to positively affect outcomes. If measure specifications (algorithms, data elements) are found to cause the difference in performance, they are reviewed for possible updates.</p> <p>2f.3 Provide Measure Scores from Testing or Current Use (<i>description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance</i>): National performance rates: 2Q09: 95.4% (benchmark 100.0%) 3Q09: 95.4% (benchmark 99.8%) 4Q09: 95.9% (benchmark 99.8%) 1Q10: 96.0% (benchmark 99.9%)</p>	<p>2f</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>2g. Comparability of Multiple Data Sources/Methods</p> <p>2g.1 Data/sample (<i>description of data/sample and size</i>): Both paper records and electronic health records can be used to collect data. Some allowances have been made as facilities incorporate EHRs in their facilities because vendors do not utilize identical data fields, but customize products according to facility need and preferences.</p> <p>2g.2 Analytic Method (<i>type of analysis & rationale</i>): No tests have been performed on this measure to determine comparability of sources (paper medical record vs. EHR).</p> <p>2g.3 Testing Results (<i>e.g., correlation statistics, comparison of rankings</i>): N/A</p>	<p>2g</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>2h. Disparities in Care</p> <p>2h.1 If measure is stratified, provide stratified results (<i>scores by stratified categories/cohorts</i>): Not stratified, but results according to race, sex, etc can be determined.</p> <p>2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: Since the preliminary univariate analyses do not show a clear indication of disparities (the largest difference is less than 2.0 percentage points as described in 1b.4), further analyses are needed to control for the simultaneous effect of other potential factors such as age, gender, comorbidity, and hospital characteristics and to take into account the correlation/cluster effect of patients discharged from the same hospitals.</p>	<p>2h</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p> <p><input type="checkbox"/></p>

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 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.

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 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.

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

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 <i>Scientific Acceptability of Measure Properties</i> ?	2
Steering Committee: Overall, to what extent was the criterion, <i>Scientific Acceptability of Measure Properties</i> , met? Rationale:	2 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
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). <u>If not publicly reported</u> , state the plans to achieve public reporting within 3 years): Hospital Inpatient Quality Reporting Program: . http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1138115987129 . http://www.hospitalcompare.hhs.gov/	
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). <u>If not used for QI</u> , state the plans to achieve use for QI within 3 years): Hospital Inpatient Quality Reporting Program (Measures can be used by individual hospitals for internal quality improvement): . http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1138115987129 . http://www.hospitalcompare.hhs.gov/ Additionally, the Joint Commission also uses this measure for accreditation.	
Testing of Interpretability (Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement) 3a.4 Data/sample (description of data/sample and size): Unknown. [Feedback on the Hospital Compare website (used for public reporting) is collected through another contractor.]	
3a.5 Methods (e.g., focus group, survey, QI project): Voluntary electronic survey by visitors to website.	3a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
3a.6 Results (qualitative and/or quantitative results and conclusions): Not available.	
3b/3c. Relation to other NQF-endorsed measures	
3b.1 NQF # and Title of similar or related measures: NQF #0551: Ace Inhibitor / Angiotensin Receptor Blocker Use and Persistence Among Members with Coronary Artery Disease at High Risk for Coronary Events, NQF #0594: Post MI: ACE inhibitor or ARB therapy	
(for NQF staff use) Notes on similar/related endorsed or submitted measures:	
3b. Harmonization	3b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
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? This measure's specifications are not harmonized with NQF #0551 measure specifications. NQF #0551 is an	

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.

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.

<p>outpatient measure which assesses the use of and persistence to ACEIs and ARBs during a one year period in patients ages 18 - 75 with coronary artery disease or other atherosclerotic vascular disease (i.e., peripheral artery disease, atherosclerotic aortic disease, and carotid artery disease) who are at high risk for coronary events. High-risk comorbidities include heart failure, hypertension, diabetes, or chronic kidney disease (excluding stage V and patients on dialysis). In contrast, this measure focuses on inpatient care of the AMI patient in particular; a completely different focus in terms of setting and treatment. NQF #0551 excludes hospice patients, like this measure, but it automatically excludes many other types of patients, including those with a diagnosis of angioedema, hyperkalemia, hypotension, arterial stenosis, or renal failure (stage V or dialysis) at any time during the measurement year and patients who were pregnant during the measurement year - Conditions which our team believes are relative contraindications which require that the physician specifically document a linkage to the non-use of ACEI/ARB (vs. automatic exclusion).</p> <p>This measure's specifications are also not harmonized with NQF #0594 measure specifications. Like NQF #0551, NQF #0594 is an outpatient measure. NQF #0594 assesses the use of ACEIs and ARBs during a one year period in patients with STEMI or NSTEMI plus a history of hypertension, heart failure and/or diabetes prior to the measurement year. Again, in contrast, this measure is concentrated on care of the hospitalized AMI patient in particular; a completely different focus. NQF #0594 automatically excludes many types of patients, including those with a diagnosis of hyperkalemia, renal artery stenosis, ESRD - severe chronic kidney disease, pregnancy, or angioneurotic edema - Conditions which our team again believes are relative contraindications which require linkage in physician documentation.</p>	<p>NA <input type="checkbox"/></p>
<p>3c. Distinctive or Additive Value</p> <p>3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures: No NQF-endorsed measures with same topic and target population.</p> <p>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: No NQF-endorsed measures with same topic and target population.</p>	<p>3c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Usability</i>?</p>	<p>3</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Usability</i>, met? Rationale:</p>	<p>3 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>4. FEASIBILITY</p>	
<p>Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)</p>	<p>Eval Ratin g</p>
<p>4a. Data Generated as a Byproduct of Care Processes</p> <p>4a.1-2 How are the data elements that are needed to compute measure scores generated? Data generated as byproduct of care processes during care delivery (Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition), 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)</p>	<p>4a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>4b. Electronic Sources</p> <p>4b.1 Are all the data elements available electronically? (<i>elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims</i>) No</p> <p>4b.2 If not, specify the near-term path to achieve electronic capture by most providers. Retooling work with HHS is expected to be completed in 2011.</p>	<p>4b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>4c. Exclusions</p>	<p>4c</p>

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.

<p>4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications? No</p>	<p>C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>4c.2 If yes, provide justification.</p>	
<p>4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences</p>	
<p>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.</p> <p>1. Documentation of both a reason for not prescribing an ACEI and reason for not prescribing an ARB are required for measure exclusion (barring other exclusions). Providers challenged the need to explicitly document both a reason for not prescribing an ACEI and reason for not prescribing an ARB when the reasons for not prescribing one class often apply to the other class in many cases. This concern was rectified in the measure and abstraction specifications effective with April 1, 2007 discharges. Specifications were changed to allow documentation of a reason for not prescribing one class (either ACEI or ARB) to be considered implicit documentation of a reason for not prescribing the other class when one of the following conditions was noted to be the reason for no ACEI or the reason for no ARB: angioedema, hyperkalemia, hypotension, renal artery stenosis, and worsening renal function/renal disease/dysfunction.</p> <p>2. Since the time of last NQF endorsement (May 2007), the Heart Care measures team met with other topic teams within the Hospital Inpatient Quality Reporting Program (namely, children’s asthma and surgical care) to examine the medication constructs being used. The measure designs at that time automatically excluded patients with a documented contraindication to a medication or reason for not prescribing a medication from the measure, regardless of whether the medication ended up being prescribed. That type of design was resulting in a substantial amount of “false exclusions” from the measure. The decision was made to rearrange the measure such that patients who were prescribed the medication would remain in the measure (i.e., be included in the numerator) when a reason for not prescribing the medication was documented, effective with April 1, 2009 discharges. It is believed that the number of false exclusions has significantly decreased as a result.</p> <p>3. Because the denominator exclusion “Patients with a documented reason for no ACEI and no ARB at discharge” allows for any physician/advance practice nurse/physician assistant/pharmacist-documented “other reason” for not prescribing ACEI or ARB at discharge to count as an exclusion, overuse of this exclusion has the potential for distorting performance rates. However, overall trends in measure numerator and denominator counts do not suggest obvious gaming of the measure. There has been no increasing trend in the use of this reason data element since the logical increase which resulted when abstraction guidelines were changed to allow for the documentation of a reason for not prescribing one class (either ACEI or ARB) to be considered implicit documentation of a reason for not prescribing the other class in the cases of angioedema, hyperkalemia, hypotension, renal artery stenosis, and worsening renal function/renal disease/dysfunction. Nevertheless, exclusion rates for this measure will continue to be monitored for consistency, from quarter to quarter.</p> <p>4. The data elements used in this measure are closely tracked. Questions submitted by abstractors are recorded, and trends related to published abstraction guidelines and disagreements over measure inclusions and exclusions in general are discussed in-depth every 6 months. Revisions in measure specifications, including data element definitions, are made as issues surface (e.g., how to handle documentation of a hold on ACEI/ARB at discharge or a planned delay to start ACEI/ARB after discharge, what constitutes acceptable physician documentation of a reason for not prescribing ACEI/ARB). The frequency of questions pertaining to each data element are tracked by the Hospital Inpatient Quality Reporting Program QIOSC. Clearly the number of questions a data element receives is another indication of how difficult the specifications for the measure might be. Frequency reports are reviewed regularly, to help identify where issues in data element definitions may exist. Of note, in an August 2010 report run by the Hospital Inpatient Quality Reporting Program QIOSC, the number of questions about the abstraction of the four most unique data elements to this measure (shared with the HF ACEI/ARB for LVSD measure), ACEI Prescribed at Discharge, ARB Prescribed at Discharge, LVSD, and Reason for No ACEI and No ARB at Discharge, amounted to 142, 16.7% of the total 848 Quest questions received for AMI and HF for that month. Lastly, CDAC validation reports (which compare hospital data to CDAC data) and internal CDAC abstractor accuracy reports are monitored, to ensure good quality data. In sum, issues which may surface in questions submitted by users and CDAC validation/accuracy reports will continue to be closely monitored to identify any additional problems, and revisions will be made if warranted.</p>	

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

4d
C
P
M
N

4e. Data Collection Strategy/Implementation	
<p>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: Both the change to allow for the documentation of a reason for not prescribing one class (either ACEI or ARB) to be considered implicit documentation of a reason for not prescribing the other class in the cases of angioedema, hyperkalemia, hypotension, renal artery stenosis, and worsening renal function for April 2007+ discharges and the reordering of the "medication prescribed" and "reason for no medication" specifications done for April 2009+ discharges (as described in section 4d.1) reduce abstraction burden. Abstractors no longer have to do an exhaustive search for acceptable reasons for not prescribing ACEI and/or ARB at discharge, saving valuable abstraction time.</p> <p>4e.2 Costs to implement the measure (<i>costs of data collection, fees associated with proprietary measures</i>): Varies according to data collection method (use of vendor) and type of abstractor used to collect clinical data. We have not received feedback that this measure has caused undue burden to the facilities collecting data.</p> <p>4e.3 Evidence for costs: N/A</p> <p>4e.4 Business case documentation: N/A</p>	<p>4e</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Feasibility</i> ?	4
Steering Committee: Overall, to what extent was the criterion, <i>Feasibility</i> , met? Rationale:	<p>4</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
RECOMMENDATION	
(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.	<p>Time-limited</p> <p><input type="checkbox"/></p>
Steering Committee: Do you recommend for endorsement? Comments:	<p>Y <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>A <input type="checkbox"/></p>
CONTACT INFORMATION	
<p>Co.1 Measure Steward (Intellectual Property Owner)</p> <p>Co.1 <u>Organization</u> Centers for Medicare & Medicaid Services, 7500 Security Boulevard , Baltimore, Maryland, 21244-1850</p> <p>Co.2 <u>Point of Contact</u> Kristie, Baus, RN, MS, kristie.baus@cms.hhs.gov, 410-786-8161-</p>	
<p>Measure Developer If different from Measure Steward</p> <p>Co.3 <u>Organization</u> Centers for Medicare & Medicaid Services, 7500 Security Boulevard, Baltimore, Maryland, 21244-1850</p> <p>Co.4 <u>Point of Contact</u> Kristie, Baus, RN, MS, kristie.baus@cms.hhs.gov, 410-786-8161-</p>	
<p>Co.5 Submitter If different from Measure Steward POC Jo, DeBuhr, RN, BSN, broncosrule@att.net, 303-457-3195-, OFMQ</p>	

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).

Co.6 Additional organizations that sponsored/participated in measure development
 The Joint Commission

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.

This measure is reviewed and maintained by the Heart Care Technical Expert Panel. Quarterly teleconferences are held to discuss issues pertinent to this measure (and its specifications) and potential revisions. Current members: Frederick Masoudi, MD, MSPH Workgroup Chair: Denver Health Medical Center, University of Colorado at Denver and Health Sciences Center

Don Casey, MD, MPH, MBA: VP Quality and Chief Medical Officer, Atlantic Health, Rep. of the American College of Physicians

Elizabeth Delong, PhD: Professor and Chair, Duke University, Biostatistics and Bioinformatics, Co-Director, Outcomes Research and Assessment

Joseph Drozda, MD: Clinical Investigator, Mercy Health Research, Executive Committee Member, PCPI, Rep. of American Medical Association

John P. Erwin, III: Professor of Medicine, Co-Director, Cardiovascular Fellowship Program, Hospital Champion, Acute Myocardial Infarction Quality Improvement, Scott and White Hospital and Clinic

Kerri Fei: Senior Policy Analyst, Measure Development Operations, American Medical Association

Susan Fitzgerald, RN, MS: Associate Director, Science and Quality, American College of Cardiology

Gary Francis, MD: Professor of Medicine, University of Minnesota, Rep. of Heart Failure Society of America

David C. Goff, MD, PhD: Professor and Chair, Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine

Kathleen Grady, CNS: Administrative Director, Center for Heart Failure, Bluhm Cardiovascular Institute Division of Cardiothoracic Surgery, Northwestern Memorial Hospital

Darryl Gray, MD: Medical Officer, Agency for Healthcare Research and Quality

Lee Green, MD: Professor, University of Michigan Medical School

Ed Havranek, MD: Professor of Medicine, Denver Health Medical Center, University of Colorado School of Medicine

Paul A. Heidenreich: Assistant Professor of Medicine, Associate Professor by courtesy of Health Research and Policy at the VA Palo Alto Health Care System and CHP/PCOR Fellow

Alice C. Jacobs, MD: Professor of Medicine, Director, Cardiac Cath Lab, Boston University Medical Center

Marvin Konstam, MD: Director, Cardiovascular Center, Tufts Medical Center, Rep. of Heart Failure Society of America

Harlan Krumholz, MD: Harold H. Hines, Jr. Professor of Medicine and Epidemiology and Public Health, Yale University School of Medicine

Jerod Loeb, PhD: Executive Vice President, Quality Measurement & Research, The Joint Commission

Ann [Hiniker] Loth, RN, MS, CNS: Certified Clinical Nurse Specialist, Mayo Foundation

Joseph Messer, MD, MACC: Professor of Medicine, Rush University Medical Center, Rep. of American Medical Association

Eric Peterson, MD, MPH: Professor of Medicine, Director Cardiovascular Research, Duke Clinical Research Institute, Duke University Medical Center

Martha Radford, MD: Chief Quality Officer, Professor of Medicine, New York University School of Medicine

Rose Marie Robertson, MD: Chief Science Officer, American Heart Association

John Rumsfeld, MD, PhD, FACC, FAHA: Staff Cardiologist, Cardiovascular Outcomes Researcher, Denver Veterans Affairs Medical Center

David Shahian, MD: Research Director, Center for Quality and Safety, Massachusetts General Hospital

Melanie Shahriary, RN, BSN: Associate Director, Performance Measures and Data Standards, American College of Cardiology

John Spertus, MD, MPH, FACC: Director of Cardiovascular Education and Outcomes Research, Mid America Heart Institute, University of Missouri

Samantha Tierney: Senior Policy Analyst I, American Medical Association

Gayle Whitman, PhD, RN, FAAN, FAHA: Sr Vice President, Office of Science Operations, American Heart Association

Janet Wright, MD, FACC: Senior Vice President for Science and Quality, American College of Cardiology

Contractor Staff:

Dale Bratzler, DO, MPH: CEO, Principal Clinical Coordinator, Oklahoma Foundation for Medical Quality

Jo DeBuhr, RN: Project Specialist, AMI/HF Inpatient Measures, Oklahoma Foundation for Medical Quality/Colorado

<p>Foundation for Medical Care Chris Leber, RN: Project Specialist, AMI/HF Inpatient Measures, Oklahoma Foundation for Medical Quality/Colorado Foundation for Medical Care CMS Staff: Kristie Baus, MS, RN: Government Task Leader, Centers for Medicare and Medicaid Services David Nilasena, MD: Chief Medical Officer, Region VI, Centers for Medicare and Medicaid</p>
<p>Ad.2 If adapted, provide name of original measure: N/A Ad.3-5 If adapted, provide original specifications URL or attachment</p>
<p>Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 Year the measure was first released: 1999 Ad.7 Month and Year of most recent revision: 10, 2010 Ad.8 What is your frequency for review/update of this measure? Every 6 months Ad.9 When is the next scheduled review/update for this measure? 07, 2011</p>
<p>Ad.10 Copyright statement/disclaimers:</p>
<p>Ad.11 -13 Additional Information web page URL or attachment:</p>
<p>Date of Submission (MM/DD/YY): 12/27/2010</p>

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.

THE NATIONAL QUALITY FORUM

COMPOSITE MEASURE SUBMISSION FORM Version 4.1 January 2010

This form will be used by stewards to submit composite measures and by reviewers to evaluate the measures.

Measure Stewards: Check with NQF staff before using this form. Complete all non-shaded areas of the form. All requested information should be entered directly into this form. The information requested is directly related to NQF's composite measure evaluation criteria and will be used by reviewers to determine if the evaluation criteria have been met. The specific relevant subcriteria language is provided in a Word comment within the form and will appear if your cursor is over the highlighted area (or in balloons).

The measure steward has the opportunity to identify and present the information that demonstrates the measure meets the criteria. Additional materials will only be considered supplemental. Do not rely solely on materials provided at URLs or in attached documents to provide measure specifications or to demonstrate meeting the criteria. If supplemental materials are provided, be sure to indicate specific page numbers/ web page locations for the relevant information (web page links preferred).

For questions about completing this form, contact the project director at 202-783-1300. Please email this form to the appropriate contact listed in the corresponding call for measures.

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 #:	NQF Project:
De.1 Title of Measure: <u>Composite Measure of Hospital Quality for Acute Myocardial Infarction (AMI)</u>	
De.2 Brief description of measure (including type of score, measure focus, target population, time, e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year): <u>A composite measure of in-hospital process- and outcome-of-care for Acute Myocardial Infarction (AMI) patients.</u>	
De.3 Type of Measure: <input type="checkbox"/> Composite with component measures combined at patient-level (e.g., all-or-none) <input checked="" type="checkbox"/> Composite with component measures combined at aggregate-level	
Select the most relevant priority area(s), quality domain(s), and consumer need(s).	
De.4 National Priority Partners Priority Area <input type="checkbox"/> patient and family engagement <input type="checkbox"/> population health <input type="checkbox"/> <input checked="" type="checkbox"/> safety <input type="checkbox"/> care coordination <input type="checkbox"/> palliative and end of life care <input type="checkbox"/> overuse	
De.5 IOM Quality Domain <input checked="" type="checkbox"/> effectiveness <input type="checkbox"/> efficiency <input type="checkbox"/> equity <input type="checkbox"/> patient-centered <input type="checkbox"/> safety <input checked="" type="checkbox"/> timeliness	

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

De.6 Consumer Care Need Getting Better Living With Illness Staying Healthy

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
<p>A. The measure is in the public domain or an intellectual property agreement (measure steward agreement) is signed. <i>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.</i></p> <p>A.1 Do you attest that the measure steward holds intellectual property rights to the measure <u>and</u> the right to use any aspects of the measure owned by another entity (e.g., component measures, risk model, code set)? <input checked="" type="checkbox"/> Yes</p> <p>A.2 Measure Steward Agreement <input type="checkbox"/> Signed and Submitted OR <input checked="" type="checkbox"/> Government entity-public domain (If measure steward agreement not signed for non-government entities, do not submit)</p> <p>A.3 Please check if either of the following apply: <input type="checkbox"/> Proprietary Measure <input type="checkbox"/> Proprietary Complex Measure w/fees</p>	A Y <input type="checkbox"/> N <input type="checkbox"/>
<p>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. B.1 <input checked="" type="checkbox"/> Yes (If no, do not submit)</p>	B Y <input type="checkbox"/> N <input type="checkbox"/>
<p>C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement.</p> <p>C.1 Purpose: <input checked="" type="checkbox"/> Public reporting <input checked="" type="checkbox"/> Internal quality improvement</p> <p>C.2 <input checked="" type="checkbox"/> Accountability <input type="checkbox"/> Accreditation <input type="checkbox"/> Payment incentive <input type="checkbox"/> Other, describe: (If not intended for <u>both</u> public reporting <u>and</u> quality improvement, do not submit)</p>	C Y <input type="checkbox"/> N <input type="checkbox"/>
<p>D. The requested measure submission information is complete. Composite 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.</p> <p>D.1 Testing: <input checked="" type="checkbox"/> Fully developed and tested (If composite measure not tested, do not submit)</p> <p>D.2 Have NQF-endorsed measures been reviewed to identify if there are similar or related measures? <input checked="" type="checkbox"/> Yes (If no, do not submit) If there are similar or related measures, be sure to address items 3b and 3c with specific information. ► Is all requested information entered into this form? <input checked="" type="checkbox"/> Yes (If no, do not submit)</p>	D Y <input type="checkbox"/> N <input type="checkbox"/>
<p>De.7 If component measures of the composite are aggregate-level measures, <u>all</u> must be either NQF-endorsed or submitted for consideration for NQF endorsement (<i>check one</i>)</p> <p><input checked="" type="checkbox"/> All component measures are <u>NQF-endorsed</u> measures</p> <p><input type="checkbox"/> Some or all component measures are <u>not NQF-endorsed</u> and have been submitted using the online measure submission tool (If not, do not submit)</p> <p>(for NQF staff use) Have all conditions for consideration been met? Staff Notes to Steward (if submission returned):</p>	Y <input type="checkbox"/> N <input type="checkbox"/> Met Y <input type="checkbox"/> N <input type="checkbox"/>
Staff Notes to Reviewers (issues or questions regarding any criteria):	
Staff Reviewer Name(s):	

Comment [KP1]: The individual measures included in the composite or subcomposite measures must be either: NQF-endorsed; OR assessed to have met the individual measure evaluation criteria as the first step in evaluating the composite measure. (This does not apply to subscales of a scale/instrument that cannot be used independently of the total scale.)

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	

<p>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. <i>Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (composite measure evaluation criteria)</i></p>	<p>Eval</p>
<p>(for NQF staff use) <u>Specific NPP goal:</u></p>	
<p>1d. Purpose/objective of the Composite 1d.1 Describe the purpose/objective of the composite measure:</p> <p>This measure was designed specifically for use in the Centers for Medicare & Medicaid Services' (CMS) public reporting efforts for measures used in CMS' Hospital Inpatient Quality Reporting Program (formerly RHQDAPU). This program is required to publicly report the adopted measures in particular focus areas related to the quality of hospital inpatient care. The number of measures in the program has expanded considerably, and in the latest inpatient prospective payment system (IPPS) rule, CMS further expanded the measure set to include 60 measures over the next few years. The volume of measures presents a challenge for the public reporting requirement of the program to present this information in a manner that is understandable and useful. The primary objective of this measure is to summarize the measures for the Acute Myocardial Infarction (AMI) focus area into a single composite that is useful, understandable, and acceptable to a wide range of stakeholders. As a result, it is a so-called formative measure. Further discussion of the construction of formative composite measures appears in Appendix B.</p> <p>Specifically, this measure summarizes both clinical process- and outcome-of-care indicators associated with the treatment of AMI and reported for CMS' Hospital Inpatient Quality Reporting Program. Measures were adopted for this program because, based on a consensus process, they were deemed to be indicators of well-coordinated, high-quality care in the hospital inpatient setting for the clinical condition of interest. In addition, CMS sought an approach to composite methodology that was flexible and adaptable to changes in the sets of measures and clinical conditions included now and in the future of the Hospital Inpatient Quality Reporting program.</p> <p>A condition-specific composite is useful for three reasons. First, in any composite, information from a number of component measures is summarized into a single measure for more effective communication. Second, in a condition-specific composite, the component measures are aggregated at a level that is relevant to both consumers and providers. A condition-specific composite strikes a useful balance between creating one global hospital measure, which might not be relevant to individual consumers or providers with specific needs or practice spheres, and offering only the component measures, which some stakeholders could find overwhelming or contradictory and thus unhelpful. Third, condition-specific composite measures respond simply and directly to a key patient-centered question: "Which hospital should I go to, given my condition?" Moreover, the use of condition-specific composite measures permits disease-specific care teams and their management within hospitals to answer the following question: "Overall, how well is our system serving patients with this condition?"</p> <p>As background, the Hospital Inpatient Quality Reporting Program was initially developed as a result of the Medicare Prescription Drug, Improvement and Modernization Act (MMA) of 2003. Section 5001(a) of Pub. 109-171 of the Deficit Reduction Act (DRA) of 2005 set out new requirements for the program, which built on the ongoing voluntary Hospital Quality Initiative. The Hospital Inpatient Quality Reporting Program is the main effort of CMS to communicate hospital-level quality to patients and providers.</p> <p>1d.2 Describe the quality construct used in developing the composite:</p> <p>The composite measure of quality of hospital care for AMI aims to be a comprehensive indicator of hospital performance that will be of special value to consumers as a summary means of evaluating alternative hospitals. The quality construct is thus formative rather than reflective in nature. At present, CMS publishes seven individual process-of-care indicators and two outcome-of-care indicators meant to capture the quality of hospital care provided to patients with AMI. The proposed composite combines these in the form of process- and outcome-of-care domains.</p> <p>CMS developed the composite measure to achieve the following goals for reporting hospital quality measures composite methodology:</p>	<p>1d C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

Comment [KP2]: 1d. The purpose/objective of the composite measure and the construct for quality are clearly described.

- Summarize measures on Hospital Compare in a single, useful, condition-specific composite
- Produce composite values that show differences in hospital performance that are clinically and statistically meaningful and reflect true underlying differences in quality
- Enable the calculation of results for most hospitals
- Employ a method that accommodates changes in the set of measures on Hospital Compare and can be used for multiple conditions
- Employ a method that is relatively simple, so hospitals can duplicate results

These goals can be achieved by a method that is consistent with that of other widely used composites; in this case the method used for the Agency for Healthcare Research and Quality (AHRQ) composites. The National Quality Forum (NQF) has endorsed those composites and CMS, states, and other organizations use them widely.

The current Hospital Inpatient Quality Reporting Program construct domains focus on diseases important to the Medicare population: AMI, Heart Failure (HF), and Pneumonia (PN), and on quality indicators related to the Surgical Care Improvement Project (SCIP). The first three have separate sub-composites in processes-and outcomes-of-care. This system of domains and sub-composites allows addition or removal of measures without changes in methodology or weighting, as well as the publication or analysis of separate process and outcome composites within a condition if desired.

In the development of this composite, certain methodological decisions were made to satisfy the policy goals outlined above. First, we entered individual measures as values, rather than ranks, to reduce the likelihood that very small differences in absolute performance lead to large differences in ranking composite scores. Second, we imputed values for missing indicators so that the composite would define as many hospitals as possible. Third, we adjusted individual measures for reliability, a process that leads to a more accurate measure of true underlying performance and avoids extreme values for small hospitals due to random variation. Lastly, we used denominator weighting so that the composite places more weight on measures that are reported for relatively more patients nationally. In Table 1d.2.1 of Appendix A, we present the mapping between CMS' policy goals and methodological decisions in tabular form.

1e. Components and conceptual construct for quality

1e.1 Describe how the component measures/items are consistent with and representative of the quality construct:

As indicated previously, this composite measure is primarily a formative summary of the measures on Hospital Compare. Thus, the composite includes all measures associated with this condition that are reported on Hospital Compare.

That said, measures were adopted for the Hospital Inpatient Quality Reporting Program because, based on a consensus process, they were deemed to be indicators of well-coordinated, high-quality care in the hospital inpatient setting for the clinical condition of interest. For the AMI, HF, and PN composite measures, the measures that make up the composite include both process- and outcome-of-care indicators; the SCIP composite is made up of process-of-care indicators only.

The composite includes both process- and outcome-of-care indicators, because both types of indicators contain information about quality of care. While it is not possible to directly assess an abstract concept such as quality of care, process-of-care indicators that evaluate whether certain best practices were executed provide critical insight into a hospital's care delivery system. For example, for the AMI composite measure, the component process-of-care indicators evaluate whether a patient received:

- Aspirin on arrival
- Aspirin at discharge
- ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD)
- Smoking Cessation advice/counseling
- Beta Blocker at discharge
- Fibrinolytic medication within 30 minutes of arrival
- PCI within 90 minutes of arrival

These NQF-endorsed process-of-care indicators represent established best practices for AMI care (1, 2) and

1e
C
P
M
N

Comment [KP3]: 1e. The component items/measures (e.g., types, focus) that are included in the composite are consistent with and representative of the conceptual construct for quality represented by the composite measure. Whether the composite measure development begins with a conceptual construct or a set of measures, the measures included must be conceptually coherent and consistent with the purpose.

<p>CMS adopted them for the Hospital Inpatient Quality Reporting Program initiative. As standards in clinical practice evolve, additions or changes to these component measures are likely to follow, as well as developing expansions into other conditions and disease states.</p> <p>In addition to reflecting current clinical guidelines, studies have shown a clear relationship between execution of these practices and decreased mortality for AMI patients (3-5), one of the two outcome-of-care indicators also included in the proposed AMI composite measure. The two AMI outcome-of-care component measures are: 1) 30-day risk-standardized mortality and 2) 30-day risk standardized all-cause readmission. Similar to the process-of-care indicators, these two outcome-of-care indicators are NQF-endorsed and part of CMS' Hospital Inpatient Quality Reporting Program initiative. They directly report the rate of the undesired outcomes (mortality or readmission) that AMI patients at a given hospital experience, and therefore might be critical to understanding the quality of care received.(i)</p> <p>The combination of these component indicators, each of which is intended to indicate the quality of care received for a subset of patients (that is, AMI, HF, PN, or SCIP), ultimately serves to deliver a single, useful, condition-specific summary for consumer use.</p> <p>Citations</p> <ol style="list-style-type: none"> Anderson, JL, Adams, CD, Antman, EM, et al. ACC/AHA 2007 guidelines for the management of patients with unstable angina/non-ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines for the Management of Patients with Unstable Angina/Non-ST-Elevation Myocardial Infarction): developed in collaboration with the American College of Emergency Physicians, American College of Physicians, Society for Academic Emergency Medicine, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol. 2007; 50:e1-157. Antman, EM, Anbe, DT, Armstrong ,PW, et al. 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. Smith, SC, Allen, J, Blair, SN, et al. AHA/ACC guidelines for secondary prevention for patients with coronary and other atherosclerotic vascular disease: 2006 update. J Am Coll Cardiol. 2006; 47:2130-9. Flather, MD, Yusuf, S, Kober, L, et al. Long-term ACE-inhibitor therapy in patients with heart failure or left-ventricular dysfunction: A systematic overview of data from individual patients. ACE-Inhibitor Myocardial Infarction Collaborative Group. Lancet. 2000; 355(9215):1575-1581. Yusuf, S, Wittes, J, & Friedman, L. Overview of results of randomized clinical trials in heart disease. I. Treatments following myocardial infarction. JAMA. 1988; 260(14):2088:2093 <p>Footnotes</p> <p>i. In order to align these two indicators with the process-of-care indicators, which report desired, rather than undesired, outcomes, each outcome-of-care indicator is subtracted from 100. This produces two desired outcomes - lack of 30-day mortality and lack of 30-day readmission - which are incorporated into the composite measure.</p>	
<p>If the component measures are combined at the patient level, complete 1a, 1b, and 1c.</p>	
<p>If the component measures are combined at the aggregate level, skip to criterion 2, <i>Scientific Acceptability of Measure Properties</i> (individual measures are either NQF-endorsed or submitted individually).</p>	
<p>1a. High Impact</p> <p>1a.1 Demonstrated high impact aspect of healthcare (<i>Select the most relevant</i>)</p> <p><input type="checkbox"/> affects large numbers <input type="checkbox"/> frequently performed procedure <input type="checkbox"/> leading cause of morbidity/mortality</p> <p><input type="checkbox"/> high resource use <input type="checkbox"/> severity of illness <input type="checkbox"/> patient/societal consequences of poor quality</p> <p><input type="checkbox"/> other, describe: 1a.2</p> <p>1a.3 Summary of Evidence of High Impact:</p> <p>1a.4 Citations for Evidence of High Impact:</p>	<p>1a</p> <p>H <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>L <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>1b. Opportunity for Improvement</p> <p>1b.1 Briefly explain benefits (improvements in quality) envisioned by use of this measure:</p>	<p>1b</p> <p>H <input type="checkbox"/></p>

Comment [KP4]: 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).

Comment [KP5]: 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).

<p>1b.2 Summary of data demonstrating performance gap (<i>variation or overall poor performance across providers</i>):</p> <p>1b.3 Citations for data on performance gap:</p> <p>1b.4 Summary of Data on disparities by population group:</p> <p>1b.5 Citations for data on Disparities:</p>	<p>M <input type="checkbox"/></p> <p>L <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>1c. Evidence-based</p> <hr style="border-top: 1px dashed #ccc;"/> <p>1c.1 Relationship to Outcomes (<i>For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population.</i>)</p> <p>1c.2 Type of Evidence (<i>Check all that apply</i>)</p> <p><input type="checkbox"/> Cohort study <input type="checkbox"/> Evidence-based guideline <input type="checkbox"/> Expert opinion <input type="checkbox"/> Meta-analysis</p> <p><input type="checkbox"/> Observational study <input type="checkbox"/> Randomized controlled trial <input type="checkbox"/> Systematic synthesis of research</p> <p><input type="checkbox"/> Other (<i>Please describe</i>): 1c.3</p> <p>1c.4 Summary of Evidence as described above for type of measure; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome):</p> <p>1c.5 Rating of strength/quality of evidence (<i>also provide narrative description of the rating and by whom</i>)</p> <p>1c.6 Method for rating evidence:</p> <p>1c.7 Summary of Controversy/Contradictory Evidence:</p> <p>1c.8 Citations for Evidence (<i>other than guidelines</i>)</p> <p>1c.9 Quote the Specific guideline recommendation (<i>including guideline number and/or page number</i>)</p> <p>1c.10 Clinical Practice Guideline Citation:</p> <p>1c.11 National Guideline Clearinghouse or other URL:</p> <p>1c.12 Rating of strength of recommendation (<i>also provide narrative description of the rating and by whom</i>)</p> <p>1c.13 Method for rating strength of recommendation (<i>If different from USPSTF system, also describe rating and how it relates to USPSTF</i>):</p> <p>1c.14 Rationale for using this guideline over others:</p>	<p>1c</p> <p>H <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>L <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Importance to Measure and Report</i>?</p>	<p>1</p>
<p>Steering Committee: Was the threshold criterion, <i>Importance to Measure and Report</i>, met? Rationale:</p>	<p>1</p> <p>Y <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES</p>	
<p>Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (composite measure evaluation criteria)</p>	<p>Eval</p>
<p>2a. COMPOSITE MEASURE SPECIFICATIONS</p>	
<p><i>In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained?</i></p> <p>S.1 Do you have a web page where current detailed measure specifications can be obtained? Upon endorsement, the proposed measure specifications will be posted on the Hospital Compare website: http://www.hospitalcompare.hhs.gov/</p>	<p>2a-specs</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p>

Comment [KP6]: 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:
 - Intermediate outcome - evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
 - Process - 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).
 - Structure - 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.
 - Patient 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.
 - Access - evidence that an association exists between access to a health service and the outcomes of, or experience with, care.

Efficiency - 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.

<p>S.2 If yes, provide web page URL: http://www.hospitalcompare.hhs.gov/</p>	<p><input type="checkbox"/> N</p>
<p>2a. Precisely Specified</p> <p>2a.0.1 Components of the Composite (<i>List the components, i.e., domains/sub-composites, individual measures. If component measures are <u>NQF-endorsed</u>, include NQF measure number; if <u>not NQF-endorsed</u>, provide date of submission to NQF</i>)</p> <p>HOSPITAL PROCESS-OF-CARE INDICATORS</p> <ol style="list-style-type: none"> 1. Percent of AMI Patients Given Aspirin on Arrival (NQF #0132; Endorsed May 9, 2007) 2. Percent of AMI Patients Given Aspirin at Discharge (NQF #0142; Endorsed May 9, 2007) 3. Percent of AMI Patients Given ACE Inhibitor or ARB for LVSD (NQF #0137; Endorsed May 9, 2007) 4. Percent of AMI Patients Given Smoking Cessation Advice/Counseling (NQF #0027; Endorsed May 1, 2006) 5. Percent of AMI Patients Given Beta Blocker at Discharge (NQF #0160; Endorsed May 9, 2007) 6. Percent of AMI Patients Given Fibrinolytic Medication within 30 Min. of Arrival (NQF #0164; Endorsed May 9, 2007) 7. Percent of AMI Patients Given PCI within 90 Min. of Arrival (NQF #0163; Endorsed May 9, 2007) <p>HOSPITAL OUTCOME-OF-CARE INDICATORS</p> <ol style="list-style-type: none"> 1. AMI 30-day Risk-Standardized Mortality (NQF #0230; Endorsed May 9, 2007) 2. AMI 30-day Risk-Standardized Readmission (NQF #0505; Endorsed Oct. 28, 2008) 	
<p><i>If the composite measure cannot be specified with a numerator and denominator, please consult with NQF staff.</i></p>	
<p><i>If the component measures are combined at the aggregate level, do not include the individual measure specifications below.</i></p>	
<p>2a.1 Composite Numerator Statement: The sum of all successes for acute myocardial infarction process-of-care indicators, weighted by one-half the reciprocal of the share of opportunities represented by acute myocardial infarction process-of-care indicators in total opportunities, plus the sum of all successes for acute myocardial infarction outcome-of-care indicators, weighted by one-half the reciprocal of the share of opportunities represented by acute myocardial infarction outcome-of-care indicators in total opportunities.</p>	
<p>2a.2 Numerator Time Window: July 2006 - June 2009</p>	
<p>2a.3 Numerator Details: Successes in the following acute myocardial infarction process-of-care and outcome of care indicators:</p>	
<p>HOSPITAL PROCESS-OF-CARE INDICATORS</p> <ol style="list-style-type: none"> 1. Percent of AMI Patients Given Aspirin on Arrival (NQF #0132) 2. Percent of AMI Patients Given Aspirin at Discharge (NQF #0142) 3. Percent of AMI Patients Given ACE Inhibitor or ARB for LVSD (NQF #0137) 4. Percent of AMI Patients Given Smoking Cessation Advice/Counseling (NQF #0027) 5. Percent of AMI Patients Given Beta Blocker at Discharge (NQF #0160) 6. Percent of AMI Patients Given Fibrinolytic Medication within 30 Min. of Arrival (NQF #0164) 7. Percent of AMI Patients Given PCI within 90 Min. of Arrival (NQF #0163) <p>HOSPITAL OUTCOME-OF-CARE INDICATORS</p> <ol style="list-style-type: none"> 1. AMI 30-day Risk-Standardized Mortality (NQF #0230) 2. AMI 30-day Risk-Standardized Readmission (NQF #0505) 	
<p>2a.4 Composite Denominator Statement: The total number of opportunities for success on all acute myocardial infarction indicators used in the composite.</p>	

Comment [KP7]: 2a. The composite measure is well defined and precisely specified so that it can be implemented consistently within and across organizations and allow for comparability. Composite specifications include methods for standardizing scales across component scores, scoring rules (i.e., how the component scores are combined or aggregated), weighting rules (i.e., whether all component scores are given equal or differential weighting when combined into the composite), handling of missing data, and required sample sizes.

2a.5 Target Population Gender Female Male

2a.6 Target Population Age range Aged 18 and over.

2a.7 Denominator Time Window: July 2006 - June 2009.

2a.8 Denominator Details: Counts of process-of-care opportunities are based on hospital acute myocardial infarction quality reports. Counts of outcome-of-care opportunities are based on claims data.

2a.9 Composite Denominator Exclusions: Hospitals missing three or more acute myocardial infarction process-of-care indicators and one or more outcome-of-care indicator were excluded.

2a.10 Denominator Exclusion Details: Hospitals missing three or more of the acute myocardial infarction process-of-care indicators and one or more of the outcome-of-care indicators listed below were excluded from the composite calculation.

HOSPITAL PROCESS-OF-CARE INDICATORS

1. Percent of AMI Patients Given Aspirin on Arrival (NQF #0132)
2. Percent of AMI Patients Given Aspirin at Discharge (NQF #0142)
3. Percent of AMI Patients Given ACE Inhibitor or ARB for LVSD (NQF #0137)
4. Percent of AMI Patients Given Smoking Cessation Advice/Counseling (NQF #0027)
5. Percent of AMI Patients Given Beta Blocker at Discharge (NQF #0160)
6. Percent of AMI Patients Given Fibrinolytic Medication within 30 Min. of Arrival (NQF #0164)
7. Percent of AMI Patients Given PCI within 90 Min. of Arrival (NQF #0163)

HOSPITAL OUTCOME-OF-CARE INDICATORS

1. AMI 30-day Risk-Standardized Mortality (NQF #0230)
2. AMI 30-day Risk-Standardized Readmission (NQF #0505)

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions*):
None.

2a.18 Type of Score: Weighted score/comosite/scale 2a.19 If "Other", please describe: N/A

2a.20 Interpretation of Score (*Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score*)
Better quality = Higher score

2a.42 Method of Scoring/Aggregation: other 2a.43 If "other" scoring method, describe:

The composite measure was calculated with a method that we have termed "Absolute Scoring Index with Reliability Weighting" (ASI-RW). The composite is actually derived by combining two sub-composites, one incorporating the process-of-care indicators and the other incorporating the outcome-of-care indicators.

The process-of-care sub-composite is derived by applying reliability weights to each individual process-of-care indicator, such that each hospital-specific indicator is based on the actual reported data for that indicator as well as the national mean for that indicator. The resulting adjusted rates are then weighted and added together to form the process-of-care sub-composite. The weight used to combine indicators is based on the national number of patients included in the indicator (denominator weighting), so that if one indicator is relevant to twice as many patients as another, the weight of that indicator in the composite is twice as large as the weight of the other. Many composite measures that NQF has approved use this patient measure opportunity basis; it has the advantage of focusing the outcome of the measurement process on the places where opportunities to provide appropriate evidence-based process care are greatest.

To calculate the outcome-of-care sub-composite, we first subtracted the individual indicators from 100 to create two desired outcomes: 1) survival rates, which replaced mortality and 2) absence of readmission, which replaced readmission. We then applied denominator weighting, once again, to estimate the outcome-

of-care sub-composite.

After generating process- and outcome-of-care sub-composite scores, each is scaled by subtracting the overall domain mean and dividing by the standard deviation (a statistical process to derive a standardized score). The two sub-composites are combined using a simple average. To map the standardized composite score to a scale between zero to one hundred, we then add the lowest possible score a hospital can receive (i.e., a hospital scores zero percent on all process- and outcome-of-care indicators) and divide by the range of potential hospital scores (i.e., the difference between the highest possible score a hospital can receive, which is if a hospital scores 100 percent on all process- and outcome-of-care indicators, and the lowest possible score).

2a.44 Missing Component Scores (*Indicate how missing component scores are handled*):

The AMI composite measure is generated for all hospitals that reported data for at least four of the seven process-of-care indicators and one of the two outcome-of-care indicators. For hospitals that meet these criteria but are missing data for some of the component measures, missing values are imputed using the national mean.

2a.45 Weighting: Equal Differential **2a.46** If differential weighting, describe:

Consistent with the approach used for the AHRQ measures, CMS used denominator weighting in constructing the process-of-care sub-composite. Denominator weighting places relatively more weight on measures that apply to relatively more patients nationally. Please see Appendix A for complete details on weighting methodology.

2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps*):

Please note: Complete information on the calculation algorithm, including equations, are contained in Appendix A. The text summary follows below.

STEP 1

Hospital process-of-care indicators for AMI, with a data collection period of July 2008 to June 2009, and outcome-of-care indicators for AMI, with a data collection period of July 2006 to June 2009, that are publicly reported on Hospital Compare (<http://www.hospitalcompare.hhs.gov/>), are combined into a single data set using hospital provider identification numbers.

STEP 2

Process-of-care indicators are reliability-weight adjusted. That is, the value of each process-of-care indicator is set equal to the weighted average of the hospital's own mean for the indicator and the national mean for the indicator. The weights are based on the between-hospital variance and the within-hospital variance in indicator scores (for more information on this adjustment, see the "Estimation of the Reliability-Weight-Adjusted Measures," which follows).(ii)

STEP 3

Hospitals missing process- or outcome-of-care indicators are imputed with the national mean.(iii) The national mean of the process-of-care indicators are estimated as a simple average of the indicators. The national mean of the outcome-of-care indicators are provided by Hospital Compare.

STEP 4

The process-of-care sub-composite score is computed using denominator weights, where the denominator weight is based on the number of hospital cases for each process-of-care indicator (see Appendix A, "Estimation of the Absolute Score Index with Reliability Weighting Composite Measure," eq. 2a.21.5).

STEP 5

The outcome-of-care sub-composite score is also computed using denominator weights (see Appendix A, "Estimation of the Absolute Score Index with Reliability Weighting Composite Measure," Equation 2a.21.6).

STEP 6

To standardize the process- and outcome-of-care sub-composite measures, each are scaled by subtracting

the overall sub-composite mean and dividing by the standard deviation. Then the average of the process- and outcome-of-care sub-composites is estimated (see Appendix A, “Estimation of the Absolute Score Index with Reliability Weighting Composite Measure,” Equation 2a.21.8).

STEP 7

Lastly, in order to have a composite measure with values between zero and 100, we add the lowest possible score a hospital can receive (i.e., a hospital scores zero percent on all process- and outcome-of-care indicators) and divide by the range of potential hospital scores (i.e., the difference between the highest possible score a hospital can receive, which is if a hospital scores 100 percent on all process- and outcome-of-care indicators, and the lowest possible score) (see Appendix A, “Estimation of the Absolute Score Index with Reliability Weighting Composite Measure,” Equation 2a.21.7).

ESTIMATION OF RELIABILITY-WEIGHT-ADJUSTED MEASURES

For each process-of-care indicator, the reliability-weight-adjusted indicator is equal to a weighted average of the hospital’s own measure and the national mean value of the measure. In each case, the weight is a measure of the precision with which a hospital’s measure has been estimated. This weighted average has been shown to be more accurate, on average, than using each hospital’s individual value for the measure.

The weight is made up of two parts—the variability of the measure within each hospital, termed the “within variance” or “noise variance,” and the variability across hospitals, known as the “signal variance.” The weight attached to each hospital’s own value for process measure k is equal to the ratio of the signal variance to the sum of the signal variance and the noise variance. As the number of observations for a hospital (nj_k) increases, the weight approaches one. Please see Appendix A for complete calculation details.

ESTIMATION OF THE ABSOLUTE SCORE INDEX WITH RELIABILITY WEIGHTING (ASI-RW) COMPOSITE MEASURE

We estimate the composite measure using an approach that we have termed absolute score index with reliability weighting (ASI-RW). To compute the ASI-RW, we first computed process- and outcome-of-care sub-composite scores. Using process-of-care indicators that are set equal to the weighted average of the hospital’s own mean for the indicator and the national mean for the indicator (that is, reliability-weight adjusted), the process-of-care sub-composite score is computed as a denominator-weighted average of the process-of-care indicators. That is, weights of each process-of-care indicator are based on the opportunities for providing a specific recommended treatment and greater weights are placed on measures that apply to relatively more patients nationally. Similarly, the outcome-of-care sub-composite score is also estimated as a denominator-weighted average of the outcome-of-care indicators, which are reported on Hospital Compare and are risk-adjusted.

To standardize each measure, the process- and outcome-of-care sub-composite scores are scaled by subtracting the overall sub-composite mean and dividing by the standard deviation. The ASI-RW composite measure is computed using two steps. First, the average of the process- and outcome-of-care sub-composites is estimated. Then, to map the standardized composite score to a scale between zero and 100, we add the lowest possible score a hospital can receive (i.e., a hospital scores zero percent on all process- and outcome-of-care indicators) and divide by the range of potential hospital scores (i.e., the difference between the highest possible score a hospital can receive, which is if a hospital scores 100 percent on all process- and outcome-of-care indicators, and the lowest possible score). Please see Appendix A for complete calculation details.

Footnotes

- ii. Hospital outcome-of-care indicators are not reliability-weight adjusted because they have been risk-standardized using a method that accounts for reliability previously, before public reporting on Hospital Compare.
- iii. The use of the national mean is consistent with the approach used for the AHRQ quality composites. It is simple, already in use, and perceived as fair by providers.

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

Please note: Complete information on the method for discriminating performance, including equations, are contained in Appendix A. The text summary follows below.

To examine meaningful differences in composite measures among hospitals, for the purpose of internal

analysis, we compared hospitals' confidence interval estimates with the overall mean and assigned hospitals into one of three performance categories: better than hospitals, if the interval estimate is entirely above the mean; no different than hospitals, if the interval estimate includes the mean; and worse than hospitals, if the interval estimate is entirely below the mean. These categories were used for illustrative analyses only and should not be assumed to be the manner in which these composites will be publicly reported.

The hospital-specific standard error is estimated by computing the variance of the composite measure and computing a square root of the variance. After we derive the standard errors for each hospital, we estimate an interval estimate around each hospital's mean composite measure. The interval estimate is a range of probable values for the composite measure that characterizes the amount of uncertainty associated with the estimate. We apply a 95 percent interval estimate, which indicates a 95 percent confidence level that the true composite measure is between the lower and upper limits of the interval.

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample (or conducting the survey) and guidance on minimum sample size (response rate):*
N/A

2a.24 Data Source *Check all the source(s) used in the component measures.*

- | | |
|--|--|
| <input type="checkbox"/> Documentation of original self-assessment (e.g., SF-36) | <input type="checkbox"/> Paper Medical Record/flowsheet |
| <input checked="" type="checkbox"/> Electronic administrative data/ claims | <input type="checkbox"/> Pharmacy data |
| <input type="checkbox"/> Electronic Clinical Data (e.g., MDS) | <input type="checkbox"/> Public health data/vital statistics |
| <input type="checkbox"/> Electronic Health/Medical Record | <input type="checkbox"/> Registry data |
| <input type="checkbox"/> External audit | <input type="checkbox"/> Survey-patient (e.g., CAHPS) |
| <input type="checkbox"/> Lab data | <input type="checkbox"/> Survey-provider |
| <input type="checkbox"/> Management data | <input type="checkbox"/> Special or unique data, specify: |
| <input type="checkbox"/> Organizational policies and procedures | |

2a.25 Data source or collection instrument *(Identify the specific data source or data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.):*

The composite is constructed from component measures posted on the Hospital Compare website.

2a.26 Data source/data collection instrument attached OR **2a.27** at web page URL:
<http://www.hospitalcompare.hhs.gov/>

2a.29 Data dictionary/code table attached OR **2a.30** at web page URL:
<http://www.hospitalcompare.hhs.gov/>

2a.32 Level of Measurement/Analysis *(Check the level for which the measure is specified and tested)*

- | | |
|---|---|
| Clinicians: <input type="checkbox"/> Individual <input type="checkbox"/> Group <input type="checkbox"/> Other | <input type="checkbox"/> Prescription drug plan |
| <input checked="" type="checkbox"/> Facility/Agency (e.g., hospital, nursing home) | Program: <input type="checkbox"/> Disease management <input type="checkbox"/> QIO |
| <input type="checkbox"/> Health plan | <input type="checkbox"/> Other |
| <input type="checkbox"/> Integrated delivery system | |
| <input type="checkbox"/> Multi-site/corporate chain | <input type="checkbox"/> Measured at all levels |
| Population: <input type="checkbox"/> National <input type="checkbox"/> Regional/network | <input type="checkbox"/> Other (Please describe): |
| <input type="checkbox"/> State <input type="checkbox"/> Counties/Cities | |

2a.26 Care Settings *(Check the settings for which the measure is specified and tested; check all that apply)*

- | | |
|---|---|
| Ambulatory Care: <input type="checkbox"/> Amb Surgery Center <input type="checkbox"/> Office <input type="checkbox"/> Clinic <input type="checkbox"/> Emergency Dept <input type="checkbox"/> Hospital Outpatient | <input checked="" type="checkbox"/> Hospital |
| <input type="checkbox"/> Assisted Living | <input type="checkbox"/> Long term acute care hospital |
| <input type="checkbox"/> Behavioral health/psychiatric unit | <input type="checkbox"/> Nursing home/ Skilled Nursing Facility (SNF) |
| <input type="checkbox"/> Dialysis Facility | <input type="checkbox"/> Rehabilitation Facility |
| <input type="checkbox"/> Emergency medical services/ambulance | <input type="checkbox"/> All settings |
| <input type="checkbox"/> Group Home | <input type="checkbox"/> Unspecified or "not applicable" |
| <input type="checkbox"/> Home | <input type="checkbox"/> Other (Please describe): |
| <input type="checkbox"/> Hospice | |

2a.38 Clinical Services *(Healthcare services being measured; all that apply.)*

<p>Behavioral Health:</p> <input type="checkbox"/> Mental health <input type="checkbox"/> Substance use treatment <input type="checkbox"/> Other	<input checked="" type="checkbox"/> Physicians (MD/DO) <input type="checkbox"/> Podiatrist <input type="checkbox"/> Psychologist/LCSW <input type="checkbox"/> PT/OT/Speech <input type="checkbox"/> Respiratory Therapy <input type="checkbox"/> Other
<p>Clinicians:</p> <input type="checkbox"/> Audiologist <input type="checkbox"/> Chiropractor <input type="checkbox"/> Dentist/Oral surgeon <input type="checkbox"/> Dietician/Nutritional professional <input checked="" type="checkbox"/> Nurses <input type="checkbox"/> Optometrist <input type="checkbox"/> PA/NP/Advanced Practice Nurse <input type="checkbox"/> Pharmacist	<input type="checkbox"/> Dialysis <input type="checkbox"/> Home health <input type="checkbox"/> Hospice/Palliative care <input type="checkbox"/> Imaging services <input type="checkbox"/> Laboratory <input type="checkbox"/> Other

If the component measures are combined at the patient level and include outcomes, complete the following

2a.12 Risk Adjustment Type: No risk adjustment necessary analysis by subgroup case-mix adjustment paired data at patient level risk-adjustment devised specifically for this measure/condition risk adjustment method widely or commercially available Other (specify) 2a.13

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

2a.15 Detailed risk model attached OR 2a.16 at web page URL:

TESTING/ANALYSIS

2i. Component item/measure analysis to justify inclusion in composite

2i.1 Data/sample:

As noted in Section 1d, the purpose of the proposed composite is to summarize the process- and outcome-of-care indicators associated with treatment of AMI that are now reported under the Hospital Inpatient Quality Reporting Program. Because we do not justify the composite in terms of the behavior of those indicators, our analysis aims to document the strength of associations among them; we are interested in the extent to which our formative measure does in fact represent a single construct of well-coordinated, high-quality care.

The analysis reported here relies on data that are publicly reported on Hospital Compare. We merged process-of-care indicators for AMI collected between July 2008 and June 2009 and outcome-of-care indicators for AMI collected between July 2006 and June 2009. A total of 4,990 hospitals were reported on Hospital Compare during this time period. Of these, we estimated AMI composite measures for 2,738 hospitals, with non-missing data for at least four of the seven process-of-care indicators and at least one of the two outcome-of-care indicators.

The seven AMI hospital process-of-care indicators used in the construction of composites were drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2008 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI were based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009. It is important to bear in mind that process-of-care indicators were reported for all patients and that outcome-of-care indicators were computed from claims data for Medicare patients only.

2i.2 Analytic Method:

We carried out two analyses to explore the structure of the AMI indicators. First, we examined correlations among all process- and outcome-of-care indicators. Second, we conducted an exploratory factor analysis on the same process- and outcome-of-care indicators. Results appear in Appendix A, Tables 2i.3.1 and 2i.3.2

2i.3 Results:

Comment [KP8]: 2i. Component item/measure analysis (e.g., various correlation analyses such as internal consistency reliability), demonstrates that the included component items/measures fit the conceptual construct; OR justification and results for alternative analyses are provided.

2i
 C
 P
 M
 N

<p>Please see Appendix A for complete details on results. The text summary follows below.</p> <p>All correlations are positive, as Table 2i.3.1 (see Appendix A) shows, though many are weak, with values below 0.10. The two time-sensitive indicators (AMI 7A and AMI 8A) exhibit low correlation with other indicators. This is probably due to the high frequency of missing values for these two measures and their replacement with the overall mean. Correlations between process- and outcome-of-care indicators are low, though consistently positive. In addition, the Cronbach's alpha is 0.48, which is slightly below the commonly desired value of 0.70.</p> <p>The factor analysis of component measures (Table 2i.3.2, see Appendix A) produced a single factor with an eigenvalue greater than one. The eigenvalue for the first factor was more than 10 times that of the second factor, strongly suggesting that the component indicators represent a single underlying construct.</p>	
<p>2j. Component item/measure analysis of contribution to variability in composite score</p> <p>2j.1 Data/sample:</p> <p>As noted in Section 1d, the purpose of the proposed composite is to summarize the process- and outcome-of-care indicators associated with treatment of AMI that are now reported under the Hospital Inpatient Quality Reporting Program. Because we do not justify the composite in terms of the behavior of individual indicators, our analysis aims to document their contributions to the measure.</p> <p>Analysis of the contribution of component items to the variability in composite scores uses data that are publicly reported on Hospital Compare. We merged process-of-care indicators for AMI with a data collection period of July 2008 to June 2009 and outcome-of-care indicators for AMI with a data collection period of July 2006 to June 2009. A total of 4,990 hospitals were reported on Hospital Compare during this time period. Of these, we estimated composite measures for 2,738 hospitals, for which less than or equal to three process-of-care indicators and less than or equal to one outcome-of-care indicator is missing.</p> <p>The seven hospital process-of-care indicators related to AMI that are used in the construction of the AMI composite are calculated from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2008 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI are based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.</p> <p>2j.2 Analytic Method:</p> <p>We compare the percentage change in (1) the variance and (2) the inter-quartile range (IQR) of the process- and outcome-of-care sub-composites when a process- or outcome-of-care indicator is removed before normalization. Results appear in Appendix A, Table 2j.3.1.</p> <p>2j.3 Results:</p> <p>Please see Appendix A for complete details on results. The text summary follows below.</p> <p>In Table 2j.3.1 (Appendix A), the positive values indicate that addition of the component indicator tends to reduce the variance or IQR. Only one indicator, AMI2 (aspirin at discharge), exhibits a nontrivial positive effect on the composite variance, probably because of its relatively strong positive correlation with other component indicators (see Table 2i.3.1). Because the outcome domain contains only two component indicators, readmission and mortality both have strong negative effects on the variance of the sub-composite measure. The strong variance-reducing effect of readmission appears to be the result of its tight distribution (see Table 2l.3.2, Appendix A).</p>	<p>2j C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>2k. Analysis to support differential weighting of component scores</p> <p>2k.1 Data/sample:</p> <p>In constructing the composite, individual component indicators are weighted, in each instance, by the number of observations for the indicator. The most frequently reported indicators therefore affect the composite most strongly. In addition, the weighting scheme tends to reduce the variance of the composite,</p>	<p>2k C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

Comment [KP9]: 2j. Component item/measure analysis demonstrates that the included components contribute to the variation in the overall composite score; OR if not, justification for inclusion is provided.

Comment [KP10]: 2k. The scoring/aggregation and weighting rules are consistent with the conceptual construct. (Simple, equal weighting is often preferred unless differential weighting is justified. Differential weights are determined by empirical analyses or a systematic assessment of expert opinion or values-based priorities.)

though this effect might be muted if individual indicators have similar distributions.

Testing to support differential weighting of composite scores uses data that are publicly reported on Hospital Compare by CMS. We merged process-of-care indicators for AMI with a data collection period of July 2008 to June 2009 and outcome-of-care indicators for AMI with a data collection period of July 2006 to June 2009. A total of 4,990 hospitals were reported on Hospital Compare during this period. Of these, we estimated AMI composite measures for 2,738 hospitals, for which less than or equal to three process-of-care indicators and less than or equal to one outcome-of-care indicator is missing.

The seven hospital process-of-care indicators related to AMI that are used in the construction of composites are drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2008 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI are based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.

2k.2 Analytic Method:

We compare the distribution of the AMI composite measure with equal and differential weighting. Please see Appendix A for complete details on the analytic method, including equations.

2k.3 Results:

Please see Appendix A for complete details on results. The text summary follows below.

Table 2k.3.1 (Appendix A) displays the distribution of the AMI composite measure with equal and differential weighting. As the table shows, denominator weighting has little effect on the distribution of the composite. The median is slightly larger when denominator weighting is used, and the inter-quartile range is somewhat smaller.

2k.4 Describe how the method of scoring/aggregation achieves the stated purpose and represents the quality construct:

The objective of the composite is to summarize the component measures in a useful and scientifically acceptable manner.

Because composites are most useful to consumers if differences in composite values are clinically and statistically meaningful and reflect true differences in underlying quality, CMS entered component measures as values, not ranks, and adjusted those values for reliability. CMS entered component measures as values rather than ranks to prevent slight differences in composite values from producing large differences in composite values, as can occur when indicators are tightly distributed across hospitals. CMS also adjusted the component indicators for reliability so that random variation did not drive small hospitals to extremes; 30-day outcome measures are adjusted for reliability before publication on Hospital Compare. Process measures are not adjusted for reliability before publication; the adjustment is made as part of the compositing process.

In addition, because composites are more useful to consumers if they emphasize measures that are relevant to a larger number of consumers, CMS constructed the process- and outcome-of-care composite scores using weights based on national denominators.

When sample sizes are equal, each component process measure contributes equally to the AMI process-of-care domain score. The same is true for each component outcome-of-care indicator. Thus a hospital that improves in any component will necessarily produce an increase in its composite score. Hospitals can therefore choose where to focus improvement efforts in evidence-based processes of care. Similar logic applies to the outcome-of-care domain score. The composite thus fully reflects the AMI process and outcome-of-care indicators and represents the quality construct expressed earlier.

2k.5 Indicate if any alternative scoring/aggregation methods were tested and why not chosen:

In addition to the preferred compositing approach, ASI-RW, two alternative scoring methods were analyzed.

These are referred to as (1) the absolute scoring index (ASI) and (2) the modified relative quality index (MRQI).

1. Absolute Scoring Index

ASI is similar to CMS' preferred approach but component indicators are not reliability-adjusted. To compute the ASI composite measure, process- and outcome-of-care sub-composite scores are first computed as the equally-weighted average of the indicators. These process and outcome scores are then scaled by subtracting the overall domain mean and dividing by the standard deviation. The ASI is then computed as the average of these two scaled means. The measure is then mapped to a scale between zero and 100, for ease of interpretation.

2. Modified Relative Quality Index

MRQI is similar to CMS' preferred approach but component indicators are not reliability-adjusted and enter the composite as ranks, not values. To compute the MRQI composite measure, scores for process- and outcome-of-care sub-composites are computed as the mean of the hospital's ranks for each indicator. The composite score is then computed as the simple mean of the two domain scores. It is closely related to the relative quality index (RQI) as described by Tompkins et al. (1) (January 2009, August 2009).

In Table 2k.5.1 (see Appendix A), we present distributions of the three alternative scoring methods. Broadly speaking, the distributions for ASI-RW, the preferred approach, look quite similar to the distribution for ASI.(iv) The difference is that the reliability adjustment has reduced the likelihood of erroneously classifying small hospitals as outliers due to random variation in measured performance by pulling them toward the mean of the distribution, though this is not visible in the table itself.

Results for MRQI show a more balanced distribution with medians close to means and less pronounced clustering in the upper half of the distribution, although there is still some clustering. Note that although this approach makes the distribution look more balanced, it does not address the fundamental problems of highly clustered performance on the underlying measures, small numbers of observations, and difficulty identifying meaningful differences in performance.

Citations

1. Tompkins, Chris, Grant Ritter, Andrew Ryan, Wato Nsa, James Burgess, and Dale Bratzler. Composite Measures for Public Reporting: Final Report. January 16, 2009 (updated August, 2009).

Footnotes

- iv. Although the shrinking process pulls the process-of-care indicators toward the mean, shrinking does not result in a smaller standard deviation of the distribution of the final composite values because each domain is normalized.

2I. Analysis of missing component scores

2I.1 Data/sample:

The seven hospital process-of-care indicators used in the construction of the composite are drawn from reports for patients discharged between July 2008 and June 2009. Outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI are based on Medicare claims for stays with discharge dates between July 2006 and June 2009.(v) Process and outcome indicators were reported on Hospital Compare for 4,990 hospitals during this period.

Because some hospitals did not report all indicators, some method for dealing with missing data was required. In order to compute a composite AMI measure for the greatest possible number of hospitals, we followed two principles:

1. All seven process-of-care indicators and both outcome-of-care indicators were included.
2. A composite measure was computed for each hospital that reported four or more process-of-care indicators and at least one outcome-of-care indicator.

AMI composites were computed for all hospitals that satisfied the second item. The national mean was used to impute a value for any missing process- or outcome-of-care indicators. AMI composite measures were

2I
C
P
M
N

Comment [KP11]: 2I. Analysis of missing component scores supports the specifications for scoring/aggregation and handling of missing component scores.

<p>computed for 2,738 hospitals.</p> <p>Footnotes v. The reporting periods for the process- and outcome-of-care measures represent the most recent version of both measures available on the Hospital Compare website, at the time of this NQF submission.</p> <p>2l.2 Analytic Method:</p> <p>We used two approaches to conduct analysis of missing component scores. First, we tested to identify if there were differences in the composite measure for hospitals that did not fit the criteria stated previously. That is, we assessed whether composite measures differed significantly when our sample included composites with data for at least four process-of-care indicators and at least one outcome-of-care indicator (2,738 hospitals in total), compared with hospitals that did not have data for at least four process-of-care indicators and at least one outcome-of-care indicator (2,252 hospitals). Distributions of hospital composite scores were compared by the number of hospitals missing process- and outcome-of-care indicators. Second, we compared (1) distributions of process- and outcome-of-care indicators; (2) distributions of composite measures; (3) Spearman (rank) correlations; and (4) kappa statistics for hospital quartiles, with and without imputation of the national mean. Results appear in Appendix A, Tables 2l.3.1 to 2l.3.4.</p> <p>2l.3 Results:</p> <p>Please see Appendix A for complete details on results. The text summary follows below.</p> <p>Hospitals are more likely to fail to meet the required minimum number of outcome-of-care indicators (one) than to fail to meet the minimum number of process-of-care indicators (four). That is, four or more process-of-care indicators are missing for 1,783 hospitals (35.7%), while both outcome-of-care indicators are missing for 2,102 hospitals (42.1%). Note that all component measures are missing for 1,190 hospitals (23.9%).</p> <p>The distributions of component measures, shown in Table 2l.3.2 (Appendix A), are largely the same whether or not missing values are imputed. The clear exception is the distribution of process-of-care indicators AMI 7A and AMI 8A (fibrinolytic medication within 30 minutes of arrival and PCI within 90 minutes of arrival, respectively). We imputed these two measures far more often than other measures.</p> <p>Table 2l.3.3 (Appendix A) shows the distribution of composites for the 2,738 hospitals for which at least four process measures and one outcome measure are reported. Composites with no imputation simply drop missing component measures from the calculation; composites with imputation use the national mean in place of the missing measure. As the table shows, the two different procedures yield nearly identical distributions for the AMI composite.</p> <p>Table 2l.3.4 (Appendix A) shows the association between imputed and non-imputed measures by quartile. More than three-fourths of hospitals lie on the diagonal, occupying the same quartile for composite values using imputed and non-imputed component measures. The Spearman correlation coefficient and the kappa statistic both indicate a strong positive relationship between the two.</p>	
<p>2b. Reliability testing of composite score</p> <p>2b.1 Data/sample (<i>description of data/sample and size</i>):</p> <p>The reliability of the proposed AMI composite measure is informed by the reliability of the component scores on which it is based. Two reports, one by Williams et al. (2006) and the other by the Government Accountability Office (GAO) (2006), provide insight into component measure reliability:</p> <p>Williams, SC, Watt, A, Schmaltz, SP, Koss, RG, & Loeb, JM. Assessing the reliability of standardized performance indicators. <i>Int J Qual Health Care</i>. 2006 Jun;18(3):246-55. Epub 2006 Jan 23.</p> <p>□ Williams et al. (2006) examined the reliability of all seven AMI process-of-care indicators that make up the AMI composite. Their sample included 30 hospitals, representing a diverse range of geographic locations, sizes, settings (urban/rural), and ownership categories (profit/not-for-profit); 19 of these collected AMI data. A randomly selected set of deidentified, previously abstracted medical records was</p>	<p>2b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

Comment [KP12]: 2b. Reliability testing of the composite measure demonstrates the results are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period.

transmitted from the hospitals' performance measurement vendors and AMI process-of-care indicators were reabstracted following guidelines from the Specification Manual for National Implementation of Hospital Core Measures. Sample sizes used to calculate each measure generally ranged from 100-200 cases, though for AMI-4 (smoking cessation counseling) and AMI-8A (first PCI time) the sample size was fewer than 50.

United States. Government Accountability Office. Report to the Committee on Finance, U.S. Senate. Hospital Quality Data: CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data. Report No. GAO-06-54, Jan. 31, 2006

□ The 2006 GAO report summarizes CMS' process to assess the reliability of the measures currently reported on Hospital Compare and reports the results of this process for hospital discharges between January 1, 2004, and June 30, 2004. CMS' contractor, CDAC (Clinical Data Abstraction Center), assesses the reliability of the component measures on a quarterly basis. This assessment uses a sample of five randomly selected patient records from each hospital participating in the Hospital Inpatient Quality Reporting Program, which includes hospitals from all states but Maryland and Puerto Rico. (vi)

Footnotes

vi. As a result of the GAO report, in 2010 this process changed so that CDAC instead reviews 12 patient records from a randomly selected sample of 800 hospitals.

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

Williams, SC, Watt, A, Schmaltz, SP, Koss, RG, & Loeb, JM. Assessing the reliability of standardized performance indicators. Int J Qual Health Care. 2006 Jun;18(3):246-55. Epub 2006 Jan 23.

□ Reliability was assessed using percentage agreement for continuous variable elements and chance-corrected agreement using Cohen's kappa for binary data elements.

United States. Government Accountability Office. Report to the Committee on Finance, U.S. Senate. Hospital Quality Data: CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data. Report No. GAO-06-54, Jan. 31, 2006

□ For each hospital, data are deemed reliable if there is 80 percent or greater agreement between the hospital quality data previously submitted to CMS and the CDAC reabstraction results.

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

Williams, SC, Watt, A, Schmaltz, SP, Koss, RG, & Loeb, JM. Assessing the reliability of standardized performance indicators. Int J Qual Health Care. 2006 Jun;18(3):246-55. Epub 2006 Jan 23.

□ Table 2b.3.1 (Appendix A) summarizes the reliability statistics for the AMI measures that are included in the proposed composite. Using the standards proposed by Landis & Koch (1977),⁽¹⁾ the resulting kappas indicate almost perfect agreement (kappa > 0.81) for three of the measures, substantial agreement (kappa ranging from 0.61 to 0.80) for one measure, and moderate agreement (kappa ranging from 0.41 to 0.60) for two measures. Although a kappa was not calculated for AMI-8A (first PCI time), the authors report 64.7 percent agreement for this measure.

United States Government Accountability Office. Report to the Committee on Finance, U.S. Senate. Hospital Quality Data: CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data. Report No. GAO-06-54, Jan. 31, 2006

□ The GAO report, which looked at reporting from January 1, 2004, through June 30, 2004, found that 90 percent of hospitals exceeded the 80 percent reliability threshold.

Citations

1. Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics 33: 159-174

<p>2c. Validity testing of composite score</p> <p>2c.1 Data/sample <i>(description of data/sample and size):</i></p> <p>The testing of the validity of the component scores uses two sets of data. The first data set merges process- and outcome-of-care indicators for AMI with a data collection period of July 2008 to June 2009. The second data set merges process- and outcome-of-care indicators for AMI with a data collection period of July 2007 to June 2008. Composite measures are calculated from these two separate periods and compared, with the assumption that a valid composite measure should show minimal change on a year-to-year basis.</p> <p>Across these two data collection periods, 1,747 hospitals had valid composite measures for AMI.</p> <p>2c.2 Analytic Method <i>(type of validity & rationale, method for testing):</i></p> <p>Using the two sets of data, we compared composite measures across the two years using the Spearman (rank) correlation coefficient to evaluate the predictive validity of the composite measure over time.</p> <p>2c.3 Testing Results <i>(statistical results, assessment of adequacy in the context of norms for the test conducted):</i></p> <p>Please see Appendix A for complete details on results. The text summary follows below.</p> <p>The Spearman correlation between composite measures computed in 2007-2008 and 2008-2009 was 0.43 (p < 0.001), indicating moderate predictive validity of the composite. (See Appendix A, Table 2c.3.1.) A large number of hospitals (about 40 percent) lie on the diagonal, such that the same hospital quartiles for composite values were occupied during 2007-2008 and 2008-2009. In contrast, very few hospitals (about 5 percent) occupy the first quartile in 2007-2008 and the fourth quartile in 2008-2009, and vice versa. Across the two separate periods, about 40 percent of hospitals' categorizations differ by one quartile (that is, during 2008-2009, a hospital was one quartile above or below its categorization in 2007-2008). This discrepancy appears to be a result of the tight distribution of the process- and outcome-of-care indicators, as shown in Table 2l.3.2 (Appendix A).</p>	<p>2c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>2f. Identification of Meaningful Differences in Performance Across Entities</p> <p>2f.1 Data/sample from Testing or Current Use <i>(description of data/sample and size):</i></p> <p>Testing to identify meaningful differences in performance of composite scores uses data that are publicly reported on Hospital Compare by CMS. We merged process-of-care indicators for AMI with a data collection period of July 2008 to June 2009 and outcome-of-care indicators for AMI with a data collection period of July 2006 to June 2009. A total of 4,990 hospitals were reported on Hospital Compare during this period. Of these hospitals, we estimated composite measures for 2,738, for which less than or equal to three process-of-care indicators and less than or equal to one outcome-of-care indicator is missing.</p> <p>The seven hospital process-of-care indicators related to AMI that are used in the construction of composites are drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2008 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI are based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.</p> <p>2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance <i>(type of analysis & rationale):</i></p> <p>To examine meaningful differences in composite measures across hospitals, we compared hospitals' confidence interval estimates with the overall mean and assigned hospitals into one of three performance categories: better than hospitals, if the interval estimate is entirely above the mean; no different than hospitals, if the interval estimate includes the mean; and worse than hospitals, if the interval estimate is entirely below the mean. These performance categories do not reflect how the composites will ultimately be displayed on Hospital Compare.</p>	<p>2f C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

Comment [KP13]: 2c. Validity testing of the composite measure 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 [KP14]: 2f. Methods for scoring and analysis of the composite measure allow for identification of statistically significant and practically/ clinically meaningful differences in performance.

<p>2f.3 Provide Measure Scores from Testing or Current Use (<i>description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningful differences in performance</i>):</p> <p>Please see Appendix A for complete results of measure scores from testing. The text summary follows below.</p> <p>CMS has not decided how it will ultimately display hospital performance to consumers on Hospital Compare or to providers in hospital-specific reports. Table 2f.3.1 in Appendix A provides the number of hospitals in each of the three performance categories (better/no different/worse than the mean). These performance categories do not reflect how the composites will ultimately be displayed on Hospital Compare.</p>	
<p>2h. Disparities in Care</p> <p>2h.1 If measure is stratified, provide stratified results (<i>scores by stratified categories/cohorts</i>): The measure is not stratified.</p> <p>2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: No disparities have been reported/identified.</p>	<p>2h</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>If the component measures are <u>combined at the patient level</u>, complete 2d.</p>	
<p>2d. Exclusions Justified</p> <p>2d.1 Summary of Evidence supporting exclusion(s):</p> <p>2d.2 Citations for Evidence:</p> <p>2d.3 Data/sample (<i>description of data/sample and size</i>):</p> <p>2d.4 Analytic Method (<i>type analysis & rationale</i>):</p> <p>2d.5 Testing Results (<i>e.g., frequency, variability, sensitivity analyses</i>):</p>	<p>2d</p> <p>H <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>L <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>If the component measures are <u>combined at the patient level and include outcomes</u>, complete 2e.</p>	
<p>2e. Risk Adjustment</p> <p>2e.1 Data/sample (<i>description of data/sample and size</i>):</p> <p>2e.2 Analytic Method (<i>type of risk adjustment, analysis, & rationale</i>):</p> <p>2e.3 Testing Results (<i>risk model performance metrics</i>):</p> <p>2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:</p>	<p>2e</p> <p>H <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>L <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Scientific Acceptability of Measure Properties</i>?</p> <p>Steering Committee: Overall, to what extent was the criterion, <i>Scientific Acceptability of Measure Properties</i>, met? Rationale:</p>	<p>2</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>3. USABILITY</p>	
<p>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. (composite measure evaluation criteria)</p>	<p>Eval</p>
<p>3a. Meaningful, Understandable, and Useful Information</p>	<p>3a</p> <p>C <input type="checkbox"/></p>

Comment [KP15]: 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.

Comment [KP16]: 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).

Comment [KP17]: 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; OR
 rationale/data support no risk adjustment.

Comment [KP18]: 3a. Demonstration that information produced by the composite 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).

3a.1 Current Use: In use Not in use

P
M
N

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):

Following NQF endorsement, the proposed measure will undergo a national dry run in advance of implementation on Hospital Compare. The dry run is currently slated for the second quarter of 2011. The dry run will include the following steps:

- Standard Data Processing System (SDPS) memos will be sent to hospitals and QIOs with public reporting contacts announcing the dry run.
- Confidential draft hospital-specific reports (HSRs) will be made available to hospitals via the “My QualityNet” website, with supporting materials describing the methods and handling of constituent measures.
- A mock report containing simulated data but describing methods in full will also be published on QualityNet.
- A 30-day comment period will be opened in order to receive hospital feedback.
- Nationwide webinars will be held in order to review the dry run process and the methodology used to derive the composite measures, and to explain the summary data provided in the HSRs.
- A summary report of the dry run will be produced, including implications for reporting the measures.

Following this process, public reporting is expected on Hospital Compare sometime in 2011.

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):

Following NQF endorsement and a national dry run, CMS plans to report this composite publicly on Hospital Compare. CMS’ current timetable calls for this public reporting to occur in 2011. CMS’ experience indicates that hospitals closely scrutinize measures reported on Hospital Compare and consider these results as part of their quality improvement efforts.

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

3a.4 Data/sample (description of data/sample and size):

Several studies suggest that the proposed composite measure will improve consumer understanding of hospital performance for AMI patients and will be an asset to clinicians. In work that is directly relevant to the proposed measure, Borck et al. (2009) held a series of focus groups that evaluated consumer and clinician understanding of condition-specific composite measures for AMI, HF, PN, and SCIP that are very similar to the proposed measure. Their work also evaluated understanding of AHRQ and HCAHPS composite measures. In addition, work by Smith et al. (2005) examined the interpretability of Hospital Compare data, including several of the component measures in the proposed composite. A further study by Peters et al. (2007) also provides insight into consumer understanding of publicly reported hospital quality measures; L&M Policy Research LLC specifically reports on consumer understanding of the readmissions outcome measure, one of two possible outcome-of-care indicators included in this composite.

Borck, M, Thomas, C, & Gerteis, M. Transparency in Public Reporting: Consumer Testing and Enhancements to CMS’s Compare Tools: Topline Summary of Findings from Round #1 Interviews with Consumers, April 9, 2009, and Topline Summary of Findings from Round #2 Interviews with Consumers and Physicians, Composite measures of quality for Hospital Compare, June 11, 2009. Memoranda to the Centers for Medicare & Medicaid Services.

□ Round 1 - Borck et al. (2009) used a convenience sample of 21 consumers in the Baltimore, Maryland, area. Participants ranged in age from 45 to 70; 67 percent were women, and 48 percent were Medicare beneficiaries.

□ Round 2 - Borck et al. (2009) used a convenience sample of 18 consumers and five physicians from the

Miami, Florida, area. The group ranged in age from 45 to 70; most of the group's members were men and Medicare beneficiaries.

Smith, F, Gerteis, M, Burnes, A, Gerteis, J, Crelia, S, & Silva, N. Usability Testing of the "Hospital Compare" Website. Final Report to the Centers for Medicare & Medicaid Services. August 29, 2005.

▫Smith et al. (2005) used a sample of 51 consumers and 40 health care providers to assess their ability to understand Hospital Compare content and navigate the user interface website. Among the consumers, 47 of 51 (92%) were older than 65, and of the over-65 group, 53 percent were Medicare beneficiaries at risk for heart disease. Among the health care providers, 30 percent were nurses, 38 percent were primary care physicians, and the remainder were cardiologists and pulmonologists.

Peters, E, Dieckmann, N, Dixon, A, Hibbard, JH, & Mertz, CK. Less is more in presenting quality information to consumers. *Med Care Res Rev.* 2007 Apr;64(2):169-90.

▫Peters et al. (2007) employed a convenience sample of employment-age adults (ages 18 to 64, mean age of 37, 48 percent female, and 76 percent white) to determine whether providing only the most important quality information increases comprehension and information use. Half of the sample had lower levels of education (high school or less), 45 percent had health insurance, and 74 percent had an annual household income of less than \$20,000.

L&M Policy Research LLC. Report to the Centers for Medicare & Medicaid Services: Recommendations for Incorporating Hospital Readmission Data into the Hospital Compare Website. January 29, 2009.

▫This effort entailed two rounds of consumer testing, the first of which focused on general understanding of hospital readmission measures and how they are calculated, as well as the fact that the measures are for readmission within 30 days and calculated from Medicare fee-for-service data. The sample for this round included 10 adult consumers ages 50 to 70, most of whom were previously diagnosed with heart disease; 8 caregivers ages 40 to 60; and 6 physicians who were primary care physicians, cardiologists, and pulmonologists.

3a.5 Methods (methods, e.g., focus group, survey, QI project):

Borck, M, Thomas, C, & Gerteis, M. Transparency in Public Reporting: Consumer Testing and Enhancements to CMS's Compare Tools: Topline Summary of Findings from Round #1 Interviews with Consumers, April 9, 2009, and Topline Summary of Findings from Round #2 Interviews with Consumers and Physicians, Composite measures of quality for Hospital Compare, June 11, 2009. Memoranda to the Centers for Medicare & Medicaid Services.

▫Borck et al. (2009) used a mock Hospital Compare website that presented the composite quality measures of interest. Using a standard interview protocol, in-depth, one-on-one discussions assessed comprehension of composite measures, organization and presentation of the site, and composite labels and descriptions.

Smith, F, Gerteis, M, Burnes, A, Gerteis, J, Crelia, S, & Silva, N. Usability Testing of the "Hospital Compare" Website. Final Report to Centers for Medicare & Medicaid Services. August 29, 2005.

▫Smith et al. (2005) tested consumers' and health providers' ability to understand and use the Hospital Compare website using both in-depth, one-on-one interviews and dyads (interviews that involve two respondents and one interviewer). Using a Hospital Compare website prototype, participants first navigated the website independently and then responded to a series of open-ended questions using an approved protocol during an approximately two-hour period.

Peters, E, Dieckmann, N, Dixon, A, Hibbard, JH, & Mertz, CK. Less is more in presenting quality information to consumers. *Med Care Res Rev.* 2007 Apr;64(2):169-90.

▫Peters et al. (2007) assigned participants to one of three groups, each of which was presented with hospital quality data in a different format. In the first group, data on cost, quality, and nonquality information was unordered. In the second, cost and quality data was highlighted and presented first and nonquality information was presented last and not emphasized. In the final group, only cost and quality information was shown, and quality information was highlighted. Within each of these groups, respondents were then shown information about three hospitals and asked to choose a hospital and answer a series of questions.

L&M Policy Research LLC. Report to the Centers for Medicare & Medicaid Services: Recommendations for Incorporating Hospital Readmission Data into the Hospital Compare Website. January 29, 2009.
 □ Participants were shown paper-based mock-ups of hospital quality data and asked to compare hospitals and select a hospital for themselves and their family members.

3a.6 Results (qualitative and/or quantitative results and conclusions):

Borck, M, Thomas, C, & Gerteis, M. Transparency in Public Reporting: Consumer Testing and Enhancements to CMS’s Compare Tools: Topline Summary of Findings from Round #1 Interviews with Consumers, April 9, 2009, and Topline Summary of Findings from Round #2 Interviews with Consumers and Physicians, Composite measures of quality for Hospital Compare, June 11, 2009. Memoranda to the Centers for Medicare & Medicaid Services.

□ This work yielded several important results that are directly relevant to the proposed condition-specific composite measure. Most significantly, all respondents from Round 1 correctly interpreted the star ratings for the condition-specific composites (AMI, HF, PN, and SCIP) and the HCAHPS composite measure. Round 1 also revealed that almost all participants preferred more descriptive definitions of the composites; specifically, those included a list of all the component measures making up the composite. Similarly, in Round 2 respondents were also able to interpret the star ratings for condition-specific quality ratings of composites and the HCAHPS composite correctly. However, some respondents in Round 2 did not understand that the condition-specific composite ratings included all of the individual component measures. These results indicate that the proposed condition-specific composite, which is very similar to the condition-specific measures evaluated by Borck et al. (2009), should also be easy for consumers to use. Moreover, any composite definition posted on Hospital Compare should include a list of all component measures.

Smith, F, Gerteis, M, Burnes, A, Gerteis, J, Crelia, S, & Silva, N. Usability Testing of the “Hospital Compare” Website. Final Report to Centers for Medicare & Medicaid Services. August 29, 2005.

□ This early analysis of Hospital Compare’s usability revealed that the amount of information available on the website tended to overwhelm consumers and that detailed information about interpretation added to this sense of overload. The provider participants concurred with this sentiment. Although these results certainly suggest certain challenges in making hospital quality data user friendly, the proposed composite measure aims to address this issue by creating a single benchmark that enables consumers to evaluate the quality of care at a given hospital for a given condition.

Peters, E, Dieckmann, N, Dixon, A, Hibbard, JH, & Mertz, CK. Less is more in presenting quality information to consumers. Med Care Res Rev. 2007 Apr;64(2):169-90.

□ Similar to Smith et al. (2005), Peters et al. (2007) determined that less is more with regard to consumer understanding of hospital quality data. They found that consumer comprehension was highest when only the most relevant quality information was shown and highlighted relevant to the other information. Specifically, 62 percent of respondents chose the highest quality hospital Y when only the quality information was shown; in the other two formats it was selected by 48 percent (ordered group) and 40 percent (unordered group). Such results reinforce the idea that a composite measure can enhance the utility of hospital quality data for consumers.

L&M Policy Research LLC. Report to the Centers for Medicare & Medicaid Services: Recommendations for Incorporating Hospital Readmission Data into the Hospital Compare Website. January 29, 2009.

□ This work suggests that a readmission measure is open to misinterpretation by consumers. For example, many participants in this study thought that readmission was a positive outcome because it meant that the hospital was providing follow-up care. In the proposed composite measure, discharges not followed by readmission improve the composite score. Although it is important to describe how the composite is created, this example highlights the need to define the composite in a simple, direct manner.

3b/3c. Relation to other NQF-endorsed measures
Identify similar or related [NQF-endorsed measures](#) to components and/or composite

3b.1 NQF # and Title of similar or related measures:

All components of this composite measure are NQF-endorsed. However there are currently no NQF-endorsed composite measures that provide a single indication of a hospital’s quality of care for AMI patients. In that they also serve to provide a single, consumer-friendly indication of a hospital’s quality of care as it relates to either patient safety or mortality for selected conditions, the proposed measure is similar in intent to the

<p>following:</p> <p>NQF #0531. Patient Safety for Selected Indicators (endorsed June 19, 2009/AHRQ) NQF #0530. Mortality for Selected Conditions (endorsed June 19, 2009/AHRQ)</p> <p>However, the proposed measure is condition-specific and intended to summarize the measures on Hospital Compare; thus, it provides unique and additive value above and beyond these measures.</p> <p>(for NQF staff use) Notes on similar/related <u>endorsed</u> or submitted measures:</p>	
<p>3b. Harmonization</p> <p>3b.2 Are the component measure specifications harmonized, or if not, why? The component measures are harmonized within each distinct domain of the composite (that is, processes of care and outcomes of care). Within the process domain, all component measures are reported as percentages; in the outcomes domain, both component measures are reported as rates.</p>	<p>3b</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>3c. Distinctive or Additive Value</p> <p>3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures: The proposed composite measure offers a condition-specific summary of the inpatient quality measures that CMS has adopted for its Hospital Inpatient Quality Reporting Program, related to the quality of care for AMI patients.</p> <p>5.1 Competing Measures 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: There are no currently endorsed composite measures on this topic or population.</p>	<p>3c</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>3d. Decomposition of Composite</p> <p>3d.1 Describe the information that is available from decomposing the composite into its components: The component measures include the following information:</p> <ol style="list-style-type: none"> Percent of AMI Patients Given Aspirin on Arrival Percent of AMI Patients Given Aspirin at Discharge Percent of AMI Patients Given ACE Inhibitor or ARB for LVSD Percent of AMI Patients Given Smoking Cessation Advice/Counseling Percent of AMI Patients Given Beta Blocker at Discharge Percent of AMI Patients Given Fibrinolytic Medication within 30 Minutes of Arrival Percent of AMI Patients Given PCI within 90 Minutes of Arrival Acute Myocardial Infarction (AMI) 30-day Mortality Acute Myocardial Infarction (AMI) 30-day Readmission 	<p>3d</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>3e. Achieved stated purpose</p> <p>3e.1 Describe how the scores from testing or use reported in 2f demonstrate that the composite achieves the stated purpose: The scores demonstrate a range of performance on the AMI process and outcome quality measures. Testing of composite scores identified hospitals that perform significantly above and below the national mean of these scores. The scores thus reflect the underlying hospital performance regarding the quality measures for AMI, achieving the purpose of the composite.</p>	<p>3e</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Usability</i>?</p>	<p>3</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Usability</i>, met? Rationale:</p>	<p>3</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>4. FEASIBILITY</p>	

Comment [KP19]: 3b. The component measure specifications are harmonized.

Comment [KP20]: 3c. Review of existing endorsed measures and measure sets demonstrates that the composite 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).

Comment [k21]: 5. Demonstration that the measure is superior to competing measures - new submissions and/or endorsed measures (e.g., is a more valid or efficient way to measure).

Comment [KP22]: 3d. Data detail is maintained such that the composite measure can be decomposed into its components to facilitate transparency and understanding.

Comment [KP23]: 3e. Demonstration (through pilot testing or operational data) that the composite measure achieves the stated purpose/objective.

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (composite measure evaluation criteria)	Eval
<p>4a. Data Generated as a Byproduct of Care Processes</p> <p>4a.1 How are <u>all</u> the data elements that are needed to compute measure scores generated? (<i>Check all that apply</i>)</p> <p><input type="checkbox"/> Data are generated as a byproduct of care processes <u>during</u> care delivery (<i>Data are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition</i>)</p> <p><input checked="" type="checkbox"/> Coding/abstraction performed by someone other than person obtaining original information (<i>e.g., DRG, ICD-9 codes on claims; chart abstraction for quality measure, registry</i>)</p> <p><input type="checkbox"/> Survey</p> <p><input type="checkbox"/> Other (<i>e.g., patient experience of care surveys, provider surveys, observation</i>), Please describe:</p>	<p>4a</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>4b. Electronic Sources</p> <p>4b.1 Are <u>all</u> the data elements available electronically? (<i>elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims</i>)</p> <p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>4b.2 If no, specify the near-term path to achieve electronic capture by most providers.</p> <p>N/A</p> <p><i>Note: Measure stewards will be asked to specify the data elements for electronic health records at a later date</i></p>	<p>4b</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences</p> <p>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.</p> <p>Our measures are not susceptible to inaccuracies, errors, or unintended consequences; the component outcomes are well-specified in hospital administrative data.</p>	<p>4d</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>4e. Data Collection Strategy/Implementation</p> <p>4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the composite/component measures regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:</p> <p>Outcome component measures are derived from Medicare hospital claims, which are believed to be complete. All process component measures are reported as part of the Hospital Inpatient Quality Reporting Program in order for hospitals to receive the full annual Medicare payment update. Hospitals therefore have a strong financial incentive to provide process-of-care indicators. Continued availability of component measures for the AMI composite is therefore assured.</p> <p>4e.2 Costs to implement the measure (<i>costs of data collection, fees associated with proprietary measures</i>): The composite measure is calculated from process- and outcome-of-care indicators that are already publicly reported by hospitals. Hospitals and providers should not experience any additional costs or burden from the calculation of this measure.</p> <p>4e.3 Evidence for costs: N/A</p> <p>4e.4 Business case documentation: N/A</p>	<p>4e</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>If the component measures are <u>combined at the patient level</u>, complete 4c.</p> <p>4c. Exclusions</p> <p>4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications? <input type="checkbox"/> No <input type="checkbox"/> Yes ► If yes, provide justification</p>	<p>4c</p> <p>H <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>L <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Feasibility</i>?</p>	<p>4</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Feasibility</i>, met? Rationale:</p>	<p>4</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p>

Comment [KP24]: 4a. For clinical composite measures, overall the required data elements are routinely generated concurrent with and as a byproduct of care processes during care delivery.

Comment [KP25]: 4b. The required data elements for the composite overall are available in electronic sources.

Comment [KP26]: 4d. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.

Comment [KP27]: 4e. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, etc.) for obtaining all component measures can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).

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.

NQF Review #:

RECOMMENDATION	N <input type="checkbox"/>
Steering Committee: Do you recommend for endorsement? Comments:	Y <input type="checkbox"/> N <input type="checkbox"/> A <input type="checkbox"/>
CONTACT INFORMATION	
Co.1 Measure Steward (Intellectual Property Owner) Organization: Centers for Medicare & Medicaid Services Street Address: 7500 Security Boulevard, Mail Stop S3-02-01 City: Baltimore State: MD ZIP: 21244	
Co.2 Point of Contact: First Name: Shaheen Last Name: Halim Credentials (MD, MPH, etc.): Ph.D., CPC-A Email: Shaheen.Halim@cms.hhs.gov Telephone: (410) 786-0641 ext:	
Co.3 Measure Developer <i>If different from Measure Steward</i> Organization: Mathematica Policy Research Street Address: 955 Massachusetts Avenue, Suite 801 City: Cambridge State: MA ZIP: 02139	
Co.4 Point of Contact: First Name: Marian Last Name: Wrobel Credentials (MD, MPH, etc.): Ph.D. Email: MWrobel@mathematica-mpr.com Telephone: 617-301-8971 ext:	
Co.5 Submitter Organization: <input type="checkbox"/> Measure Steward <input type="checkbox"/> Measure Developer First Name: Last Name: Credentials (MD, MPH, etc.): Email: Telephone: ext:	
Co.6 List any additional organizations that sponsored/participated in measure development:	
ADDITIONAL INFORMATION	
Ad.1 Workgroup/Expert Panel involved in measure development <i>Provide a list of workgroup/panel member names and organizations. Describe the group's role in measure development.</i> <p>On October 20, 2009, CMS convened an Advisory Panel on Medicare Education (APME) that included healthcare professionals involved with communication of quality information to consumers. CMS provided this panel with an overview of plans to include new composite measures on the Hospital Compare website, and solicited feedback from the group. In general, the group was supportive of CMS' plans to pursue composites and encouraged further development in this area.</p> <p>APME Panel Members Gwendolyn T. Bronson, SHINE/SHIP Counselor, Massachusetts SHINE Program Yanira Cruz, Ph.D., President and Chief Executive Officer, National Hispanic Council on Aging Nan-Kirsten Forté, Executive Vice President, Consumer Services, WebMD Cathy C. Graeff, R.Ph., M.B.A., Partner, Sonora Advisory Group Carmen R. Green, M.D., Professor, Anesthesiology and Associate Professor, Health, Management, and Policy, University of Michigan Jessie C. Gruman, Ph.D., President, Center for Advancing Health Cindy Hounsell, J.D., President, Women's Institute for a Secure Retirement Gail Hunt, President and Chief Executive Officer, National Alliance for Caregiving Deeanna Jang, Policy Director, Asian and Pacific Islander American Health Forum Andrew Kramer, M.D., Professor of Medicine, Division of Health Care Policy and Research, University of Colorado, Denver Sandy Markwood, Chief Executive Officer, National Association of Area Agencies on Aging David W. Roberts, M.P.A., Vice President, Government Relations, Healthcare Information and Management System Society Julie Bodén Schmidt, M.S., Associate Vice President, Training and Technical Assistance, National Association of Community Health Centers Rebecca P. Snead, Chief Executive Officer and Executive Vice President, National Alliance of State Pharmacy </p>	

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

<p>Associations and APME Chair</p> <p>In 2006, CMS partnered with the Hospital Quality Alliance (HQA) in order to explore and assess strategies for improving the consumer friendliness of the Hospital Compare website. Staff representing the HQA principal organizations, which include the American Hospital Association, the Federation of American Hospitals, and the Association of American Medical Colleges, convened a working group charged with determining how to make Hospital Compare more consumer friendly over the short and long term. One of the key long-term recommendations from this group was to direct CMS/HQA to create condition- or procedure-specific composites related to current measures on Hospital Compare. Indeed, the group noted that such summary measures may help condense a large volume of information into a smaller, more manageable amount that is easier for decision-making.</p>
<p>Ad.2 If adapted, name of original measure: N/A</p> <p>Ad.3 If adapted, original specifications <input type="checkbox"/> attachment or Ad.4 web page URL:</p>
<p>Measure Developer/Steward Updates and Ongoing Maintenance</p> <p>Ad.6 Year the measure was first released: N/A</p> <p>Ad.7 Month and Year of most recent revision: N/A</p> <p>Ad.8 What is the frequency for review/update of this measure? Annually</p> <p>Ad.9 When is the next scheduled review/update for this measure? 2012</p>
<p>Ad.10 Copyright statement/disclaimers:</p>
<p>Ad.11 Additional Information <input type="checkbox"/> attachment or web page URL:</p>
<p>I have checked that the submission is complete and all the information needed to evaluate the measure is provided in the form; any blank fields indicate that no information is provided. <input checked="" type="checkbox"/></p>
<p>Date of Submission (MM/DD/YY): 11/3/10</p>

The National Quality Forum
Composite Measure of Hospital Quality for AMI
Appendix A

Please note: This document contains results and equations intended to accompany the NQF submission form for the proposed AMI composite measure. For ease of review, associated text descriptions and other relevant information are provided both on the NQF form and in this document.

1. Importance to Measure and Report

1d. Purpose/objective of the Composite

1d.2 Describe the quality construct used in developing the composite:

The composite measure of quality of hospital care for AMI aims to be a comprehensive indicator of hospital performance that will be of special value to consumers as a summary means of evaluating alternative hospitals. The quality construct is thus formative rather than reflective in nature. At present, CMS publishes seven individual process-of-care indicators and two outcome-of-care indicators meant to capture the quality of hospital care provided to patients with AMI. The proposed composite combines these in the form of process- and outcome-of-care domains.

CMS developed the composite measure to achieve the following goals for reporting hospital quality measures composite methodology:

- Summarize measures on Hospital Compare in a single, useful, condition-specific composite
- Produce composite values that show differences in hospital performance that are clinically and statistically meaningful and reflect true underlying differences in quality
- Enable the calculation of results for most hospitals
- Employ a method that accommodates changes in the set of measures on Hospital Compare and can be used for multiple conditions
- Employ a method that is relatively simple, so hospitals can duplicate results

These goals can be achieved by a method that is consistent with that of other widely used composites; in this case the method used for the Agency for Healthcare Research and Quality (AHRQ) composites. The National Quality Forum (NQF) has endorsed those composites and CMS, states, and other organizations use them widely.

The current Hospital Inpatient Quality Reporting Program construct domains focus on diseases important to the Medicare population: AMI, Heart Failure (HF), and Pneumonia (PN), and on quality indicators related to the Surgical Care Improvement Project (SCIP). The first three have separate sub-composites in processes- and outcomes-of-care. This system of domains and sub-composites allows addition or removal of measures without changes in methodology or weighting, as well as the publication or analysis of separate process and outcome composites within a condition if desired.

In the development of this composite, certain methodological decisions were made to satisfy the policy goals outlined above. First, we entered individual measures as values, rather than ranks, to reduce the likelihood that very small differences in absolute performance lead to large differences in ranking composite scores. Second, we imputed values for missing indicators so that the composite would define as many hospitals as possible. Third, we adjusted individual measures for reliability, a process that leads to a more accurate measure of true underlying performance and avoids extreme values for small hospitals due to random variation. Lastly, we used denominator weighting so that the composite places more weight on measures that are

Production and Implementation of the CMS Hospital Outcomes and Efficiency Measures
 Appendix A: Composite Measure of Hospital Quality for AMI
 11.3.10

reported for relatively more patients nationally. In Table 1d.2.1, we present the mapping between CMS’ policy goals and methodological decisions in tabular form.

Table 1d.2.1. CMS Policy Goals for Composite Measures and Associated Methodological Decisions

Policy Goals	Methodological Decisions
Summarize measures on Hospital Compare in a single, useful, condition-specific composite	<ul style="list-style-type: none"> • Include the same set of process and outcome measures as Hospital Compare • Use same data periods as Hospital Compare
Produce differences in composite values that are clinically and statistically meaningful and reflect true differences in underlying quality	<ul style="list-style-type: none"> • Enter component measures as values, not ranks, so that slight differences in measured performance do not potentially lead to large differences in the composite value for topped-off measures • For process measures, adjust component measures for reliability so that random variation does not drive small hospitals to extremes
Results available for a large number of hospitals	<ul style="list-style-type: none"> • Define composites when at least half of both process and outcome variables are available • No minimum sample applied • Note: Process measures are available when the number of eligible discharges is one or more; outcome variables are available when the number of eligible discharges is 25 or more
Results do not differ dramatically depending on whether a hospital has one eligible discharge or zero eligible discharges	<ul style="list-style-type: none"> • Adjust process measures for reliability • Use the national mean in both the reliability adjustment and in cases in which a component measure is missing
Treat all hospitals equally regardless of hospital characteristics	<ul style="list-style-type: none"> • Use the national mean (not a mean based on performance of similar hospitals) in imputation
Focus more on measures relevant to more patients	<ul style="list-style-type: none"> • Construct process and outcome composites using weights based on national denominators
Method is scientifically acceptable and acceptable to stakeholders	<ul style="list-style-type: none"> • Adopt an approach that is similar to that used for AHRQ quality indicators (QIs) • Note: AHRQ QIs are NQF-endorsed and widely reported
Method accommodates changes in the set of measures on Hospital Compare Method can be used for multiple conditions Relative weighting of process and outcome domains does change when measures are added to or deleted from one domain	<ul style="list-style-type: none"> • Method is based on general principles, not on the specific statistical performance of a group of measures • Process and outcome domains are statistically standardized before they are added together

<p>Method is relatively simple</p> <p>Hospitals can duplicate results</p>	<ul style="list-style-type: none"> • Use equal weighting to combine process and outcome domains • Use the national mean (not a hospital-specific predicted value) in imputation • Reliability weights are a function of a hospital’s number of cases and national parameters
---	---

2. Scientific Acceptability of Measure Properties

2a. Composite Measure Specifications

2a.46 If differential weighting, describe:

Consistent with the approach used for the AHRQ measures, CMS used denominator weighting in constructing the process-of-care sub-composite. Denominator weighting places relatively more weight on measures that apply to relatively more patients nationally. Specifically, the process-of-care sub-composite score is estimated as

$$P_j^* = \sum_{k=1}^K \left(\frac{\sum_{j=1}^J n_{jk}}{\sum_{k=1}^K \sum_{j=1}^J n_{jk}} \right) P_{jk}^* \tag{eq. 2a.46.1}$$

where n_{jk} is the number of hospital cases for AMI process-of-care indicator $k=1, \dots, K$, in hospital j and P_{jk}^* is the reliability-adjusted process-of-care indicator k , in hospital $j=1, \dots, J$.¹

Similarly, the outcome-of-care sub-composite score is estimated used denominator weighting. That is

$$O_j^* = \sum_{l=1}^L \left(\frac{\sum_{j=1}^J n_{jl}}{\sum_{l=1}^L \sum_{j=1}^J n_{jl}} \right) O_{jl}^* \tag{eq. 2a.46.2}$$

where n_{jl} is the number of hospital cases for AMI outcome-of-care indicator $l=1, \dots, L$, in hospital $j=1, \dots, J$ and O_{jl}^* is the risk-standardized outcome-of-care score.

The overall composite score (C_j^*) is estimated using two steps. First, a standardized composite measure (\widetilde{C}_j^*) is estimated as a simple average of standardized process- and outcome-of-care sub-composites

¹ For more information about the reliability-adjusted process-of-care indicator, please see the discussion “Estimation of the Reliability-Weight-Adjusted Measure,” under Section 2a.21: Calculation Algorithm.

$$\widetilde{C}_j^* = \frac{1}{2} \left(\frac{P_j^* - \widehat{\mu}_P}{\widehat{\sigma}_P} \right) + \frac{1}{2} \left(\frac{O_j^* - \widehat{\mu}_O}{\widehat{\sigma}_O} \right) \quad (\text{eq. 2a.46.3})$$

where $\widehat{\mu}_P$ is the average of the process-of-care sub-composite, $\widehat{\sigma}_P$ is the standard deviation of the process-of-care sub-composite, $\widehat{\mu}_O$ is the average of the outcome-of-care sub-composite, and $\widehat{\sigma}_O$ is the standard deviation of the outcome-of-care sub-composite.

Then, to map the standardized composite score to a scale between zero and 100, we add the lowest possible score a hospital can receive (i.e., a hospital scores zero percent on all process- and outcome-of-care indicators) and divide by the range of potential hospital scores (i.e., the difference between the highest possible score a hospital can receive, which is if a hospital scores 100 percent on all process- and outcome-of-care indicators, and the lowest possible score). Specifically, the overall composite score (C_j^*) is

$$C_j^* = \frac{\widetilde{C}_j^* + \frac{1}{2} \left(\frac{\widehat{\mu}_P}{\widehat{\sigma}_P} + \frac{\widehat{\mu}_O}{\widehat{\sigma}_O} \right)}{\frac{1}{2} \left(\frac{1}{\widehat{\sigma}_P} + \frac{1}{\widehat{\sigma}_O} \right)} \quad (\text{eq. 2a.46.4})$$

2a.21 Calculation Algorithm (Describe the calculation of the measure as a flowchart or series of steps):

STEP 1

Hospital process-of-care indicators for AMI, with a data collection period of July 2008 to June 2009, and outcome-of-care indicators for AMI, with a data collection period of July 2006 to June 2009, that are publicly reported on Hospital Compare (<http://www.hospitalcompare.hhs.gov/>), are combined into a single data set using hospital provider identification numbers.

STEP 2

Process-of-care indicators are reliability-weight adjusted. That is, the value of each process-of-care indicator is set equal to the weighted average of the hospital’s own mean for the indicator and the national mean for the indicator. The weights are based on the between-hospital variance and the within-hospital variance in indicator scores (for more information on this adjustment, see the “Estimation of the Reliability-Weight-Adjusted Measures,” which follows).ⁱⁱ

STEP 3

ⁱⁱ Hospital outcome-of-care indicators are not reliability-weight adjusted because they have been risk-standardized using a method that accounts for reliability previously, before public reporting on Hospital Compare.

Hospitals missing process- or outcome-of-care indicators are imputed with the national mean.ⁱⁱⁱ The national mean of the process-of-care indicators are estimated as a simple average of the indicators. The national mean of the outcome-of-care indicators are provided by Hospital Compare.

STEP 4

The process-of-care sub-composite score is computed using denominator weights, where the denominator weight is based on the number of hospital cases for each process-of-care indicator (see the “Estimation of the Absolute Score Index with Reliability Weighting Composite Measure,” eq. 2a.21.5, which follows).

STEP 5

The outcome-of-care sub-composite score is also computed using denominator weights (see the “Estimation of the Absolute Score Index with Reliability Weighting Composite Measure,” Equation 2a.21.6, which follows).

STEP 6

To standardize the process- and outcome-of-care sub-composite measures, each are scaled by subtracting the overall sub-composite mean and dividing by the standard deviation. Then the average of the process- and outcome-of-care sub-composites is estimated (see the “Estimation of the Absolute Score Index with Reliability Weighting Composite Measure,” Equation 2a.21.8, which follows).

STEP 7

Lastly, in order to have a composite measure with values between zero and 100, we add the lowest possible score a hospital can receive (i.e., a hospital scores zero percent on all process- and outcome-of-care indicators) and divide by the range of potential hospital scores (i.e., the difference between the highest possible score a hospital can receive, which is if a hospital scores 100 percent on all process- and outcome-of-care indicators, and the lowest possible score) (see the “Estimation of the Absolute Score Index with Reliability Weighting Composite Measure,” Equation 2a.21.7, which follows).

ⁱⁱⁱ The use of the national mean is consistent with the approach used for the AHRQ quality composites. It is simple, already in use, and perceived as fair by providers.

ESTIMATION OF RELIABILITY-WEIGHT-ADJUSTED MEASURES

For each process-of-care indicator, the reliability-weight-adjusted indicator is equal to a weighted average of the hospital’s own measure and the national mean value of the measure. In each case, the weight is a measure of the precision with which a hospital’s measure has been estimated. This weighted average has been shown to be more accurate, on average, than using each hospital’s individual value for the measure.

The weight is made up of two parts—the variability of the measure within each hospital, termed the “within variance” or “noise variance,” and the variability across hospitals, known as the “signal variance.” The weight attached to each hospital’s own value for process measure k is equal to the ratio of the signal variance to the sum of the signal variance and the noise variance. As the number of observations for a hospital (n_{jk}) increases, the weight approaches one.

First, let:

σ_{sk}^2	Signal variance
σ_{wjk}^2	Within variance
P_{jk}	Hospital-specific rate for process-of-care indicator k
P_k^n	National rate for process-of-care indicator k
n_{jk}	Total number of cases in hospital j for indicator k
N_k	Total number of hospitals for indicator k
$k = 1, \dots, K$	Process-of-care indicator
$j = 1, \dots, J$	Hospital index

Then the reliability-weight adjusted estimator (P_{jk}^*) is

$$P_{jk}^* = W_{jk}P_{jk} + (1 - W_{jk})P_k^n \tag{eq. 2a.21.1}$$

where W_{jk} is the reliability-weight:

$$W_{jk} = \frac{\sigma_{sk}^2}{\sigma_{sk}^2 + \sigma_{wjk}^2} \tag{eq. 2a.21.2}$$

σ_{sk}^2 is the signal variance:

$$\sigma_{sk}^2 = \frac{\sum_{i=1}^J (P_{ik} - P_k^n)^2}{N_k} - \frac{\sum_{i=1}^J P_{ik}(1 - P_{ik})}{\sum_{i=1}^J n_{ik}} \quad (\text{eq. 2a.21.3})$$

and σ_{wjk}^2 is the within variance:

$$\sigma_{wjk}^2 = \frac{\sum_{i=1}^J P_{ik}(1 - P_{ik}) \frac{n_{ik}}{\sum_{l=1}^J n_{lk}}}{n_{jk}} \quad (\text{eq. 2a.21.4})$$

ESTIMATION OF THE ABSOLUTE SCORE INDEX WITH RELIABILITY WEIGHTING (ASI-RW) COMPOSITE MEASURE

We estimate the composite measure using an approach that we have termed absolute score index with reliability weighting (ASI-RW). To compute the ASI-RW, we first computed process- and outcome-of-care sub-composite scores. Using process-of-care indicators that are set equal to the weighted average of the hospital's own mean for the indicator and the national mean for the indicator (that is, reliability-weight adjusted), the process-of-care sub-composite score is computed as a denominator-weighted average of the process-of-care indicators. That is, weights of each process-of-care indicator are based on the opportunities for providing a specific recommended treatment and greater weights are placed on measures that apply to relatively more patients nationally. Similarly, the outcome-of-care sub-composite score is also estimated as a denominator-weighted average of the outcome-of-care indicators, which are reported on Hospital Compare and are risk-adjusted.

To standardize each measure, the process- and outcome-of-care sub-composite scores are scaled by subtracting the overall sub-composite mean and dividing by the standard deviation. The ASI-RW composite measure is computed using two steps. First, the average of the process- and outcome-of-care sub-composites is estimated. Then, to map the standardized composite score to a scale between zero and 100, we add the lowest possible score a hospital can receive (i.e., a hospital scores zero percent on all process- and outcome-of-care indicators) and divide by the range of potential hospital scores (i.e., the difference between the highest possible score a hospital can receive, which is if a hospital scores 100 percent on all process- and outcome-of-care indicators, and the lowest possible score).

The estimation of the ASI-RW can be described in greater detail as follows. First, let:

P_{jk}^* Reliability-weight-adjusted rate for process-of-care indicator k , in hospital j

O_{jl}^*	Risk-standardized rate for outcome-of-care indicator l , in hospital j
$k = 1, \dots, K$	Process-of-care indicator
$l = 1, \dots, L$	Outcome-of-care indicator
$j = 1, \dots, J$	Hospital index

The process-of-care sub-composite score (P_j^*) is

$$P_j^* = \sum_{k=1}^K \left(\frac{\sum_{j=1}^J n_{jk}}{\sum_{k=1}^K \sum_{j=1}^J n_{jk}} \right) P_{jk}^* \quad (\text{eq. 2a.21.5})$$

The outcome-of-care sub-composite score (O_j^*) is

$$O_j^* = \sum_{l=1}^L \left(\frac{\sum_{j=1}^J n_{jl}}{\sum_{l=1}^L \sum_{j=1}^J n_{jl}} \right) O_{jl}^* \quad (\text{eq. 2a.21.6})$$

Lastly, the composite score (C_j^*) is

$$C_j^* = \frac{\widetilde{C}_j^* + \frac{1}{2} \left(\frac{\widehat{\mu}_P}{\widehat{\sigma}_P} + \frac{\widehat{\mu}_O}{\widehat{\sigma}_O} \right)}{\frac{1}{2} \left(\frac{1}{\widehat{\sigma}_P} + \frac{1}{\widehat{\sigma}_O} \right)} \quad (\text{eq. 2a.21.7})$$

where,

$$\widetilde{C}_j^* = \frac{1}{2} \left(\frac{P_j^* - \widehat{\mu}_P}{\widehat{\sigma}_P} \right) + \frac{1}{2} \left(\frac{O_j^* - \widehat{\mu}_O}{\widehat{\sigma}_O} \right) \quad (\text{eq. 2a.21.8})$$

and μ_P is the average of the process-of-care sub-composite, σ_P is the standard deviation of the process-of-care sub-composite, μ_O is the average of the outcome-of-care sub-composite, and σ_O is the standard deviation of the outcome-of-care sub-composite.

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

To examine meaningful differences in composite measures among hospitals, for the purpose of internal analysis, we compared hospitals' confidence interval estimates with the overall mean and assigned hospitals into one of three performance categories: better than hospitals, if the

interval estimate is entirely above the mean; no different than hospitals, if the interval estimate includes the mean; and worse than hospitals, if the interval estimate is entirely below the mean. These categories were used for illustrative analyses only and should not be assumed to be the manner in which these composites will be publicly reported.

The hospital-specific standard error is estimated by computing the variance of the composite measure and computing a square root of the variance. After we derive the standard errors for each hospital, we estimate an interval estimate around each hospital’s mean composite measure. The interval estimate is a range of probable values for the composite measure that characterizes the amount of uncertainty associated with the estimate. We apply a 95 percent interval estimate, which indicates a 95 percent confidence level that the true composite measure is between the lower and upper limits of the interval.

More specifically, the standard error for a specific hospital is calculated as follows. First, we let:

P_{jk}^*	Hospital-specific reliability-weight-adjusted rate for process-of-care indicator k
O_{jl}^*	Risk-standardized hospital-specific rate for process-of-care indicator l
n_{jk}	Total number of cases in hospital j for indicator k
N_k	Total number of hospitals for indicator k
μ_P	Mean of process domain composite
μ_O	Mean of outcome domain composite
σ_P	Standard deviation of process domain composite
σ_O	Standard deviation of outcome domain composite
$k = 1, \dots, K$	Process-of-care indicator
$l = 1, \dots, L$	Outcome-of-care indicator
$j = 1, \dots, J$	Hospital index

The hospital’s process-of-care domain composite score (P_j^*) is estimated as a denominator weighted average of the reliability-weight-adjusted process-of-care indicator rates.

$$P_j^* = \sum_{k=1}^K \left(\frac{\sum_{j=1}^J n_{jk}}{\sum_{k=1}^K \sum_{j=1}^J n_{jk}} \right) P_{jk}^* \tag{eq. 2a.22.1}$$

The hospital’s outcome-of-care domain composite score (O_j^*) is estimated as an equally weighted average of the outcome-of care indicators.

$$O_j^* = \sum_{l=1}^L \left(\frac{\sum_{j=1}^J n_{jl}}{\sum_{l=1}^L \sum_{j=1}^J n_{jl}} \right) O_{jl}^*$$

(eq. 2a.22.2)

The composite measure (C_j) is a simple average of the normalized process-of-care and outcome-of-care sub-composites mapped to a scale between zero and 100. Specifically, we add the lowest possible score a hospital can receive (i.e., a hospital scores zero percent on all process- and outcome-of-care indicators) to the standardized composite (\widetilde{C}_j^*) and divide by the range of potential hospital scores (i.e., the difference between the highest possible score a hospital can receive, which is if a hospital scores 100 percent on all process- and outcome-of-care indicators, and the lowest possible score). Specifically, the overall composite score (C_j^*) is

$$C_j^* = \frac{\widetilde{C}_j^* + \frac{1}{2} \left(\frac{\widehat{\mu}_P}{\widehat{\sigma}_P} + \frac{\widehat{\mu}_O}{\widehat{\sigma}_O} \right)}{\frac{1}{2} \left(\frac{1}{\widehat{\sigma}_P} + \frac{1}{\widehat{\sigma}_O} \right)} \quad (\text{eq. 2a.22.3})$$

where,

$$\widetilde{C}_j^* = \frac{1}{2} \left(\frac{P_j^* - \widehat{\mu}_P}{\widehat{\sigma}_P} \right) + \frac{1}{2} \left(\frac{O_j^* - \widehat{\mu}_O}{\widehat{\sigma}_O} \right) \quad (\text{eq. 2a.22.4})$$

Therefore, the variance of the composite measure $\text{Var}(C_j)$ can be estimated as

$$\begin{aligned} \text{Var}(C_j^*) &= \text{Var} \left[\frac{\widetilde{C}_j^* + \frac{1}{2} \left(\frac{\widehat{\mu}_P}{\widehat{\sigma}_P} + \frac{\widehat{\mu}_O}{\widehat{\sigma}_O} \right)}{\frac{1}{2} \left(\frac{1}{\widehat{\sigma}_P} + \frac{1}{\widehat{\sigma}_O} \right)} \right] \\ &= \text{Var} \left[\left(\frac{P_j^* - O_j^*}{\frac{1}{\widehat{\sigma}_P} + \frac{1}{\widehat{\sigma}_O}} \right) \right] \\ &= \text{Var} \left[\left(\frac{\sum_{k=1}^K \left(\frac{\sum_{j=1}^J n_{jk}}{\sum_{k=1}^K \sum_{j=1}^J n_{jk}} \right) P_{jk}^* - \sum_{k=1}^K \left(\frac{\sum_{j=1}^J n_{jl}}{\sum_{l=1}^L \sum_{j=1}^J n_{jl}} \right) O_{jl}^*}{\frac{1}{\widehat{\sigma}_P} + \frac{1}{\widehat{\sigma}_O}} \right) \right] \end{aligned}$$

$$\begin{aligned}
 &= \left(\frac{1}{\frac{1}{\sigma_P} + \frac{1}{\sigma_O}} \right)^2 \text{Var} \left[\left(\sum_{k=1}^K \left(\frac{\sum_{j=1}^J n_{jk}}{\sum_{k=1}^K \sum_{j=1}^J n_{jk}} \right) P_{jk}^* - \sum_{k=1}^K \left(\frac{\sum_{j=1}^J n_{jl}}{\sum_{l=1}^L \sum_{j=1}^J n_{jl}} \right) O_{jl}^* \right) \right] \\
 &= \left(\frac{1}{\left(\frac{1}{\sigma_P} + \frac{1}{\sigma_O} \right)} \right)^2 \left\{ \frac{1}{\left(\sum_{k=1}^K \sum_{j=1}^J n_{jk} \right)^2} \sum_{k=1}^K \left[\left(\sum_{j=1}^J n_{jk} \right)^2 \frac{P_{jk}^*(1 - P_{jk}^*)}{n_{jk}} \right] \right. \\
 &\quad \left. + \frac{1}{\left(\sum_{l=1}^L \sum_{j=1}^J n_{jl} \right)^2} \sum_{l=1}^L \left[\left(\sum_{j=1}^J n_{jl} \right)^2 \text{Var}(O_{jl}^*) \right] \right\}
 \end{aligned}$$

given the following assumptions^{iv}:

- A1. σ_P, μ_P and σ_O, μ_O are constants
- A2. $\text{cov}(P_{jm}^*, P_{jn}^*) = 0 \quad \forall m \neq n$
- A3. $\text{cov}(O_{jm}^*, O_{jn}^*) = 0 \quad \forall m \neq n$
- A4. $\text{cov}(P_{jm}^*, O_{jn}^*) = 0$

Testing/Analysis

2i. Component item/measure analysis to justify inclusion in composite

2i.1 Data/sample

As noted in Section 1d, the purpose of the proposed composite is to *summarize* the process- and outcome-of-care indicators associated with treatment of AMI that are now reported under the Hospital Inpatient Quality Reporting Program. Because we do not justify the composite in terms of the behavior of those indicators, our analysis aims to document the strength of associations among them; we are interested in the extent to which our formative measure does in fact represent a single construct of well-coordinated, high-quality care.

The analysis reported here relies on data that are publicly reported on Hospital Compare. We merged process-of-care indicators for AMI collected between July 2008 and June 2009 and outcome-of-care indicators for AMI collected between July 2006 and June 2009. A total of 4,990 hospitals were reported on Hospital Compare during this time period. Of these, we estimated

^{iv} These standard statistical assumptions are not ideal for small sample sizes and topped-off measures, but, in fact, many process measures are topped off or feature small sample sizes. If possible, we would like to test an alternate approach to constructing hospital-specific intervals, such as a Monte Carlo method.

AMI composite measures for 2,738 hospitals, with non-missing data for at least four of the seven process-of-care indicators and at least one of the two outcome-of-care indicators.

The seven AMI hospital process-of-care indicators used in the construction of composites were drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2008 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI were based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009. It is important to bear in mind that process-of-care indicators were reported for all patients and that outcome-of-care indicators were computed from claims data for Medicare patients only.

2i.2 Analytic Method

We carried out two analyses to explore the structure of the AMI indicators. First, we examined correlations among all process- and outcome-of-care indicators. Second, we conducted an exploratory factor analysis on the same process- and outcome-of-care indicators. Results appear in Tables 2i.3.1 and 2i.3.2

2i.3 Results

All correlations are positive, as Table 2i.3.1 shows, though many are weak, with values below 0.10. The two time-sensitive indicators (AMI 7A and AMI 8A) exhibit low correlation with other indicators. This is probably due to the high frequency of missing values for these two measures and their replacement with the overall mean. Correlations between process- and outcome-of-care indicators are low, though consistently positive. In addition, the Cronbach's alpha is 0.48, which is slightly below the commonly desired value of 0.70.

The factor analysis of component measures (Table 2i.3.2) produced a single factor with an eigenvalue greater than one. The eigenvalue for the first factor was more than 10 times that of the second factor, strongly suggesting that the component indicators represent a single underlying construct.

Production and Implementation of the CMS Hospital Outcomes and Efficiency Measures
Appendix A: Composite Measure of Hospital Quality for AMI
11.3.10

Table 2i.3.1. Correlation of Variables in AMI Composite Measure

	AMI 1	AMI 2	AMI 3	AMI 4	AMI 5	AMI 7A	AMI 8A	Mort	Read
AMI 1	1.00								
AMI 2	0.45	1.00							
AMI 3	0.25	0.30	1.00						
AMI 4	0.10	0.11	0.13	1.00					
AMI 5	0.35	0.50	0.35	0.12	1.00				
AMI 7A	0.06	0.05	0.04	0.04	0.03	1.00			
AMI 8A	0.22	0.24	0.14	0.12	0.21	0.06	1.00		
Mort ¹	0.07	0.11	0.00	0.05	0.07	0.00	0.17	1.00	
Read ¹	0.06	0.08	0.05	0.01	0.07	0.01	0.13	0.00	1.00
Cronbach's Alpha	0.48								

Notes:

1. Mort: survival rate, where Mort=100-(30-day risk-standardized mortality rate); Read: absence of readmission, where Read=100-(30-day risk standardized readmission rate).

Table 2i.3.2. Factor Analysis Results

Indicator	Factor Loadings					Uniqueness
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	
AMI 1	0.66	-0.09	0.03	-0.04	0.02	0.56
AMI 2	0.76	-0.09	0.02	-0.02	0.00	0.41
AMI 3	0.41	0.11	-0.02	0.10	-0.01	0.81
AMI 4	0.28	0.07	-0.11	0.07	-0.01	0.90
AMI 5	0.70	-0.05	0.00	0.01	-0.02	0.51
AMI 7A	0.06	0.07	0.06	0.08	0.05	0.98
AMI 8A	0.18	0.28	0.03	-0.03	0.00	0.89
Mort ¹	0.20	0.18	-0.08	-0.10	0.01	0.91
Read ¹	0.08	0.12	0.11	0.00	-0.03	0.96
Eigenvalues	1.83	0.17	0.04	0.03	0.00	
Proportion	1.18	0.11	0.02	0.02	0.00	
N	2,738					

Notes:

1. Mort: survival rate, where Mort=100-(30-day risk-standardized mortality rate); Read: absence of readmission, where Read=100-(30-day risk standardized readmission rate).

2j. Component item/measure analysis of contribution to variability in composite score

2j.1 Data/sample

As noted in Section 1d, the purpose of the proposed composite is to *summarize* the process- and outcome-of-care indicators associated with treatment of AMI that are now reported under the Hospital Inpatient Quality Reporting Program. Because we do not justify the composite in terms of the behavior of individual indicators, our analysis aims to document their contributions to the measure.

Analysis of the contribution of component items to the variability in composite scores uses data that are publicly reported on Hospital Compare. We merged process-of-care indicators for AMI with a data collection period of July 2008 to June 2009 and outcome-of-care indicators for AMI with a data collection period of July 2006 to June 2009. A total of 4,990 hospitals were reported on Hospital Compare during this time period. Of these, we estimated composite measures for 2,738 hospitals, for which less than or equal to three process-of-care indicators and less than or equal to one outcome-of-care indicator is missing.

The seven hospital process-of-care indicators related to AMI that are used in the construction of the AMI composite are calculated from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2008 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI are based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.

2j.2 Analytic Method

We compare the percentage change in (1) the variance and (2) the inter-quartile range (IQR) of the process- and outcome-of-care sub-composites when a process- or outcome-of-care indicator is removed before normalization. Results appear in Table 2j.3.1.

2j.3 Results

In Table 2j.3.1, the positive values indicate that addition of the component indicator tends to reduce the variance or IQR. Only one indicator, AMI2 (aspirin at discharge), exhibits a nontrivial positive effect on the composite variance, probably because of its relatively strong positive correlation with other component indicators (see Table 2i.3.1). Because the outcome domain contains only two component indicators, readmission and mortality both have strong negative effects on the variance of the sub-composite measure. The strong variance-reducing effect of readmission appears to be the result of its tight distribution (see Table 2i.3.2).

Table 2j.3.1. Change in Variance and Inter-quartile Range of the Process- and Outcome-of-Care Sub-Composites with the Removal of Indicators

	Change in Variance (%)	Change in Inter-Quartile Range (%)
Process Domain		
AMI 1	20.17	8.31
AMI 2	-15.60	-6.50
AMI 3	2.66	-2.60
AMI 4	10.04	4.03
AMI 5	-1.11	0.50
AMI 7A	0.28	0.29
AMI 8A	5.36	-4.19
Outcome Domain		
Mort ¹	34.53	2.32
Read ¹	163.27	68.48

Notes:

1. Mort: survival rate, where $Mort = 100 - (30\text{-day risk-standardized mortality rate})$; Read: absence of readmission, where $Read = 100 - (30\text{-day risk standardized readmission rate})$.

2k. Analysis to support differential weighting of component scores

2k.1 Data/sample

In constructing the composite, individual component indicators are weighted, in each instance, by the number of observations for the indicator. The most frequently reported indicators therefore affect the composite most strongly. In addition, the weighting scheme tends to reduce the variance of the composite, though this effect might be muted if individual indicators have similar distributions.

Testing to support differential weighting of composite scores uses data that are publicly reported on Hospital Compare by CMS. We merged process-of-care indicators for AMI with a data collection period of July 2008 to June 2009 and outcome-of-care indicators for AMI with a data collection period of July 2006 to June 2009. A total of 4,990 hospitals were reported on Hospital Compare during this period. Of these, we estimated AMI composite measures for 2,738 hospitals, for which less than or equal to three process-of-care indicators and less than or equal to one outcome-of-care indicator is missing.

The seven hospital process-of-care indicators related to AMI that are used in the construction of composites are drawn from Medicare hospital administrative claims data and medical record

documents with discharge dates between July 2008 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI are based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.

2k.2 Analytic Method

We compare the distribution of the AMI composite measure with equal and differential weighting. Specifically, the process-of-care sub-composite score and outcome-of-care sub-composite score (P_j^*) and outcome-of-care sub-composite score (O_j^*) with equal weighting is estimated respectively as,

$$P_j^* = \frac{1}{K} \sum_{k=1}^K P_{jk}^* \quad (\text{eq. 2k.2.1})$$

$$O_j^* = \frac{1}{L} \sum_{l=1}^L O_{jl}^* \quad (\text{eq. 2k.2.2})$$

where P_{jk}^* is the reliability-weight-adjusted rate for AMI process-of-care indicator $k=1, \dots, K$, in hospital j and O_{jl}^* is the risk-standardized rate for outcome-of-care indicator $l=1, \dots, L$, in hospital j .

The process- and outcome-of-care sub-composite score with differential weighting (that is, denominator weighting) is estimated respectively as

$$P_j^* = \sum_{k=1}^K \left(\frac{\sum_{j=1}^J n_{jk}}{\sum_{k=1}^K \sum_{j=1}^J n_{jk}} \right) P_{jk}^* \quad (\text{eq. 2k.2.3})$$

$$O_j^* = \sum_{l=1}^L \left(\frac{\sum_{j=1}^J n_{jl}}{\sum_{l=1}^L \sum_{j=1}^J n_{jl}} \right) O_{jl}^* \quad (\text{eq. 2k.2.4})$$

where n_{jk} is the number of hospital cases for AMI process-of-care indicator k and n_{jl} is the number of hospital cases for AMI outcome-of-care indicator l , in hospital j .

Then, overall composite score (C_j^*) is estimated using two steps. First, a standardized composite measure (\widetilde{C}_j^*) is estimated as a simple average of standardized process- and outcome-of-care sub-composites

$$\widetilde{C}_j^* = \frac{1}{2} \left(\frac{P_j^* - \widehat{\mu}_P}{\widehat{\sigma}_P} \right) + \frac{1}{2} \left(\frac{O_j^* - \widehat{\mu}_O}{\widehat{\sigma}_O} \right) \quad (\text{eq. 2k.2.5})$$

where $\widehat{\mu}_P$ is the average of the process-of-care sub-composite, $\widehat{\sigma}_P$ is the standard deviation of the process-of-care sub-composite, $\widehat{\mu}_O$ is the average of the outcome-of-care sub-composite, and $\widehat{\sigma}_O$ is the standard deviation of the outcome-of-care sub-composite. Then, the composite measure is re-standardized to a 0 to 100 scale. Given that the lowest possible composite score is if a hospital scores zero percent on all process- and outcome-of-care indicators and the highest possible composite score is if a hospital scores 100 percent on all process- and outcome-of-care indicators, we add the potential minimum composite value and divide by the range. Specifically, the overall composite score (C_j^*) is

$$C_j^* = \frac{\widetilde{C}_j^* + \frac{1}{2} \left(\frac{\widehat{\mu}_P}{\widehat{\sigma}_P} + \frac{\widehat{\mu}_O}{\widehat{\sigma}_O} \right)}{\frac{1}{2} \left(\frac{1}{\widehat{\sigma}_P} + \frac{1}{\widehat{\sigma}_O} \right)} \quad (\text{eq. 2k.2.6})$$

2k.3 Results

Table 2k.3.1 displays the distribution of the AMI composite measure with equal and differential weighting. As the table shows, denominator weighting has little effect on the distribution of the composite. The median is slightly larger when denominator weighting is used, and the inter-quartile range is somewhat smaller.

Table 2k.3.1. Comparison of Distribution of AMI Composite Measure^{1,2} by Weighting Method

Percentile	Weighting Method	
	Equal	Differential
Min	75.68	74.56
1%	79.05	79.73
5%	80.43	81.74
10%	81.09	82.45
25%	82.07	83.50
50%	82.86	84.27
75%	83.64	84.98
90%	84.40	85.66
95%	84.79	86.03
99%	85.43	86.83
Max	86.47	87.74
Mean	82.78	84.13
N	2,738	2,738

Notes:

1. Composite measures are estimated for hospitals if missing less than or equal to three process indicators and less than or equal to one outcome indicator.
2. Composite measures are estimated with imputation of the national mean, where the national mean is estimated as a simple average.

2k.4 Describe how the method of scoring/aggregation achieves the stated purpose and represents the quality construct

The objective of the composite is to summarize the component measures in a useful and scientifically acceptable manner.

Because composites are most useful to consumers if differences in composite values are clinically and statistically meaningful and reflect true differences in underlying quality, CMS entered component measures as values, not ranks, and adjusted those values for reliability. CMS entered component measures as values rather than ranks to prevent slight differences in composite values from producing large differences in composite values, as can occur when indicators are tightly distributed across hospitals. CMS also adjusted the component indicators

for reliability so that random variation did not drive small hospitals to extremes; 30-day outcome measures are adjusted for reliability before publication on Hospital Compare. Process measures are not adjusted for reliability before publication; the adjustment is made as part of the compositing process.

In addition, because composites are more useful to consumers if they emphasize measures that are relevant to a larger number of consumers, CMS constructed the process- and outcome-of-care composite scores using weights based on national denominators.

When sample sizes are equal, each component process measure contributes equally to the AMI process-of-care domain score. The same is true for each component outcome-of-care indicator. Thus a hospital that improves in any component will necessarily produce an increase in its composite score. Hospitals can therefore choose where to focus improvement efforts in evidence-based processes of care. Similar logic applies to the outcome-of-care domain score. The composite thus fully reflects the AMI process and outcome-of-care indicators and represents the quality construct expressed earlier.

2k.5 Indicate if any alternative scoring/aggregation methods were tested and why not chosen

In addition to the preferred compositing approach, ASI-RW, two alternative scoring methods were analyzed. These are referred to as (1) the absolute scoring index (ASI) and (2) the modified relative quality index (MRQI).

1. Absolute Scoring Index

ASI is similar to CMS' preferred approach but component indicators are not reliability-adjusted. To compute the ASI composite measure, process- and outcome-of-care sub-composite scores are first computed as the equally-weighted average of the indicators. These process and outcome scores are then scaled by subtracting the overall domain mean and dividing by the standard deviation. The ASI is then computed as the average of these two scaled means. The measure is then mapped to a scale between zero and 100, for ease of interpretation.

2. Modified Relative Quality Index

MRQI is similar to CMS' preferred approach but component indicators are not reliability-adjusted and enter the composite as ranks, not values. To compute the MRQI composite measure, scores for process- and outcome-of-care sub-composites are computed as the mean of the hospital's ranks for each indicator. The composite score is then computed as the simple mean of the two domain scores. It is closely related to the relative quality index (RQI) as described by Tompkins et al.¹ (January 2009, August 2009).

In Table 2k.5.1, we present distributions of the three alternative scoring methods. Broadly speaking, the distributions for ASI-RW, the preferred approach, look quite similar to the distribution for ASI.^v The difference is that the reliability adjustment has reduced the likelihood of erroneously classifying small hospitals as outliers due to random variation in measured performance by pulling them toward the mean of the distribution, though this is not visible in the table itself.

Results for MRQI show a more balanced distribution with medians close to means and less pronounced clustering in the upper half of the distribution, although there is still some clustering. Note that although this approach makes the distribution *look* more balanced, it does not address the fundamental problems of highly clustered performance on the underlying measures, small numbers of observations, and difficulty identifying meaningful differences in performance.

^v Although the shrinking process pulls the process-of-care indicators toward the mean, shrinking does not result in a smaller standard deviation of the distribution of the final composite values because each domain is normalized.

Table 2k.5.1. Comparison of Distribution of AMI Composite Measure^{1,2} by Scoring Method

Percentile	Scoring Method		
	Absolute Scoring Index	Modified Relative Quality Index	Absolute Scoring Index with Reliability Weights
Min	74.53	2.19	74.56
1%	79.39	15.14	79.73
5%	81.52	25.10	81.74
10%	82.25	30.03	82.45
25%	83.32	39.68	83.50
50%	84.09	49.83	84.27
75%	84.81	59.87	84.98
90%	85.47	68.07	85.66
95%	85.86	72.23	86.03
99%	86.65	80.42	86.83
Max	87.58	86.66	87.74
Mean	83.95	49.47	84.13
N	2,738	2,738	2,738

Notes:

1. Composite measures are estimated for hospitals if missing less than or equal to three process indicators and less than or equal to one outcome indicator.
2. Composite measures are estimated with imputation of the national mean, where the national mean is estimated as a simple average.

Citations

1. Tompkins, Chris, Grant Ritter, Andrew Ryan, Wato Nsa, James Burgess, and Dale Bratzler. *Composite Measures for Public Reporting: Final Report*. January 16, 2009 (updated August, 2009).

2l. Analysis of missing component scores

2l.1 Data/sample

The seven hospital process-of-care indicators used in the construction of the composite are drawn from reports for patients discharged between July 2008 and June 2009. Outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI are based on Medicare

claims for stays with discharge dates between July 2006 and June 2009.^{vi} Process and outcome indicators were reported on Hospital Compare for 4,990 hospitals during this period.

Because some hospitals did not report all indicators, some method for dealing with missing data was required. In order to compute a composite AMI measure for the greatest possible number of hospitals, we followed two principles:

1. All seven process-of-care indicators and both outcome-of-care indicators were included.
2. A composite measure was computed for each hospital that reported four or more process-of-care indicators and at least one outcome-of-care indicator.

AMI composites were computed for all hospitals that satisfied the second item. The national mean was used to impute a value for any missing process- or outcome-of-care indicators. AMI composite measures were computed for 2,738 hospitals.

21.2 Analytic Method

We used two approaches to conduct analysis of missing component scores. First, we tested to identify if there were differences in the composite measure for hospitals that did not fit the criteria stated previously. That is, we assessed whether composite measures differed significantly when our sample included composites with data for at least four process-of-care indicators and at least one outcome-of-care indicator (2,738 hospitals in total), compared with hospitals that did not have data for at least four process-of-care indicators and at least one outcome-of-care indicator (2,252 hospitals). Distributions of hospital composite scores were compared by the number of hospitals missing process- and outcome-of-care indicators. Second, we compared (1) distributions of process- and outcome-of-care indicators; (2) distributions of composite measures; (3) Spearman (rank) correlations; and (4) kappa statistics for hospital quartiles, with and without imputation of the national mean. Results appear in Tables 21.3.1 to 21.3.4.

21.3 Results

Hospitals are more likely to fail to meet the required minimum number of outcome-of-care indicators (one) than to fail to meet the minimum number of process-of-care indicators (four). That is, four or more process-of-care indicators are missing for 1,783 hospitals (35.7%), while both outcome-of-care indicators are missing for 2,102 hospitals (42.1%). Note that all component measures are missing for 1,190 hospitals (23.9%).

^{vi} The reporting periods for the process- and outcome-of-care measures represent the most recent version of both measures available on the Hospital Compare website, at the time of this NQF submission.

Production and Implementation of the CMS Hospital Outcomes and Efficiency Measures
Appendix A: Composite Measure of Hospital Quality for AMI
11.3.10

The distributions of component measures, shown in Table 21.3.2, are largely the same whether or not missing values are imputed. The clear exception is the distribution of process-of-care indicators AMI 7A and AMI 8A (fibrinolytic medication within 30 minutes of arrival and PCI within 90 minutes of arrival, respectively). We imputed these two measures far more often than other measures.

Table 21.3.3 shows the distribution of composites for the 2,738 hospitals for which at least four process measures and one outcome measure are reported. Composites with no imputation simply drop missing component measures from the calculation; composites with imputation use the national mean in place of the missing measure. As the table shows, the two different procedures yield nearly identical distributions for the AMI composite.

Table 21.3.4 shows the association between imputed and non-imputed measures by quartile. More than three-fourths of hospitals lie on the diagonal, occupying the same quartile for composite values using imputed and non-imputed component measures. The Spearman correlation coefficient and the kappa statistic both indicate a strong positive relationship between the two.

Table 21.3.1. Number of Missing Process- and Outcome-of-Care Indicators for AMI

		Number of Missing Outcome-of-Care Indicators			Total
		0	1	2	
Number of Missing Process-of-Care Indicators	0	278	1	5	284
	1	1,405	38	56	1,499
	2	615	157	145	917
	3	102	142	263	507
	4	19	41	297	357
	5	1	0	47	48
	6	1	6	99	106
	7	52	30	1,190	1,272
Total		2,473	415	2,102	4,990

Production and Implementation of the CMS Hospital Outcomes and Efficiency Measures
Appendix A: Composite Measure of Hospital Quality for AMI
11.3.10

Table 2I.3.2. Distribution of Rates for AMI Process and Outcome Indicators by Imputation Method

		1%	5%	10%	25%	50%	75%	90%	95%	99%	Mean	N
Process-of-Care Indicators												
AMI 1	Without ¹	0.0	71.0	86.0	96.0	99.0	100.0	100.0	100.0	100.0	94.5	3,672
	National ²	39.0	80.0	90.0	94.5	97.0	100.0	100.0	100.0	100.0	94.5	4,990
AMI 2	Without	0.0	67.0	81.0	95.0	99.0	100.0	100.0	100.0	100.0	93.3	3,586
	National	0.0	73.0	88.0	93.3	96.0	100.0	100.0	100.0	100.0	93.3	4,990
AMI 3	Without	0.0	60.0	78.0	92.0	100.0	100.0	100.0	100.0	100.0	91.8	3,015
	National	0.0	75.0	87.0	91.8	91.8	100.0	100.0	100.0	100.0	91.8	4,990
AMI 4	Without	0.0	83.0	96.0	100.0	100.0	100.0	100.0	100.0	100.0	96.4	2,853
	National	0.0	95.0	96.4	96.4	98.0	100.0	100.0	100.0	100.0	96.4	4,990
AMI 5	Without	0.0	67.0	83.0	96.0	99.0	100.0	100.0	100.0	100.0	93.8	3,588
	National	33.0	75.0	88.5	93.8	97.0	100.0	100.0	100.0	100.0	93.8	4,990
AMI 7A	Without	0.0	0.0	0.0	0.0	50.0	100.0	100.0	100.0	100.0	45.7	643
	National	0.0	17.0	45.7	45.7	45.7	45.7	45.7	62.0	100.0	45.7	4,990
AMI 8A	Without	0.0	43.0	58.0	74.0	86.0	93.0	99.0	100.0	100.0	80.6	1,511
	National	33.0	67.0	79.0	80.6	80.6	80.6	91.0	96.0	100.0	80.6	4,990
Outcome-of-Care Indicators												
Mortality ³	Without	12.3	13.5	14.1	15.1	16.1	17.2	18.3	19.0	20.6	16.2	2,884
	National	12.7	14.0	14.6	15.9	16.2	16.4	17.7	18.4	20.1	16.2	4,990
Readmission ³	Without	17.0	17.8	18.4	19.2	19.9	20.8	21.5	22.1	23.5	20.0	2,477
	National	17.3	18.4	19.0	19.9	19.9	19.9	21.0	21.5	22.9	19.9	4,990

Notes:

1. Average rates are estimated without imputation.
2. Average rates are estimated with imputation of the national mean, where the national mean is estimated as a simple average.
3. Mortality and readmission rates are shown in tables and not survival or non-readmission rates, as used in the construction of the composite measure.

Table 21.3.3. Comparison of Distribution of AMI Composite Measure¹ by Imputation Method

Percentile	Imputation Method	
	None	National Mean ²
Min	73.25	74.56
1%	79.64	79.73
5%	81.99	81.74
10%	82.75	82.45
25%	83.80	83.50
50%	84.60	84.27
75%	85.33	84.98
90%	86.04	85.66
95%	86.46	86.03
99%	87.18	86.83
Max	88.29	87.74
Mean	84.46	84.13
N	2,738	2,738

Notes:

1. Composite measures are estimated for hospitals if missing less than or equal to three process-of-care indicators and less than or equal to one outcome-of-care indicator.
2. Composite measures are estimated with imputation of the national mean, where the national mean is estimated as a simple average.

Table 21.3.4. Comparison of AMI Composite Measures¹ by Imputation Method

	Imputation with National Mean ²				Total
	Quartile 1	Quartile 2	Quartile 3	Quartile 4	
Quintile 1 ³	628	54	3	0	685
Quartile 2	35	530	119	0	684
Quartile 3	21	39	494	131	685
Quartile 4	1	61	69	553	684
Total	685	684	685	684	2,738
Spearman Correlation	0.93 (0.00)				
Kappa Statistic	0.74 (0.00)				

Notes:

1. Composite measures are estimated for hospitals if missing less than or equal to three process indicators and less than or equal to one outcome indicator.
2. Composite measures are estimated with imputation of the national mean, where the national mean is estimated as a simple average.
3. Higher quintile categories indicate that the hospital had higher (i.e., better quality) composite measures.

* *p-values* in parentheses.

2b. Reliability testing of composite score

2b.1 Data/sample (description of data/sample and size):

The reliability of the proposed AMI composite measure is informed by the reliability of the component scores on which it is based. Two reports, one by Williams et al. (2006) and the other by the Government Accountability Office (GAO) (2006), provide insight into component measure reliability:

Williams, SC, Watt, A, Schmaltz, SP, Koss, RG, & Loeb, JM. *Assessing the reliability of standardized performance indicators*. Int J Qual Health Care. 2006 Jun;18(3):246-55. Epub 2006 Jan 23.

- Williams et al. (2006) examined the reliability of all seven AMI process-of-care indicators that make up the AMI composite. Their sample included 30 hospitals, representing a diverse range of geographic locations, sizes, settings (urban/rural), and ownership categories (profit/not-for-profit); 19 of these collected AMI data. A randomly selected set of deidentified, previously abstracted medical records was transmitted from the hospitals' performance measurement vendors and AMI process-of-care indicators were reabstracted following guidelines from the Specification Manual for National Implementation of Hospital Core Measures. Sample sizes used to calculate each measure generally ranged from 100–200 cases, though for AMI-4

(smoking cessation counseling) and AMI-8A (first PCI time) the sample size was fewer than 50.

United States. Government Accountability Office. Report to the Committee on Finance, U.S. Senate. *Hospital Quality Data: CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data*. Report No. GAO-06-54, Jan. 31, 2006

- The 2006 GAO report summarizes CMS' process to assess the reliability of the measures currently reported on Hospital Compare and reports the results of this process for hospital discharges between January 1, 2004, and June 30, 2004. CMS' contractor, CDAC (Clinical Data Abstraction Center), assesses the reliability of the component measures on a quarterly basis. This assessment uses a sample of five randomly selected patient records from each hospital participating in the Hospital Inpatient Quality Reporting Program, which includes hospitals from all states but Maryland and Puerto Rico.^{vii}

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

Williams, SC, Watt, A, Schmaltz, SP, Koss, RG, & Loeb, JM. *Assessing the reliability of standardized performance indicators*. Int J Qual Health Care. 2006 Jun;18(3):246-55. Epub 2006 Jan 23.

- Reliability was assessed using percentage agreement for continuous variable elements and chance-corrected agreement using Cohen's kappa for binary data elements.

United States. Government Accountability Office. Report to the Committee on Finance, U.S. Senate. *Hospital Quality Data: CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data*. Report No. GAO-06-54, Jan. 31, 2006

- For each hospital, data are deemed reliable if there is 80 percent or greater agreement between the hospital quality data previously submitted to CMS and the CDAC reabstraction results.

^{vii} As a result of the GAO report, in 2010 this process changed so that CDAC instead reviews 12 patient records from a randomly selected sample of 800 hospitals.

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

Williams, SC, Watt, A, Schmaltz, SP, Koss, RG, & Loeb, JM. *Assessing the reliability of standardized performance indicators*. Int J Qual Health Care. 2006 Jun;18(3):246-55. Epub 2006 Jan 23.

- Table 2b.3.1 summarizes the reliability statistics for the AMI measures that are included in the proposed composite. Using the standards proposed by Landis & Koch (1977),¹ the resulting kappas indicate almost perfect agreement (kappa > 0.81) for three of the measures, substantial agreement (kappa ranging from 0.61 to 0.80) for one measure, and moderate agreement (kappa ranging from 0.41 to 0.60) for two measures. Although a kappa was not calculated for AMI-8A (first PCI time), the authors report 64.7 percent agreement for this measure.

Table 2b.3.1. Reliability Findings by Williams et al. 2006

AMI Component Measure	Number	Agreement (%)	Kappa
AMI-1	200	90.5	0.54
AMI-2	156	84.6	0.52
AMI-3	101	91.1	0.82
AMI-4	44	93.2	0.85
AMI-5	156	91.0	0.76
AMI-7A	143	95.8	0.81
AMI-8A	34	64.7	Not Calculated

United States Government Accountability Office. Report to the Committee on Finance, U.S. Senate. *Hospital Quality Data: CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data*. Report No. GAO-06-54, Jan. 31, 2006

- The GAO report, which looked at reporting from January 1, 2004, through June 30, 2004, found that 90 percent of hospitals exceeded the 80 percent reliability threshold.

Citations

1. Landis, J.R., & Koch, G.G. (1977). *The measurement of observer agreement for categorical data*. Biometrics 33: 159–174

2c. Validity testing of composite score

2c.1 Data/sample (description of data/sample and size)

The testing of the validity of the component scores uses two sets of data. The first data set merges process- and outcome-of-care indicators for AMI with a data collection period of July 2008 to June 2009. The second data set merges process- and outcome-of-care indicators for AMI with a data collection period of July 2007 to June 2008. Composite measures are calculated from

these two separate periods and compared, with the assumption that a valid composite measure should show minimal change on a year-to-year basis.

Across these two data collection periods, 1,747 hospitals had valid composite measures for AMI.

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

Using the two sets of data, we compared composite measures across the two years using the Spearman (rank) correlation coefficient to evaluate the predictive validity of the composite measure over time.

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

The Spearman correlation between composite measures computed in 2007–2008 and 2008–2009 was 0.43 ($p < 0.001$), indicating moderate predictive validity of the composite. (See Table 2c.3.1.) A large number of hospitals (about 40 percent) lie on the diagonal, such that the same hospital quartiles for composite values were occupied during 2007–2008 and 2008–2009. In contrast, very few hospitals (about 5 percent) occupy the first quartile in 2007–2008 and the fourth quartile in 2008–2009, and vice versa. Across the two separate periods, about 40 percent of hospitals' categorizations differ by one quartile (that is, during 2008–2009, a hospital was one quartile above or below its categorization in 2007–2008). This discrepancy appears to be a result of the tight distribution of the process- and outcome-of-care indicators, as shown in Table 21.3.2.

Table 2c.3.1. Comparison of AMI Composite Measures¹ by Reporting Period

2007-2008 Reporting ²	2008-2009 Reporting ³				Total
	Quartile 1	Quartile 2	Quartile 3	Quartile 4	
Quintile 1 ⁴	191	107	67	35	400
Quintile 2	124	126	102	88	440
Quintile 3	62	113	153	115	443
Quintile 4	50	93	112	209	464
Total	427	439	434	447	1,747
Spearman Correlation	0.43 (0.00)				
Kappa Statistic	0.18 (0.00)				

Notes:

1. Composite Measures are estimated for hospitals if missing less than or equal to three process indicators and less than or equal to one outcome indicator.
 2. 2007-2008 reporting: process- and outcome-of-care measures for AMI with a data collection period of July 2007 to June 2008
 3. 2008-2009 reporting: process- and outcome-of-care measures for AMI with a data collection period of July 2008 to June 2009
 4. Higher quartile categories indicate that the hospital had higher (i.e., better quality) composite measures.
- *. *p-values* in parentheses.

2f. Identification of Meaningful Differences in Performance Across Entities

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

Testing to identify meaningful differences in performance of composite scores uses data that are publicly reported on Hospital Compare by CMS. We merged process-of-care indicators for AMI with a data collection period of July 2008 to June 2009 and outcome-of-care indicators for AMI with a data collection period of July 2006 to June 2009. A total of 4,990 hospitals were reported on Hospital Compare during this period. Of these hospitals, we estimated composite measures for 2,738, for which less than or equal to three process-of-care indicators and less than or equal to one outcome-of-care indicator is missing.

The seven hospital process-of-care indicators related to AMI that are used in the construction of composites are drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2008 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for AMI are based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.

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

To examine meaningful differences in composite measures across hospitals, we compared hospitals’ confidence interval estimates with the overall mean and assigned hospitals into one of three performance categories: better than hospitals, if the interval estimate is entirely above the mean; no different than hospitals, if the interval estimate includes the mean; and worse than hospitals, if the interval estimate is entirely below the mean. These performance categories do not reflect how the composites will ultimately be displayed on Hospital Compare.

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 meaningful differences in performance):

CMS has not decided how it will ultimately display hospital performance to consumers on Hospital Compare or to providers in hospital-specific reports. Table 2f.3.1 provides the number of hospitals in each of the three performance categories. These performance categories do not reflect how the composites will ultimately be displayed on Hospital Compare.

Table 2f.3.1. Number of Hospitals in Alternative Performance Categories

Performance Category	Number of Hospitals in Performance Category
Better than Mean	224
No Different than Mean	2,421
Worse than Mean	93

Contact Information

Co.1 Measure Steward (Intellectual Property Owner)
Organization: [Centers for Medicare & Medicaid Services](#)
Street Address: [7500 Security Boulevard, Mail Stop S3-02-01](#)
City: [Baltimore](#)
State: [MD](#)
Zip: [21244](#)

Co.2 Point of Contact:
First Name: [Shaheen](#)
Last Name: [Halim](#)
Credentials: [Ph.D., CPC-A](#)
Email: Shaheen.Halim@cms.hhs.gov
Telephone: [\(410\) 786-0641](#)
Ext:

Co.3 Measure Developer (If different from Measure Steward)
Organization: [Mathematica Policy Research](#)
Street Address: [955 Massachusetts Avenue, Suite 801](#)
City: [Cambridge](#)
State: [MA](#)
Zip: [02139](#)

Co.4 Point of Contact:
First Name: [Marian](#)
Last Name: [Wrobel](#)
Credentials: [Ph.D.](#)
Email: MWrobel@mathematica-mpr.com
Telephone: [617-301-8971](#)
Ext:

**The National Quality Forum
Composite Measures of Hospital Quality for Acute Myocardial
Infarction, Heart Failure, Pneumonia, and the Surgical Care
Improvement Project
Appendix B**

BACKGROUND

Composite measures are used in many contexts or settings to provide a broad picture of the performance, behavior, traits and other characteristics of individuals or other types of entities. In general, composite measures combine quantitatively two or more separate measures into a single measure or index. Within health care, a composite measure can be formed by combining quantitatively the performance data of providers across multiple measures.

Such composite measures of provider performance serve two primary goals. First it summarizes a large amount of information about the performance of a provider. This type of summary can be useful for giving consumers provider-related performance information. Much research has shown that consumers find it difficult and frustrating to sort through multiple performance measures to arrive at a conclusion regarding the performance of a provider from whom they are contemplating receiving care (Hibbard et al., 2000; Hibbard, 2001). Thus composites are a potentially useful tool for sponsors of consumer report cards and other types of vehicles for disseminating information about provider performance to consumers. Providers also may benefit when their performance information is presented in a summary form if the summary offers insight about opportunities for improvement.

Second, it increases measurement reliability for providers. As provider profiling and consumer report cards have become widely used, researchers have raised concerns about the reliability of performance measurement. Studies have demonstrated that measurement reliability is often below acceptable levels because of small sample sizes for providers (Zaslavsky, 2001). The construction of composites may be used to address this problem by combining, for a given provider, the number of patients across the multiple measures.

With respect to the information summarized, composites for healthcare measures are likely to comprise process measures, outcome measures or some combination of the two. Although in the field of health services research, process measures are sometimes treated as an intermediate measure for outcomes within conceptual models of quality of care, there is no consensus that process measures are not important in their own right for assessing quality of care. First, it is not clear that process scores consistently correspond with outcomes as studies examining the statistical correlations between process and outcome measures often report mixed results. In addition, more recent studies using sophisticated measurement techniques seem to indicate that they are not related strongly (e.g. Jha et al., 2007; Ryan et al., 2009). Second, for quality improvement, processes always are much more under the control of providers than are outcomes as they offer guidance as to what actions provider can undertake to improve scores. As such, many providers appear to value process measures for purposes of quality assessment.

There are two general approaches for constructing composites (Shwartz et al., 2009). One approach is to construct “reflective” composites. A reflective composite seeks to combine multiple measures that theoretically are believed to be linked to an underlying construct that cannot be directly measured such as quality or intelligence. The construction of a reflective

construct requires that the individual measures be highly correlated as they are treated theoretically as representing different dimensions of the same construct. The other approach is to construct “formative” composites. A formative composite is essentially a combination of multiple measures that are intended to provide useful summary information but without a strong theoretical rationale that they are linked to the same construct. As such, there is no expectation that the individual measures comprising the composite will be highly correlated or meet other psychometric tests that are considered standard for the construction of a valid reflective composite. In particular, then, reflective measures may gain validity and reliability by summarizing information from individual indicators in a condensed form. Such a result may or may not hold for particular formative measures.

CMS HOSPITAL COMPARE COMPOSITES

CMS has developed composite measures for four conditions that are part of the accepted set of measures from the CMS Hospital Compare system: Acute Myocardial Infarction (AMI), Heart Failure (HF), Pneumonia (PN), and Surgical Care Improvement Project (SCIP). For three of these four conditions (i.e., AMI, HF, and PN), both process and outcome measures are available for constructing composites. For SCIP, process measures are available only. For constructing the composites, the process and outcome measures were treated as separate domains. All the measures comprising the composites have previously been reviewed and endorsed by the National Quality Forum (NQF). Because CMS plans to include these composite measures in the Hospital Compare website, which is a consumer-oriented tool for comparing provider performance, a primary goal is to summarize information in a way that will be helpful to consumers.

The construction of these composites was conducted in manner that is consistent with a formative approach. There are several considerations that are relevant to this decision. First, the process by which the measures comprising each composite evolved and were chosen for Hospital Compare did not take place with a reflective construct in mind. The measures were developed, evaluated, and considered for NQF endorsement separately, each on their own merits. Thus, we consider these constructs formative in that they summarize an array of measures for that condition. Second, each of the four conditions is complex in etiology and treatment, so that it is difficult or even impossible to condense the measures into simple and valid conceptual constructs as would be seen in reflective composites. Yet, the decisions from a patient, provider, and healthcare system level on evaluating quality for individual treatment conditions need to be made. We cannot pick and choose to take the treatment of one hospital for one measure and another hospital for another measure; the treatment comes as a package. Third, composites are intended to be flexible for future additions or deletions of measures. CMS policy on the appropriate measures for these conditions and possibilities for additional conditions will adapt to measure development opportunities and changes in the evidence base underlying both process and outcome measures over time. Finally, the process and outcome measures themselves have different theoretical constructs, are affected differently by the actions of providers, and may not be causally related to each other. As such, for each of these four conditions now, and for any

new conditions that are added, formative composites can be developed following the technical procedures that have been outlined in the initial NQF submissions for each of these composites.

A key technical decision as to the construction of the composites was to weight the process and outcome domains equally by standardizing each domain score, before combining into a single composite score. The decision to weight equally was based on the consideration that no strong theoretical foundation existed for assigning differential weights. In this sense, the rationale is similar to the decision to construct the composites as a formative measure. Since the measures are not necessarily drawn from a consistent unifying underlying construct, there may not really be a population standard deviation for each measure to be estimating by the sample standard deviation. Also, for true equal weighting to be achieved, standardization of the domain scores is necessary. This is because the impact of any measure on a composite with equal weighting will be proportional to the standard deviation of the underlying measure. Measures which vary more will have greater influence on the composite measure and the ranking of entities measured. Z-score methods to normalize measures to mean 0 and standard deviation of 1 are possible to equalize the influence across all measures, but this is undesirable since it greatly inflates the influence of measures with very small standard deviation measured differences that likely have little to no clinical or practical significance. In fact, for practical implementation of a composite measure where expert opinion is not being brought to bear on weighting, equal weighting where the standard deviation impact is allowed to pass through to the composite measure actually is more acceptable.

REFERENCES

- Hibbard, JH (2001). Being Strategic with Public Reports. *The International Journal for Quality in Health Care*, Vol. 13, No. 5, pp. 353-355.
- Hibbard, JH, Harris-Kojetin L, Mullin P, Lubalin JS, Garfinkel SA (2000). Increasing the Impact of Health Plan Report Cards by Addressing Consumer Concerns. *Health Affairs*, Vol. 19, No. 5, pp. 138-143.
- Jha, Ashish K., E. John Orav, Zhonghe Li and Arnold M. Epstein (2007). The Inverse Relationship Between Mortality Rates And Performance In The Hospital Quality Alliance Measures, *Health Affairs*, Vol. 26, No. 4, pp. 1104-1110.
- Ryan, Andrew M., James F. Burgess, Jr., Christopher P. Tompkins, Stanley S. Wallack (2009). The Relationship between Medicare's Process of Care Quality Measures and Mortality. *Inquiry*, Vol. 46, No. 3, pp. 274-290.
- Shwartz, Michael, James F. Burgess, Jr., and Dan Berlowitz (2009). Benefit-of-the-Doubt Approaches for Calculating a Composite Measure of Quality, *Health Services and Outcomes Research Methodology*, Vol. 9, No. 4, pp. 234-251.
- Zaslavsky, Alan M. (2001). Statistical issues in reporting quality data: small samples and casemix variation, *International Journal for Quality in Health Care*, Volume 13, Issue 6, pp. 481-488.

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 #: 0230	NQF Project: Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION	
De.1 Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization	
De.2 Brief description of measure: The measure estimates a hospital-level risk-standardized mortality rate (RSMR), defined as death from any cause within 30 days after the index admission date, for patients discharged from the hospital with a principal diagnosis of AMI.	
1.1-2 Type of Measure: Outcome	
De.3 If included in a composite or paired with another measure, please identify composite or paired measure This measure is paired with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission rate (RSRR) following an AMI hospitalization.	
De.4 National Priority Partners Priority Area: Safety	
De.5 IOM Quality Domain: Effectiveness, Patient-centered, Safety	
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. <i>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.</i>	A Y <input checked="" type="checkbox"/> N <input type="checkbox"/>
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:	

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 <input checked="" type="checkbox"/> N <input type="checkbox"/>
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ► Purpose: Public reporting, Internal quality improvement	C Y <input checked="" type="checkbox"/> N <input type="checkbox"/>
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 (for NQF staff use) Have all conditions for consideration been met? Staff Notes to Steward (<i>if submission returned</i>): Staff Notes to Reviewers (<i>issues or questions regarding any criteria</i>): Risk model attachments.	D Y <input checked="" type="checkbox"/> N <input type="checkbox"/> Met Y <input type="checkbox"/> N <input type="checkbox"/>
Staff Reviewer Name(s): RWinkler	

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. <i>Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)</i> 1a. High Impact	Eval Ratin g
(for NQF staff use) Specific NPP goal : Safety - improve 30-day morality for AMI	
1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, Leading cause of morbidity/mortality, High resource use, Severity of illness 1a.2 1a.3 Summary of Evidence of High Impact: Acute myocardial infarction (AMI) is one of the most common principal hospital discharge diagnoses among Medicare beneficiaries and is associated with high mortality. The high prevalence and considerable morbidity and mortality associated with AMI create an economic burden on the healthcare system. (American Heart Association, 2010) In 2005, AMI was the fourth most expensive condition treated in US hospitals, accounting for nearly 4% of the national hospital bill. It was also the fourth most expensive condition billed to Medicare that year, accounting for 4.5% of Medicare’s hospital bill (Andrews and Elixhauser, 2007). Many current hospital interventions are known to decrease the risk of death within 30 days of hospital admission. (Jha 2007, Rathore 2009) Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals. 1a.4 Citations for Evidence of High Impact: American Heart Association. Heart Disease and Stroke Statistics - 2010 Update. Dallas, Texas: American Heart Association; 2010. c2010, American Heart Association.	1a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>

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).

Andrews RM, Elixhauser A. The national hospital bill: growth trends and 2005 update on the most expensive conditions by payer. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2007 Dec. (HCUP statistical brief; no. 42).

Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. Health Aff (Millwood) 2007 Jul-Aug;26(4):1104-10.
 Rathore SS, Curtis JP, Chen J, Wang Y, Nallamothu BK, Epstein AJ, Krumholz HM; National Cardiovascular Data Registry. Association of door-to-balloon time and mortality in patients admitted to hospital with ST elevation myocardial infarction: national cohort study. BMJ. 2009 May 19;338

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized mortality rates following hospitalization for AMI. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This mortality measure was developed to identify institutions whose performance is better or worse than would be expected based on their patient case-mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:

Recent analyses show substantial variation in RSMRs among hospitals. For the most recently reported three years of data (7/2006-6/2009) the mean hospital RSMR was 15.9% with a range of 10.3% to 24.6%. The 5th percentile was 13.2 and the 95th was 18.4. The interquartile range was 15.0% to 16.8%.

This work also demonstrated ongoing geographic variation in hospital RSMRs for AMI.

Bernheim SM, Grady JN, Lin Z, Wang Y, Wang Y, Savage SV, Bhat KR, Ross JS, Desai MM, Merrill AR, Han LF, Rapp MT, Drye EE, Normand SL, Krumholz HM. National patterns of risk-standardized mortality and readmission for acute myocardial infarction and heart failure. Update on publicly reported outcomes measures based on the 2010 release. Circ Cardiovasc Qual Outcomes. 2010 Sep 1;3(5):459-67. Epub 2010 Aug 24.

1b.3 Citations for data on performance gap:

This data on the performance gap is based on RSMRs calculated for AMI hospitalizations from July 1, 2006-June 30, 2009 and includes 558,665 hospitalizations from 4569 hospitals. The index hospitalizations are those included in the measure and reported in the 2010 update to the Hospital Compare website.

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

CMS supported analyses to evaluate disparities in performance by hospitals based on the proportion of patients that they serve who are African-American. These analyses show that the range of performance is similar for hospitals with higher proportions of African-American patients compared to hospitals with lower proportions. We divided hospitals into deciles based on the proportion of their patients that were African-American and looked at hospitals across deciles. The combined lowest 5 deciles have fewer than 5% African-American patients and a median AMI RSMR of 16.3% (range 10.6- 23.2) vs hospitals in the highest decile with >25% African American patients and a median AMI RSMR 16.2% (range 11.8-24.6).

Similar analyses were completed to evaluate hospital differences in performance based on the socioeconomic status of their patients. These analyses suggest a slightly higher median AMI RSMR at the hospitals in the lowest quartile based on the socioeconomic status (SES) of their patients (as measured by median income of the patient's zip code). The lowest quartile hospitals' median RSMR was 16.8 compared to median RSMR of 15.8 for hospitals in highest quartile of patient SES. However the range for the two groups was largely overlapping (11.6-24.6 vs 10.6-22.0) demonstrating that substantial numbers of hospitals serving low SES patients perform well on the measure. A recently published study also demonstrated that patient SES

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.

1b
 C
 P
 M
 N

accounted for a very small portion of variation in hospital performance on the AMI RSMR measure (Bradley, 2010).

Reference: Bradley EH, Herrin J, Curry L, Cherlin EJ, Wang Y, Webster TR, Drye EE, Normand SL, Krumholz HM. Variation in hospital mortality rates for patients with acute myocardial infarction. *Am J Cardiol.* 2010 Oct 15;106(8):1108-12.

1b.5 Citations for data on Disparities:

The sample for the above analyses is from a similar 3 year cohort of hospitalizations as the data for the performance gap analysis above (January 2006- December 2008) but limited to hospitals with at least 25 AMI cases over the 3 year period, a total of 2943 hospitals.

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): This measure calculates hospital-level, 30-day all-cause mortality rates after hospitalization for an AMI. The goal is to directly affect patient outcomes by measuring risk-standardized rates of mortality.

1c.2-3. Type of Evidence: Systematic synthesis of research

1c.4 Summary of Evidence (as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome):

Many hospital interventions, such as use of appropriate medications, timely percutaneous coronary interventions and prevention of complications are known to decrease the risk of death within 30 days of hospital admission. (Rathore 2009, Antman 2007). Over the last 10 years, nationally, risk-standardized mortality rates have decreased for AMI (Krumholz 2009). Yet, continued variation in performance suggests continued opportunities for improvements.

In addition, recent qualitative research funded by AHRQ, Commonwealth Fund, and United Healthcare, identified common system-level approaches to care and, specifically, the tailored use of protocols in those hospitals that have low RSMRs compared with hospitals with high RSMRs. (paper in submission) These findings are being validated in a large national hospital survey.

References:

Rathore SS, Curtis JP, Chen J, Wang Y, Nallamothu BK, Epstein AJ, Krumholz HM; National Cardiovascular Data Registry. Association of door-to-balloon time and mortality in patients admitted to hospital with ST elevation myocardial infarction: national cohort study. *BMJ.* 2009 May 19;338

Antman EM, Hand M, Armstrong PW, et al. 2007 focused update of the ACC/AHA 2004 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. *J Am Coll Cardiol.* Jan 15 2008;51(2):210-247.

Jha AK, Orav EJ, Li Z, Epstein AM. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. *Health Aff (Millwood)* 2007 Jul-Aug;26(4):1104-10.

Krumholz HM, Wang Y, Chen J, Drye EE, Spertus JA, Ross JS, Curtis JP, Nallamothu BK, Lichtman JH, Havranek EP, Masoudi FA, Radford MJ, Han LF, Rapp MT, Straube BM, Normand SL. Reduction in acute myocardial infarction mortality in the United States: risk-standardized mortality rates from 1995-2006. *JAMA.* 2009 Aug 19;302(7):767-73.

1c.5 Rating of strength/quality of evidence (also provide narrative description of the rating and by whom): N/A (outcomes measure)

1c.6 Method for rating evidence: N/A (outcomes measure)

1c
 C
 P
 M
 N

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. ... [1]

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., mammography) or measures for multiple care processes that affect a single outcome.

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.

1c.7 Summary of Controversy/Contradictory Evidence: Use of Hierarchical Generalized Linear Modeling
 Hierarchical modeling for hospital outcomes measurement is the appropriate statistical approach for hospital outcomes measures given the structure of the data and the underlying assumption of such measures, which is that hospital quality of care influences 30-day mortality rates. However, CMS frequently receives comments and questions about this approach, so we are concisely reiterating the rationale for and merits of using hierarchical logistic regression. Patients are clustered within hospitals and, as such, have a shared exposure to the hospital quality and processes. The use of hierarchical modeling accounts for the clustering of patients within hospitals. Second, hierarchical models distinguish within-hospital variation and between-hospital variation to estimate the hospital's contribution to the risk of mortality. This allows for an estimation of the hospital's influence on patient outcomes. Finally, within hierarchical models we can account for both differences in case mix and sample size to fairly profile hospital performance. If we did not use hierarchical modeling we could overestimate variation and potentially misclassify hospitals' performance. Accurately estimating variation is an important objective for models used in public reporting and potentially used in value-based purchasing programs.

Effect of patient-preferences regarding end-of-life care
 In certain cases, the best quality care may ultimately be that which supports patients' goals and comfort at the end of life rather than that which prolongs life. The intent of a mortality rate is not to convey that all deaths are the result of poor care. The goal is not to have zero deaths. The premise is that there are preventable deaths. Knowledge of how an institution performs compared with what might be expected given their case mix is helpful in encouraging efforts to improve outcomes.

Some stakeholders have expressed concerns that our measure cannot adequately exclude patients who choose comfort measures or palliative care during their index hospitalization. Stakeholders are concerned that this could lead to unintended consequences, such as prolonging lives against patient wishes. To address these issues CMS has taken the following steps:

- (1) CMS added an exclusion for patients who are enrolled in the Medicare hospice program prior to, or on the day of, admission.
- (2) CMS chose not to exclude patients who are discharged to hospice or seek a palliative care consult during admission to account for the fact that the choice of palliative/comfort care may be the result of poor care.
- (3) To account for risk-factors associated with the end of life CMS included markers of frailty within our risk-adjustment variables, including: protein-calorie malnutrition, dementia or senility, and hemiplegia, paraplegia, paralysis and functional disability .
- (4) CMS is looking into the possibility of adding POA codes to the palliative care consult ICD-9 code (v. 66.7) to gather more information, but would need to give further consideration to the clinical and measurement implications before instituting any changes to the measure using this code.
- (5) Although CMS is confident in the current models, CMS will further consider clinical and measurement issues for patients for whom survival is not an objective as it maintains this mortality measure.

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

1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number):
 N/A

1c.10 Clinical Practice Guideline Citation: N/A

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):
 N/A

1c.13 Method for rating strength of recommendation (If different from [USPSTF system](#), also describe rating and how it relates to USPSTF):
 N/A

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.

1c.14 Rationale for using this guideline over others: N/A	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Importance to Measure and Report</i> ?	1
Steering Committee: Was the threshold criterion, <i>Importance to Measure and Report</i> , met? Rationale:	1 Y <input type="checkbox"/> N <input type="checkbox"/>
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, <u>as specified</u> , produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Ratin g
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: <u>2a. Precisely Specified</u>	
2a.1 Numerator Statement (<i>Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome</i>): This outcome measure does not have a traditional numerator and denominator like a core process measure (e.g., percentage of adult patients with diabetes aged 18-75 years receiving one or more hemoglobin A1c tests per year); thus, we are using this field to define the outcome. The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days of the index admission date for patients discharged from the hospital with a principal diagnosis of AMI. 2a.2 Numerator Time Window (<i>The time period in which cases are eligible for inclusion in the numerator</i>): Patients who die within 30 days of the index admission date. 2a.3 Numerator Details (<i>All information required to collect/calculate the numerator, including all codes, logic, and definitions</i>): Measure includes deaths from any cause within 30 days from admission date of index hospitalization. 2a.4 Denominator Statement (<i>Brief, text description of the denominator - target population being measured</i>): Note: This outcome measure does not have a traditional numerator and denominator like a core process measure; thus, we are using this field to define the patient cohort. The cohort includes admissions for Medicare FFS beneficiaries age 65 years or older discharged from the hospital with a principal diagnosis of AMI (ICD-9-CM codes 410.xx except for 410.x2) and with a complete claims history for the 12 months prior to admission. Patients who are transferred from one acute care facility to another must have a principal discharge diagnosis of AMI at both hospitals. The initial hospital for a transferred patient is designated as the responsible institution for the episode. If a patient has more than one AMI admission in a year, one hospitalization is randomly selected for inclusion in the measure. 2a.5 Target population gender: Female, Male 2a.6 Target population age range: The target population is age 65 years or older 2a.7 Denominator Time Window (<i>The time period in which cases are eligible for inclusion in the denominator</i>): This measure was developed with 12 months of data. Currently the measure is publicly-reported with three	2a- spec s C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>

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).

years of index hospitalizations.

2a.8 Denominator Details (All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions):

The denominator includes patients aged 65 and older admitted to non-federal acute care hospitals for an AMI defined by a principal discharge diagnosis of ICD-9-CM code 410.xx, excluding those with 410.x2 (AMI, subsequent episode of care) and with a complete claims history for the 12 months prior to admission.

ICD-9-CM codes that define the patient cohort:

- 410.00 AMI (anterolateral wall) - episode of care unspecified
- 410.01 AMI (anterolateral wall) - initial episode of care
- 410.10 AMI (other anterior wall) - episode of care unspecified
- 410.11 AMI (other anterior wall) - initial episode of care
- 410.20 AMI (inferolateral wall) - episode of care unspecified
- 410.21 AMI (inferolateral wall) - initial episode of care
- 410.30 AMI (inferoposterior wall) - episode of care unspecified
- 410.31 AMI (inferoposterior wall) - initial episode of care
- 410.40 AMI (other inferior wall) - episode of care unspecified
- 410.41 AMI (other inferior wall) - initial episode of care
- 410.50 AMI (other lateral wall) - episode of care unspecified
- 410.51 AMI (other lateral wall) - initial episode of care
- 410.60 AMI (true posterior wall) - episode of care unspecified
- 410.61 AMI (true posterior wall) - initial episode of care
- 410.70 AMI (subendocardial) - episode of care unspecified
- 410.71 AMI (subendocardial) - initial episode of care
- 410.80 AMI (other specified site) - episode of care unspecified
- 410.81 AMI (other specified site) - initial episode of care
- 410.90 AMI (unspecified site) - episode of care unspecified
- 410.91 AMI (unspecified site) - initial episode of care

Note: We do not include 410.x2 (AMI, subsequent episode of care)

2a.9 Denominator Exclusions (Brief text description of exclusions from the target population): The measures exclude admissions for patients:

- who were discharged on the day of admission or the following day and did not die or get transferred (because it is less likely they had a significant AMI);
- who were transferred from another acute care hospital (because the death is attributed to the hospital where the patient was initially admitted);
- with inconsistent or unknown mortality status or other unreliable data (e.g. date of death precedes admission date);
- enrolled in the Medicare Hospice program any time in the 12 months prior to the index hospitalization including the first day of the index admission (since it is likely these patients are continuing to seek comfort measures only);
- who were discharged alive and against medical advice (AMA) (because providers did not have the opportunity to deliver full care and prepare the patient for discharge);
- that were not the first hospitalization in the 30 days prior to a patient's death. We use this criteria to prevent attribution of a death to two admissions.

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

See "Denominator Exclusions" section.

2a.11 Stratification Details/Variables (All information required to stratify the measure including the stratification variables, all codes, logic, and definitions):

Results of this measure will not be stratified.

2a.12-13 Risk Adjustment Type: Risk-adjustment devised specifically for this measure/condition

2a.14 Risk Adjustment Methodology/Variables (List risk adjustment variables and describe conceptual

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.

models, statistical models, or other aspects of model or method):

Our approach to risk adjustment was tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

The measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital level 30-day RSMR. This approach to modeling appropriately accounts for the structure of the data (patients clustered within hospitals), the underlying risk due to patients' comorbidities, and sample size at a given hospital when estimating hospital mortality rates. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand et al., 2007). At the patient level, each model adjusts the log-odds of mortality within 30-days of admission for age, sex, selected clinical covariates and a hospital-specific intercept. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept, or hospital specific effect, represents the hospital contribution to the risk of mortality, after accounting for patient risk and sample size, and can be inferred as a measure of quality. The hospital-specific intercepts are given a distribution in order to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables: Candidate variables were patient-level risk-adjustors that are expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity. For each patient, covariates were obtained from Medicare claims extending 12 months prior to and including the index admission. The model adjusted for case differences based on the clinical status of the patient at the time of admission. We used condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes. We did not risk-adjust for CCs that were possible adverse events of care and that were only recorded in the index admission. In addition, only comorbidities that conveyed information about the patient at that time or in the 12-months prior, and not complications that arose during the course of the hospitalization were included in the risk-adjustment.

The final set of risk-adjustment variables are:

Demographic

- Age-65 (years above 65, continuous)
- Male

Cardiovascular

- History of PTCA
- History of CABG
- Congestive heart failure
- History of AMI
- Unstable angina
- Anterior myocardial infarction
- Other location of myocardial infarction
- Chronic atherosclerosis
- Cardio-respiratory failure and shock
- Valvular and rheumatic heart disease

Comorbidity

- Hypertension
- Stroke
- Cerebrovascular disease
- Renal failure
- Chronic Obstructive Pulmonary Disease
- Pneumonia
- Diabetes and DM complications

- Protein-calorie malnutrition
- Dementia and senility
- Hemiplegia, paraplegia, paralysis, functional disability
- Peripheral vascular disease
- Metastatic cancer, acute leukemia and other severe cancers
- Trauma in the last year
- Major psychiatric disorders
- Chronic liver disease

--
 References:
 Krumholz HM, Brindis RG, Brush JE, et al. 2006. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation* 113: 456-462.

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. *Stat Sci* 22 (2): 206-226.

2a.15-17 Detailed risk model available Web page URL or attachment: URL N/A
<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1163010421830>

2a.18-19 Type of Score: Rate/proportion
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*):
 The RSMR is calculated as the ratio of the number of "predicted" to the number of "expected" deaths, multiplied by the national unadjusted mortality rate. For each hospital, the "numerator" of the ratio is the number of deaths within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the "denominator" is the number of deaths expected on the basis of the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital's performance given its case-mix to an average hospital's performance with the same case-mix. Thus a lower ratio indicates lower-than-expected mortality or better quality and a higher ratio indicates higher-than-expected mortality or worse quality.

The predicted hospital outcome (the numerator) is calculated by regressing the risk factors and the hospital-specific intercept on the risk of mortality, multiplying the estimated regression coefficients by the patient characteristics in the hospital, transforming, and then summing over all patients attributed to the hospital to get a value. The expected number of deaths (the denominator) is obtained by regressing the risk factors and a common intercept on the mortality outcome using all hospitals in our sample, multiplying the subsequent estimated regression coefficients by the patient characteristics observed in the hospital, transforming, and then summing over all patients in the hospital to get a value.

To assess hospital performance in any reporting period, the model coefficients are re-estimated using the years of data in that period.

2a.22 Describe the method for discriminating performance (*e.g., significance testing*):
 CMS currently estimates an interval estimate for each risk-standardized rate to characterize the amount of uncertainty associated with the rate, compares the interval estimate to the national crude rate for the outcome, and categorizes hospitals as "better than," "worse than," or "no different than" the US national rate.

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):*
 N/A -This measure is not based on a sample or survey.

2a.24 Data Source (*Check the source(s) for which the measure is specified and tested*)
 Electronic administrative data/claims

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.*):

Two data sources were used to create the measure:

1. Medicare Part A Inpatient and Outpatient and Part B outpatient claims: This database contains claims data for fee-for-service inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, and hospice care, as well as inpatient and outpatient claims for the 12 months prior to an index admission.

2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This dataset was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming Fisher et al., 1992).

The measure was originally developed with claims data from 1998. The models have been maintained and re-evaluated each year since public reporting of the measures began in 2007.

Fleming C., Fisher ES, Chang CH, Bubolz D, Malenda J. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs Hospitals. *Medical Care*. 1992; 30(5): 377-91.

2a.26-28 Data source/data collection instrument reference web page URL or attachment:

2a.29-31 Data dictionary/code table web page URL or attachment: URL N/A Condition Category/ICD-9 Code Map available at: (<http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1182785083979>)

2a.32-35 Level of Measurement/Analysis (*Check the level(s) for which the measure is specified and tested*) Facility/Agency

2a.36-37 Care Settings (*Check the setting(s) for which the measure is specified and tested*) Hospital

2a.38-41 Clinical Services (*Healthcare services being measured, check all that apply*)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample (*description of data/sample and size*): The model was developed in a randomly selected 50% of patients in the initial one-year cohort and tested in the other 50% of patients in the initial one-year cohort. In each subsequent year of measure maintenance we recreated the cohorts in the same way or with very a little modification. The developmental cohort consisted of 134,661 cases discharged from 4,646 hospitals. The validation sample consisted of 199,978 cases discharged from 4,668 hospitals. Further validation was conducted in additional years.

Reference:
Krumholz HM, Normand S-LT, Galusha DH, Mattera JA, Rich AS, Wang YF, Wang Y. et al. Risk-Adjustment Models for AMI and HF: 30-Day Mortality: Report prepared for the Centers for Medicare & Medicaid Services; 2005. Available at: <http://www.qualitynet.org/>

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

For all cohorts, we computed diagnostics that describe their respective performance in terms of discriminative ability, overall fit, model coefficients, and generated hospital RSMRs and corresponding interval estimates for the cohort. With all this information, we can compare the changes over time as well as the performance with the model in the development cohort.

2b
C
P
M
N

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.

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.

<p>2b.3 Testing Results (<i>reliability statistics, assessment of adequacy in the context of norms for the test conducted</i>): See results under "Risk-Adjustment Strategy" below.</p>	
<p>2c. Validity testing</p> <p>2c.1 Data/sample (<i>description of data/sample and size</i>): Medical-record validation: For the derivation of the chart-based model, we used cases identified through the Health Care Financing Administration (now CMS) Cooperative Cardiovascular Project (CCP) initiative, which included more than 200,000 admissions to non-governmental, acute care hospitals in the United States and Puerto Rico. In the CCP study, CMS sampled all claims from fee-for-service Medicare patients during an approximately 8-month period (varying by state) in 1994 and 1995 who were discharged with a principal diagnosis of AMI (ICD-9-CM code 410, excluding 410.x2). These patients were matched to the Medicare enrollment database to determine survival and, where applicable, the date of death. Corresponding medical records were abstracted by 2 clinical data abstraction centers (DynKePRO [York, PA] and FMAS Corporation [Rockville, MD]), and the clinical data used to confirm the diagnosis of AMI.</p> <p>2c.2 Analytic Method (<i>type of validity & rationale, method for testing</i>): Medical-record validation: We developed a medical record measure to compare with the administrative measure. We defined the measure cohort for the medical record model using the same inclusion/exclusion criteria consistent with the claims-based administrative measure but using chart-based risk adjusters, such as blood pressure, not available in the claims data. We then matched a sample of the same patients in the administrative data for comparison. The sample included 181,032 patients. Lastly we examined the model performance and produced the hospital RSMR based on both models for comparison.</p> <p>2c.3 Testing Results (<i>statistical results, assessment of adequacy in the context of norms for the test conducted</i>): The mortality medical record model had a c-statistic of 0.77 as compare to 0.69 for the claims model. The correlation coefficient between hospital RSMR from medical record model and hospital RSMR from claims model was 0.90 indicating good consistency of the two models.</p> <p>Reference: Krumholz HM, Normand S-LT, Galusha DH, Mattera JA, Rich AS, Wang YF, Wang Y. et al. Risk-Adjustment Models for AMI and HF: 30-Day Mortality: Report prepared for the Centers for Medicare & Medicaid Services; 2005. Available at: http://www.qualitynet.org/</p> <p>Reference: Krumholz HM, Wang Y, Mattera JA, Wang YF, Han L, Ingber M, Roman S and Normand S-LT, An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-day Mortality Rate Among Patients with An Acute Myocardial Infarction. <i>Circulation</i>, 2006,113:1683-1692.</p>	<p>2c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>2d. Exclusions Justified</p> <p>2d.1 Summary of Evidence supporting exclusion(s): Rationale for exclusions described in "Denominator Exclusions"</p> <p>2d.2 Citations for Evidence: See "Denominator Exclusions"</p> <p>2d.3 Data/sample (<i>description of data/sample and size</i>): N/A</p> <p>2d.4 Analytic Method (<i>type analysis & rationale</i>): N/A</p> <p>2d.5 Testing Results (<i>e.g., frequency, variability, sensitivity analyses</i>): N/A</p>	<p>2d C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>2e. Risk Adjustment for Outcomes/ Resource Use Measures</p> <p>2e.1 Data/sample (<i>description of data/sample and size</i>): Prior years of data from Medicare Part A inpatient</p>	<p>2e C <input type="checkbox"/> P <input type="checkbox"/></p>

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) 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.

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
 •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).

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 outcome (but not disparities in care) and are present at start of care.
 OR
 rationale/data support no risk adjustment.

and outpatient data and Part B outpatient data are used to identify variables for risk-adjustment. Specifically, Medicare Part A inpatient data is used to identify variables for risk adjustment in the index admission. Part A and B outpatient data are used to identify variables for risk adjustment in the 12-month period preceding the index date of admission. The cohorts are as described above in Reliability Testing Data Sample.

M
 N
 NA

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

This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSMRs. (See "risk adjustment methodology" for additional details.)

Approach to assessing model performance:

During measure development, we computed five summary statistics for assessing model performance (Harrell, 2001) for the development and validation cohort:

- (1) over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)
- (2) predictive ability
- (3) area under the receiver operating characteristic (ROC) curve
- (4) distribution of residuals
- (5) model chi-square (A test of statistical significance usually employed for categorical data to determine whether there is a good fit between the observed data and expected values; i.e., whether the differences between observed and expected values are attributable to true differences in characteristics or instead the result of chance variation).

--
 F.E. Harrell and Y.C.T. Shih, Using full probability models to compute probabilities of actual interest to decision makers, *Int. J. Technol. Assess. Health Care* 17 (2001), pp. 17-26.

2e.3 Testing Results (risk model performance metrics):

During measure development, we tested the performance of the model developed in a randomly selected half of the 1998 hospitalizations for AMI (representing 199,978 cases discharged from the 4,668 hospitals) with hospitalizations from the other half. The performance was not substantively different in the validation sample (ROC area = 0.70) compared to the development cohort (ROC area=0.71). Further validation was done in additional years of data and these results were consistent with the development cohort.

For the development cohort, the results are summarized below:

Residuals lack of fit: <-2 = 0.00%; [-2, 0) = 81.92%; [0, 2) = 10.21%; [2+ = 7.84%
 Model Chi-square [# of covariates]: 9370 [27]
 Predictive ability (lowest decile %, highest decile %): (4.0, 40.0)
 Area under the ROC curve = 0.71

For the validation cohort, the results are summarized below:

Residuals lack of fit: <-2 = 0.00%; [-2, 0) = 81.92%; [0, 2) = 10.22%; [2+ = 7.85%
 Model Chi-square [# of covariates]: 9125 [27]
 Predictive ability (lowest decile %, highest decile %): (4.2, 40.1)
 Area under the ROC curve = 0.70

During the subsequent years of annual maintenance including the 2010 maintenance update, to test for reliability, we looked at the distributions of comorbid conditions, hospital volume, crude rates, hospital RSMR, risk-adjusted odds ratios and 95% confidence intervals, and between-hospital variance over different time periods during yearly maintenance updates and the parameters were consistent. For example, for the 2006-2008 calendar year dataset, we reported each individual year results as well as the 3-year combined results. Model performance was stable over all time periods; ROC=0.72 across all times periods.

Reference: Krumholz HM, Normand S-LT, Galusha DH, Mattera JA, Rich AS, Wang YF, Wang Y. et al. Risk-Adjustment Models for AMI and HF: 30-Day Mortality: Report prepared for the Centers for Medicare & Medicaid Services; 2005. Available at: <http://www.qualitynet.org/>

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 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.

<p>Bernheim SM, et al. 2010 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure and Pneumonia 30-day Risk Standardized Mortality Rate. 2010 Available at: http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic/Page/QnetTier3&cid=1163010421830</p> <p>2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: N/A—The measure is risk-adjusted</p>	
<p>2f. Identification of Meaningful Differences in Performance</p> <p>2f.1 Data/sample from Testing or Current Use (<i>description of data/sample and size</i>): This data below is based on RSMRs calculated for AMI hospitalizations from July 1, 2006- June 30, 2009 and includes 558,665 hospitalizations from 4569 hospitals. The index hospitalizations are those included in the measure and reported in the 2010 update to hospital compare.</p> <p>2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (<i>type of analysis & rationale</i>): For each RSMR, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national crude mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare Web site as "better than the U.S. national rate" or "worse than the U.S. national rate." If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or "the difference is uncertain." CMS also reports does not classify performance for hospitals that have fewer than 25 AMI cases in the three-year period.</p> <p>2f.3 Provide Measure Scores from Testing or Current Use (<i>description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance</i>): Recent analyses show substantial variation in RSMRs among hospitals. For the most recently reported three years of data (7/2006-6/2009) the mean hospital RSMR was 15.9% with a range of 10.3% to 24.6%. The 5th percentile was 13.2 and the 95th was 18.4. The interquartile range was 15.0% to 16.8%.</p> <p>These results also demonstrated ongoing geographic variation in hospital RSMRs for AMI.</p> <p>Reference: Bernheim SM, Grady JN, Lin Z, Wang Y, Wang Y, Savage SV, Bhat KR, Ross JS, Desai MM, Merrill AR, Han LF, Rapp MT, Drye EE, Normand SL, Krumholz HM. National patterns of risk-standardized mortality and readmission for acute myocardial infarction and heart failure. Update on publicly reported outcomes measures based on the 2010 release. <i>Circ Cardiovasc Qual Outcomes</i>. 2010 Sep 1;3(5):459-67. Epub 2010 Aug 24.</p>	<p>2f C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>2g. Comparability of Multiple Data Sources/Methods</p> <p>2g.1 Data/sample (<i>description of data/sample and size</i>): No current comparable data source was available that has complete data for a nationally representative sample.</p> <p>2g.2 Analytic Method (<i>type of analysis & rationale</i>): N/A</p> <p>2g.3 Testing Results (<i>e.g., correlation statistics, comparison of rankings</i>): N/A</p>	<p>2g C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>2h. Disparities in Care</p> <p>2h.1 If measure is stratified, provide stratified results (<i>scores by stratified categories/cohorts</i>): N/A - Measure is not stratified</p> <p>2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: Disparities in race and socioeconomic status (SES) have been reported at the patient level but our analyses</p>	<p>2h C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>

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 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.

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

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.

indicate little hospital-level disparities.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Scientific Acceptability of Measure Properties</i> ?	2
Steering Committee: Overall, to what extent was the criterion, <i>Scientific Acceptability of Measure Properties</i> , met? Rationale:	2 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
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 Ratin g
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) (<i>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</i>): The measure has been publicly reported on Hospital Compare since June 2007. Used in CMS's Hospital Inpatient Quality Reporting Program (Formerly RHQDAPU). The measure is reported on Hospital Compare, www.hospitalcompare.hhs.gov .	
3a.3 If used in other programs/initiatives (<i>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</i>):	
Testing of Interpretability (<i>Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement</i>)	
3a.4 Data/sample (<i>description of data/sample and size</i>):	
3a.5 Methods (<i>e.g., focus group, survey, QI project</i>): This measure is NQF endorsed. Prior to public reporting in 2007, CMS conducted a dry run in Dec 2006 to provide hospitals and the public with an opportunity to preview the measure methodology, proposed information for public reporting and hospital-specific information. Additionally, CMS has also conducted consumer testing of the language on Hospital Compare to ensure clarity and ease of interpretation of the information to be posted publicly.	3a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
3a.6 Results (<i>qualitative and/or quantitative results and conclusions</i>):	
3b/3c. Relation to other NQF-endorsed measures	
3b.1 NQF # and Title of similar or related measures: NQF # 0229 - Heart Failure 30-day Mortality ; NQF # 0468 - Pneumonia (PN) 30-Day Mortality Rate	
(for NQF staff use) Notes on similar/related endorsed or submitted measures:	
3b. Harmonization	3b
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):	C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/>
3b.2 Are the measure specifications harmonized? If not, why? Yes, the use a similar risk-adjustment strategy.	
3c. Distinctive or Additive Value	3c
3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-	C <input type="checkbox"/>

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.

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).

<p>endorsed measures: This measure looks at a different condition for the mortality outcome, AMI, from the two other related mortality measures.</p> <p>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: AHRQ inpatient AMI mortality measure. Inpatient mortality rates can be influenced by hospital length of stay, thus 30-day measures, that establish a standard follow-up period are more appropriate for profiling diverse group of hospitals.</p>	<p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Usability</i> ?	3
Steering Committee: Overall, to what extent was the criterion, <i>Usability</i> , met? Rationale:	<p>3</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
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
<p>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)</p>	<p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
4b. Electronic Sources	
<p>4b.1 Are all the data elements available electronically? (<i>elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims</i>) Yes</p> <p>4b.2 If not, specify the near-term path to achieve electronic capture by most providers.</p>	<p>4b</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
4c. Exclusions	4c
<p>4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications? No</p> <p>4c.2 If yes, provide justification.</p>	<p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences	
<p>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. Using administrative claims variables for risk adjustment: This measure uses variables from claims data submitted by hospitals to CMS for payment as clinical risk adjusters. Our analyses have demonstrated that administrative claims data can be used to develop risk-adjusted outcomes measures for mortality following admission for AMI and that the model produced estimates of RSMRs that are very similar to rates estimated by models based on chart data. This high level of agreement in the results based on the two different approaches supports the use of the claims-based models for public reporting. The models have also demonstrated consistent performance across years of claims data.</p> <p>The approach to gathering risk factors for patients also mitigates the potential limitations of claims data. Because not every diagnosis is coded at every visit, we use inpatient, outpatient, and physician claims data for the year prior to admission, and diagnosis codes during the index admission, for risk adjustment. This</p>	<p>4d</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>

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.

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

time frame provides a more comprehensive view of patients' medical histories than is provided by the secondary diagnosis codes from the index hospitalization alone. If a diagnosis appears in some visits and not others, it is included, minimizing the effect of incomplete coding. We were careful, however, to include information about each patient's status at admission and not to adjust for possible complications of the admission. Although some codes, by definition, represent conditions that are present before admission (e.g. cancer), other codes and conditions cannot be differentiated from complications during the hospitalization (e.g. infection or shock). If these are secondary diagnoses from the index admission, then they are not adjusted for in the analysis.	
4e. Data Collection Strategy/Implementation	
<p>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: N/A</p> <p>4e.2 Costs to implement the measure (<i>costs of data collection, fees associated with proprietary measures</i>): The measure is developed using administrative claims data and does not necessitate any additional cost/burden on hospitals.</p> <p>4e.3 Evidence for costs: N/A</p> <p>4e.4 Business case documentation: N/A</p>	<p>4e</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Feasibility</i>?	4
Steering Committee: Overall, to what extent was the criterion, <i>Feasibility</i> , met? Rationale:	<p>4</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
RECOMMENDATION	
(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.	<p>Time-limited</p> <p><input type="checkbox"/></p>
Steering Committee: Do you recommend for endorsement? Comments:	<p>Y <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>A <input type="checkbox"/></p>
CONTACT INFORMATION	
<p>Co.1 Measure Steward (Intellectual Property Owner)</p> <p>Co.1 <u>Organization</u> Centers for Medicare & Medicaid Services (CMS), 7500 Security Boulevard , Mail Stop S3-02-01, Baltimore, Maryland, 21244-9045</p> <p>Co.2 <u>Point of Contact</u> Lein, Han, PhD, Government Task Leader, lein.han@cms.hhs.gov, 410-786-0205-</p> <p>Measure Developer If different from Measure Steward</p> <p>Co.3 <u>Organization</u> Yale New Haven Health Services Corporation (YNHSC), 1 Church Street, Suite 200, New Haven, Connecticut, 06510</p> <p>Co.4 <u>Point of Contact</u> Susannah, Bernheim, MD, MHS, susannah.bernheim@yale.edu, 203-764-3271-</p>	

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).

<p>Co.5 Submitter If different from Measure Steward POC Susannah, Bernheim, MD, MHS, susannah.bernheim@yale.edu, 410-764-7231-, YNHHS</p>
<p>Co.6 Additional organizations that sponsored/participated in measure development MPR-Mathematica Policy Research; RTI-Research Triangle Institute</p>
<p>ADDITIONAL INFORMATION</p>
<p>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. The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org</p>
<p>Ad.2 If adapted, provide name of original measure: Acute Myocardial Infarction 30-day Mortality Ad.3-5 If adapted, provide original specifications URL or attachment URL www.qualitynet.org</p>
<p>Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 Year the measure was first released: 2007 Ad.7 Month and Year of most recent revision: 03, 2010 Ad.8 What is your frequency for review/update of this measure? Yearly Ad.9 When is the next scheduled review/update for this measure? 06, 2011</p>
<p>Ad.10 Copyright statement/disclaimers: N/A</p>
<p>Ad.11 -13 Additional Information web page URL or attachment: URL N/A www.qualitynet.org for Measure Methodology report and Maintenance reports</p>
<p>Date of Submission (MM/DD/YY): 10/28/2010</p>

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:
 - o Intermediate outcome - evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
 - o Process - 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).
 - o Structure - 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.
 - o Patient 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.
 - o Access - evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
 - o Efficiency - 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.

Circulation

JOURNAL OF THE AMERICAN HEART ASSOCIATION



An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With an Acute Myocardial Infarction

Harlan M. Krumholz, Yun Wang, Jennifer A. Mattera, Yongfei Wang, Lein Fang Han, Melvin J. Ingber, Sheila Roman and Sharon-Lise T. Normand

Circulation 2006;113:1683-1692; originally published online Mar 20, 2006;

DOI: 10.1161/CIRCULATIONAHA.105.611186

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214

Copyright © 2006 American Heart Association. All rights reserved. Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://circ.ahajournals.org/cgi/content/full/113/13/1683>

Subscriptions: Information about subscribing to *Circulation* is online at
<http://circ.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax: 410-528-8550. E-mail:
journalpermissions@lww.com

Reprints: Information about reprints can be found online at
<http://www.lww.com/reprints>

An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With an Acute Myocardial Infarction

Harlan M. Krumholz, MD, SM; Yun Wang, PhD; Jennifer A. Mattera, MPH; Yongfei Wang, MS; Lein Fang Han, PhD; Melvin J. Ingber, PhD; Sheila Roman, MD, MPH; Sharon-Lise T. Normand, PhD

Background—A model using administrative claims data that is suitable for profiling hospital performance for acute myocardial infarction would be useful in quality assessment and improvement efforts. We sought to develop a hierarchical regression model using Medicare claims data that produces hospital risk-standardized 30-day mortality rates and to validate the hospital estimates against those derived from a medical record model.

Methods and Results—For hospital estimates derived from claims data, we developed a derivation model using 140 120 cases discharged from 4664 hospitals in 1998. For the comparison of models from claims data and medical record data, we used the Cooperative Cardiovascular Project database. To determine the stability of the model over time, we used annual Medicare cohorts discharged in 1995, 1997, and 1999–2001. The final model included 27 variables and had an area under the receiver operating characteristic curve of 0.71. In a comparison of the risk-standardized hospital mortality rates from the claims model with those of the medical record model, the correlation coefficient was 0.90 (SE=0.003). The slope of the weighted regression line was 0.95 (SE=0.007), and the intercept was 0.008 (SE=0.001), both indicating strong agreement of the hospital estimates between the 2 data sources. The median difference between the claims-based hospital risk-standardized mortality rates and the chart-based rates was <0.001 (25th and 75th percentiles, –0.003 and 0.003). The performance of the model was stable over time.

Conclusions—This administrative claims-based model for profiling hospitals performs consistently over several years and produces estimates of risk-standardized mortality that are good surrogates for estimates from a medical record model. (*Circulation*. 2006;113:1683-1692.)

Key Words: health policy ■ quality of health care ■ myocardial infarction

Acute myocardial infarction (AMI), a common, high-risk event that requires timely intervention and extensive coordination among hospital clinicians and personnel, is the focus of several national efforts to improve quality of care.¹ The Centers for Medicare & Medicaid Services (CMS) and the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) publicly report 7 process measures for AMI, including the use of aspirin on admission and discharge, β -blockers on admission and discharge, angiotensin-converting enzyme inhibitors or angiotensin receptor blockers on discharge, time to reperfusion, and smoking cessation counseling.^{2,3} These measures convey important information about healthcare quality but focus on a narrow spectrum of the overall care provided to patients and thus explain a relatively small portion of the variation across hospitals in risk-adjusted mortality rates.⁴

Clinical Perspective p 1692

Outcomes, in contrast to structure and process, provide a more global assessment of healthcare quality and represent what matters most to patients.⁵ Process measurement is susceptible to the diversion of resources to what is being measured at the expense of what is not, potentially worsening overall quality of care and outcomes. Although outcomes are not entirely under the control of clinicians and hospitals, quality of care and safety can influence the risk of adverse events. Moreover, outcomes have the most relevance to patients.

Outcome measurement is challenging, however, because of variation among institutions in the risk profile of their patients.⁶ Statistical methods can adjust for observed differ-

Received December 28, 2005; revision received February 13, 2006; accepted February 16, 2006.

From the Section of Cardiovascular Medicine, Department of Medicine (H.M.K., Yongfei Wang), Section of Health Policy and Administration, Department of Epidemiology and Public Health (H.M.K.), and Robert Wood Johnson Clinical Scholars Program (H.M.K.), Yale University School of Medicine, New Haven, Conn; Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, Conn (H.M.K., Yun Wang, J.A.M.); Centers for Medicare & Medicaid Services, Baltimore, Md (L.F.H., M.J.I., S.R.); Department of Health Care Policy, Harvard Medical School, Boston, Mass (S.T.N.); and Department of Biostatistics, Harvard School of Public Health, Boston, Mass (S.T.N.).

Guest Editor for this article was Donna K. Arnett, PhD.

Correspondence to Dr Harlan M. Krumholz, Yale University School of Medicine, Room I-456 SHM, 333 Cedar St, PO Box 208088, New Haven, CT 06520-8088. E-mail harlan.krumholz@yale.edu

© 2006 American Heart Association, Inc.

Circulation is available at <http://www.circulationaha.org>

DOI: 10.1161/CIRCULATIONAHA.105.611186

ences in patient risk, but the only nationally available data on hospitalizations in the United States are Medicare claims, which do not always accurately reflect the information in the medical record.⁷ Thus, for a claims-based model to be suitable for public reporting, it should ideally be validated against a similar approach using medical record data.⁸ The validation should not be assessed at the patient level but instead should assess how well the characterization of hospital performance by administrative data compares with that of the model based on medical record data.

We developed a hierarchical regression model using Medicare claims data that produces hospital risk-adjusted 30-day mortality rates. We aligned our approach with a recent American Heart Association Scientific Statement that defined standards for statistical models used for the public reporting of health outcomes.⁸ This document recommends that such models be in the public domain, not misclassify complications as comorbidities, have a standardized period of outcome assessment, use statistical techniques that account for clustering of the data, and be validated with various cohorts and against the results of a model based on medical record data. We compared hospital risk-standardized mortality estimates derived from a claims-based model with rates determined from a model based on medical record data for 181 032 patients discharged from 4322 US hospitals from 1994 to 1996. To assess the stability of the model over time, we also assessed model performance in multiple years of Medicare claims data.

Methods

Derivation and Validation Cohorts

The Derivation Cohort

We randomly sampled half of the hospitalizations for AMI (*International Classification of Diseases, 9th Revision, Clinical Modification [ICD-9-CM]* codes 410.xx except for 410.x2) in the 1998 Medicare Provider Analysis and Review (MEDPAR) files, clustered within hospitals. For risk adjustment we used information in the MEDPAR files, physician files, and hospital outpatient files. The MEDPAR claims have data on each hospitalization for fee-for-service Medicare enrollees and include demographic information, principal and secondary diagnosis codes, and procedure codes. Diagnosis codes for comorbidities were also collected from physician and hospital outpatient files. These data were collected for the year before the index hospitalization.

We retained hospitalizations in which the patient was aged >64 years because these patients are representative of the older AMI population. We linked hospitalizations as an episode of care and attributed the outcome to the admitting hospital. To confirm the diagnosis, patients with AMI who were transferred from one facility to another were required to have a principal discharge diagnosis of AMI at both hospitals. For transferred patients, comorbid ("preexisting") conditions were identified from the initial (index) admission only so that these patients would not have the opportunity to have more conditions coded than the patients who were not transferred. We excluded patients with a total length of stay of ≤ 1 day and who were discharged alive and not against medical advice because it is unlikely that these patient suffered an AMI. We also excluded patients without 1 year of history in Medicare fee-for-service.

The Validation Cohorts

We constructed a linked sample that contains both claims and medical chart abstracted data for each patient. The medical record data were obtained from the Cooperative Cardiovascular Project (CCP), a national AMI quality improvement project in which

>200 000 medical records were abstracted.⁹ Hospitalizations for CCP were identified from hospital bills in the Medicare National Claims History File of claims submitted under fee-for-service. Hospitalizations for CCP occurred during an 8-month period between February 1994 and July 1995, except for the states in the pilot study (Alabama, Connecticut, Iowa, and Wisconsin), in which sampling took place during a 4-month period from August through November 1995. Predefined variables were abstracted from copies of hospital records. The reliability of the data was monitored by means of monthly reabstractions of randomly selected records; the accuracy of abstraction with respect to treatment variables was 95%.

To evaluate the stability of the claims model over time, we also evaluated the performance of the Medicare claims model using the other half of the 1998 MEDPAR data and data for each of years 1995, 1996, 1997, 1999, 2000, and 2001. In each case we constructed the sample in the same manner as for the derivation cohort.

Outcome

The primary outcome was hospital-specific risk-standardized all-cause 30-day mortality, defined as death from any cause 30 days after the index admission date. Mortality information was obtained from the Medicare enrollment files by linking unique patient identifiers.

Model Derivation: Patient Predictors of Mortality

Candidate variables for the Medicare claims model were developed primarily from the administrative diagnostic codes. Because there are >15 000 *ICD-9-CM* codes, we used the Hierarchical Condition Categories (HCC) to assemble clinically coherent codes into single variables.¹⁰ This system, which includes 189 categories, was developed by physician and statistical consultants under a contract to CMS and is publicly available. The HCC candidate variables considered for this model were derived from the secondary diagnosis and procedure codes from the index hospitalization (all the principal diagnoses were 410.xx, except for 410.x2) and from the principal and secondary diagnosis codes from hospitalizations, hospital outpatient visits, and outpatient office encounters in the 12 months before the index hospitalization.

We conducted a clinical review of the candidate variables to eliminate the secondary diagnoses from the index hospitalization that could have represented complications rather than conditions present on admission. For example, we did not include hemorrhage as a secondary diagnosis on the index admission because it may have been present on admission or occurred during the hospitalization. We combined categories of HCC variables on the basis of clinical judgment and bivariate associations and eliminated candidate variables with <1% frequency. Additional candidate variables, based on clinical judgment and a review of the literature, included demographic factors (age, sex), location of the AMI, and procedural factors (history of bypass surgery or percutaneous coronary intervention in the past year).

Model

Because of the natural clustering of the observations within hospitals, we estimated hierarchical generalized linear models (HGLM).¹¹⁻¹³ We modeled the log-odds of mortality within 30 days of admission as a function of patient demographic and clinical variables and a random hospital-specific effect. This strategy accounts for within-hospital correlation of the observed outcomes, separates within-hospital variation from between-hospital variation, and models the assumption that underlying differences in quality among the institutions lead to systematic differences among hospital outcomes. The covariates for the model were first selected with the use of a backward elimination procedure through the generalized linear model (GLM) with a logit link function approach. Because of the large number of patient observations that heavily influences probability values, we chose an exit criterion for a variable of $P > 0.01$. We also evaluated the full model compared with the model that contained only age and sex as covariates. We calculated the area under the receiver operating characteristic (ROC) curve and the percentage of explained variation (R^2). Finally, we reestimated the

regression coefficients of the covariates identified from our backward elimination strategy using a HGLM.

Model Validation

Medical Record Model

Candidate variables for the medical record model were selected on the basis of a literature review and clinical experience.^{4,14} Unlike the claims data, some covariates could be missing for patients in the sample. When there were missing data for a continuous-valued variable, we created an additional variable that assumed a value of 0 if the variable was measured and a value of 1 if missing, and set the value of the original variable to its mean when missing. This method of modeling missing data assumes that data are missing at random and permits inclusion of all available cases, although it is not as efficient as multiple imputation procedures. For discrete-valued variables, we included an additional level that indicated the variable was missing. We computed measures of model fit and discrimination for the medical record model similar to those computed for the claims-based models.

Hospital Risk-Standardized Mortality Rates

We calculated risk-standardized mortality rates for each hospital using the estimated hospital-specific parameters from the respective hierarchical models. These rates are obtained as the ratio of “predicted” to expected mortality, multiplied by the national unadjusted rate.¹⁵ Although other researchers have calculated the ratio of observed to expected outcomes, we use the predicted rates to avoid several analytic problems that have been cited.^{11,13,16} The expected outcome for each hospital is the number of 30-day deaths expected at the hospital if the hospital’s patients were treated at a “reference” hospital. Operationally this was accomplished by regressing the risk factors on the mortality with all hospitals in our sample, applying the subsequent estimated regression coefficients to the patient characteristics observed at the hospital, and then summing. This is a form of indirect standardization. The predicted hospital outcome is the number of expected mortalities at the “specific” hospital and not at a reference hospital. Operationally this was accomplished by estimating a hospital-specific random effect that represented baseline mortality risk for the hospital, applying the hospital-specific regression coefficients to the patient characteristics at the hospital, and then summing.

Using the 1994 to 1995 hospitalizations, we used 2 approaches to examine the relationship between the risk-standardized rates obtained from using administrative data and those using chart data. First, after creating a linked sample of admissions between the administrative claims data and the medical record data, we assessed the relationship between the risk-standardized mortality rates from the administrative claims model and from the chart model for each hospital through graphical and regression techniques. We estimated a linear regression equation describing the association between the 2 rates, weighting each hospital by the number of hospitalizations, and calculated the intercept and the slope of this equation. A slope close to 1 and an intercept close to 0 would provide evidence that the hospital rates from the 2 sources are very similar. Second, for each hospital we calculated the difference between the risk-standardized mortality rate based on the claims data and the medical record data and then summarized the distribution of these differences among the hospitals using the average, median, and maximum differences.

Stability of the Model Over Time

We validated the model over time by comparing its performance in the derivation set with various validation cohorts, as described above. To assess whether we included too many risk factors in our final model, we calculated indices that quantify overfitting. Specifically, we used the coefficients estimated from the derivation model to predict the log-odds of mortality in the validation cohorts. This was accomplished by multiplying the observed risk factors in each validation cohort and summing over the covariates for a subject to obtain a mortality score. Using these scores for each subject, we then estimated a logistic regression model in which the outcome was observed mortality and the single covariate was the risk score. The

intercept and slope obtained from this model are referred to as overfitting indices. If there is overfitting, we would expect the slopes to be different from 1 and the intercepts to be different from 0. We repeated this process for each validation data set, each time calculating a risk score using the regression estimates from our derivation model.

After assessing overfitting, we recalibrated the models in each of the validation data sets so that we used the same variables but reestimated the regression coefficients to the data for each specific cohort. We then calculated several indices for assessing model performance¹⁸: the area under the ROC curve, explained variation as measured by the generalized R^2 statistic, and the observed outcomes in strata defined by the lowest and highest deciles based on predictive probabilities. Model fit was further assessed through examination of Pearson residuals.

All analyses were conducted with the use of SAS version 8.02 (SAS Institute Inc, Cary, NC). Models were fitted separately to each year of data. The hierarchical models were estimated with the use of the GLIMMIX macro in SAS.

The authors had full access to the data and take responsibility for its integrity. All authors have read and agree to the manuscript as written.

Results

Patient Characteristics and Administrative Model: Derivation Sample

The 1998 sample included 387 081 AMI discharges from 4828 hospitals that were retrieved from the national fee-for-service administrative claims database, of which 8.3%, 10.7%, and 2.0% of the discharges were excluded for age <65 years, incomplete information in the 12 months before admission, and length of stay of ≤ 1 day, respectively (Table 1). Another 10.7% of the hospitalizations represented transfer in admission and were combined with the admission at the initial hospital to create an episode of care.

The derivation sample consisted of 140 120 cases with an unadjusted 30-day mortality rate of 18.0%. The mean age of the cohort was 77.8 ± 7.4 years. The cohort included 50.5% women and 9.7% nonwhite patients. There were 4664 hospitals in the derivation cohort, with a median annual number of Medicare AMI hospitalizations of 17 (25th and 75th percentiles, 6 and 40, respectively). The observed mortality rate ranged from 0.0% to 100.0% across these hospitals, and the 25th, 50th, and 75th percentiles were 13.1%, 16.9%, and 21.1%, respectively.

On the basis of a clinical review of the candidate variables, bivariate analysis, and stepwise GLM procedure, we identified 27 variables, including 2 demographic, 10 cardiovascular, and 15 comorbidity variables, for the final model. The model had good discrimination, calibration, and fit (Table 2). The area under the ROC curve was 0.71. Model discrimination was good, with the observed mortality rate ranging from 4.0% in the lowest predicted decile to 40.0% in the highest predicted decile, a range of 36.0%. The adjusted R^2 was 0.12. Figure 1A illustrates the overall distribution of risk-standardized 30-day mortality rates at the hospital level, and Figure 1B shows the distributions of risk-standardized 30-day mortality rates stratified by hospital volume. The 25th and 75th percentiles were 16.8% and 18.4%, respectively. The 95th percentile was 19.9%, and the 5th percentile was 15.5%.

The model that included only age and sex had worse model fit with an area under the ROC curve of 0.62. Model

TABLE 1. AMI Initial Administration Claims Sample

Data Source	Total	Exclusion, %				Final Sample	
		Age <65 y	Incomplete Information*	Transfer†	LOS ≤1 d‡	n	% of Total
1995	386 521	7.9	10.7	10.7	1.0	283 578	73.4
1996	393 845	8.1	10.9	10.9	1.3	285 953	72.6
1997	391 812	8.2	11.2	10.9	1.7	282 127	72.0
1998	387 081	8.3	10.7	10.7	2.0	280 098	72.4
1999	387 018	8.5	10.4	10.7	2.3	280 319	72.4
2000	346 595	8.3	5.2	11.6	2.6	263 124	75.9
2001	354 402	8.6	6.6	11.4	3.0	264 191	74.6

LOS indicates length of stay.

*Incomplete information in the 12-month, preindex admission period was excluded for the AMI sample.

†After linking the “transfer-in” hospital with the “transfer-out” (index admission) hospital, the records of the “transfer-in” hospital were deleted so that the case was assigned to the index admission hospital.

‡Discharged within first day of admission and alive, not against medical advice, not transferred.

discrimination was not as good: The observed mortality rate ranged from 10.2% in the lowest predicted decile to 31.0% in the highest predicted decile, a range of 20.8%. Explained variation, R^2 , was only 0.04.

The model that did not include institutional outpatient and physician data as a source of data had a lower prevalence of comorbid conditions. For example, diabetes was present in 32.9% on the basis of inpatient, outpatient, and physician data but was only 28.9% for inpatient alone. Hypertension was present in 46.8% on the basis of inpatient, outpatient, and physician data but was only 31.7% for inpatient alone. Overall, among 27 variables, except age, male gender, history of percutaneous transluminal coronary angioplasty, history of percutaneous coronary intervention, and location of AMI in the index admission, many variables were affected by physician data; the absolute percent change in frequency ranged from 0.17% (chronic liver disease) to 15.9% (trauma in last year), with a mean of $5.8 \pm 4.6\%$. However, the model using only inpatient data had model fit that was close to the model based on inpatient, outpatient, and physician data, with an area under the ROC curve of 0.70. The observed mortality rate ranged from 3.9% in the lowest predicted decile to 39.8% in the highest predicted decile, a range of 35.9%. The adjusted R^2 was 0.11.

Medical Record Model

The final CCP validation sample contained 181 032 hospitalizations and a crude 30-day mortality rate of 18.8%. In this cohort, the administrative model had an area under the ROC curve of 0.69, an observed mortality rate ranging from 5.3% in the lowest predicted decile to 38.6% in the highest predicted decile, and an adjusted R^2 of 0.10. The medical record comparison model in this cohort included 31 variables (Table 3). The area under the ROC curve was 0.77. The observed mortality rate ranged from 2.9% in the lowest predicted decile to 59.0% in the highest. Explained variation, R^2 , was 0.24. For all covariates that represented similar information in the 2 data sources, the respective estimated regression coefficients were in a similar direction. Bypass

surgery was positive in the medical record model and negative in the claims model, but the definitions were different: In the claims model there was a requirement for a billing code in the year before the AMI, whereas in the medical record model there was written documentation of bypass surgery at any time.

Comparison of Hospital Mortality Rates: Claims and Medical Record Data

The estimated hospital-specific standardized 30-day mortality rates derived from each model are displayed in Figure 2A and stratified by volume in Figure 2B. The slope of the weighted regression line is 0.95 (SE=0.007) and the intercept is 0.008 (SE=0.001), both indicating strong agreement of the hospital risk-standardized mortality estimates between the 2 data sources. The correlation coefficient of the standardized mortality rates from the 2 models is 0.90 (SE=0.003). The median difference between the hospital-specific risk-standardized mortality rates estimated from the claims data and those estimated from the medical record data was <0.001 (25th and 75th percentiles, -0.003 and 0.003, respectively; 10th and 90th percentiles, -0.007 and 0.007, respectively).

Model Performance in Administrative Validation Set

In each validation cohort the model fit was similar to that of the derivation cohort (Table 4). These comparisons spanned 7 years of Medicare admissions for AMI. The unadjusted mortality ranged from 18.1% to 19.0%. The percent explained variation ranged from 0.11 to 0.12, and the area under the ROC curve ranged from 0.69 to 0.71. The overfitting statistics were all within an acceptable range, indicating that we had not overfitted the models.

Discussion

This study introduces an administrative claims-based model for reporting hospital-specific 30-day AMI mortality rates for Medicare beneficiaries with output that is an excellent surrogate for that produced by a model based on medical record

TABLE 2. AMI Administrative Model Based on 1998 Derivation Sample

Predictor	Estimate	<i>t</i>	Odds Ratio	95% CI
Intercept	-2.61	-99.78		
Demographic				
Age, years over 65	0.05	49.45	1.05	1.05-1.05
Male	0.06	4.34	1.07	1.04-1.10
Cardiovascular				
History of PTCA	-0.52	-9.83	0.60	0.54-0.66
History of CABG	-0.10	-3.23	0.91	0.85-0.96
History of heart failure (HCC 80)	0.42	23.03	1.52	1.47-1.58
History of AMI (HCC 81)	-0.44	-22.72	0.65	0.62-0.67
Anterior MI (<i>ICD-9</i> 410.00-410.19)	0.64	34.74	1.89	1.83-1.96
Inferior/lateral/posterior MI (<i>ICD-9</i> 410.20-410.69)	0.44	24.49	1.56	1.51-1.62
Unstable angina (HCC 82)	-0.13	-6.14	0.88	0.84-0.91
Chronic atherosclerosis (HCC 83 and 84)	-0.41	-25.91	0.67	0.65-0.69
Cardiopulmonary-respiratory failure and shock (HCC 79)	0.30	10.95	1.35	1.28-1.42
Valvular heart disease (HCC 86)	0.09	4.29	1.09	1.05-1.13
Comorbidity				
Hypertension (HCC 89 and 91)	-0.13	-8.43	0.88	0.85-0.91
Stroke (HCC 95 and 96)	0.23	8.49	1.26	1.20-1.33
Cerebrovascular disease (HCC 97, 98, 99, 103)	0.08	3.76	1.09	1.04-1.13
Renal failure (HCC 131)	0.36	13.25	1.43	1.36-1.51
COPD (HCC 108)	0.15	8.66	1.16	1.12-1.20
Pneumonia (HCC 111, 112, 113)	0.13	5.85	1.14	1.09-1.19
Diabetes (HCC 15-20, 120)	0.24	15.59	1.28	1.24-1.32
Protein-calorie malnutrition (HCC 21)	0.51	13.27	1.66	1.54-1.79
Dementia (HCC 49-50)	0.41	18.64	1.51	1.45-1.58
Hemiplegia, paraplegia, paralysis, functional disability (HCC 100, 101, 102, 68, 69, 177, 178)	0.31	9.60	1.36	1.28-1.45
Peripheral vascular disease (HCC 104, 105)	0.21	10.62	1.23	1.18-1.28
Metastatic cancer (HCC 7, 8)	0.58	14.89	1.78	1.65-1.92
Trauma in last year (HCC 154-156, 158-162)	0.10	5.73	1.11	1.07-1.15
Major psychiatric disorders (HCC 54, 55, 56)	0.23	7.01	1.25	1.18-1.34
Chronic liver disease (HCC 25, 26, 27)	0.61	8.10	1.84	1.59-2.13

PTCA indicates percutaneous transluminal coronary angioplasty; CABG, coronary artery bypass graft surgery; and COPD, chronic obstructive pulmonary disease. Estimate of between-hospital variance=0.041 (SE=0.0032).

data. From prior work we know that at the patient level, medical record data are better for discriminating patients who survive and those who do not. In profiling institutions, however, the emphasis is on a measure that averages information for all patients within the hospital rather than on the specific agreement of individual variables or patient-level discrimination. Because the claims-based model and a medical record model classified hospitals similarly with respect to their standardized mortality rates, this approach, which has been endorsed by the National Quality Forum, may be suitable for the public reporting of hospital outcomes for patients with AMI. We note that our comparison between data sources focused on risk-standardized estimates; investigators who wish to use the results in different ways will need to undertake an assessment of the comparability of the 2 data sources for that purpose.

The development of this model included several methodological improvements on currently utilized administrative

data risk-adjustment models. The model was designed to include only diagnosis codes that indicate conditions present on admission, thus avoiding the problem of unwittingly crediting a hospital for more ill patients who just may have had more complications during the hospitalization, possibly as a result of worse care. We used clinical judgment to exclude secondary diagnosis codes in which complications could not be distinguished from preexisting conditions. In addition, we defined the outcome with a standardized period of follow-up rather than relying on the hospital stay that may vary by institution. In addition, we made use of healthcare utilization in the year before the index admission to improve the predictive ability of the model and minimize the potential for the gaming of codes at the admitting hospital.

A notable aspect of our approach is the validation achieved by comparing the output of the administrative claims model with that from a model based on medical record review data.

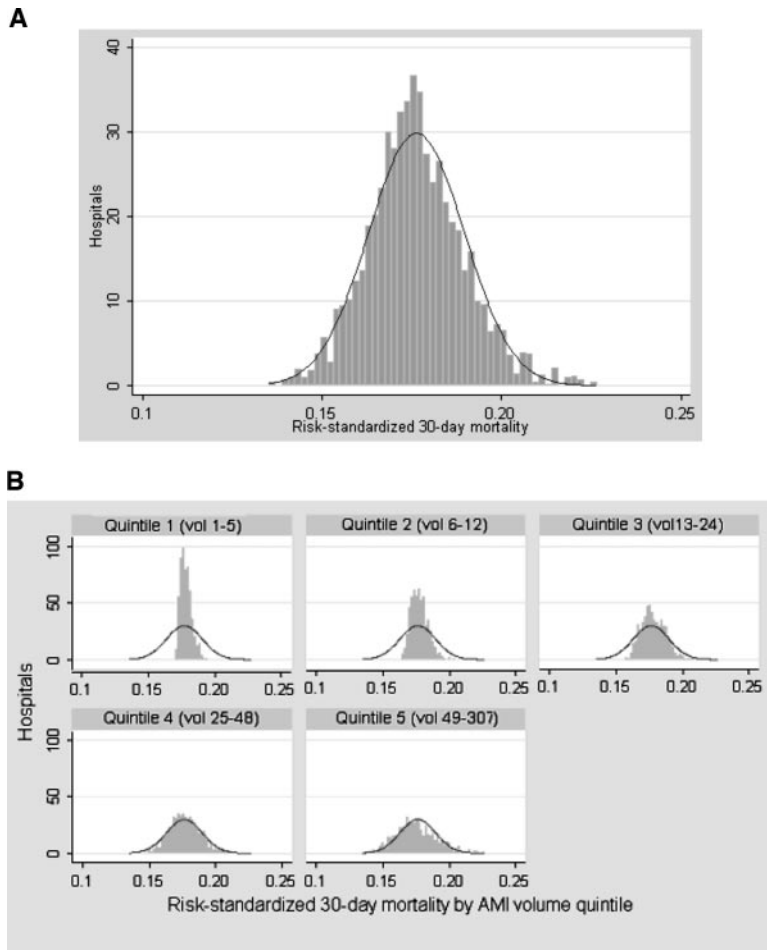


Figure 1. Distributions of risk-standardized 30-day AMI mortality rates overall (A) and stratified by volume (B).

Although even a model based on detailed information from medical records is not truly a gold standard, models based on this information currently provide the best opportunity to characterize the baseline risk of patients admitted to the hospital with an AMI. The Medicare claims data model did not perform as well as the medical record model at the patient level; however, the output of the models, ie, the profiling of the hospital performance, is the focus of this effort. We show that at the hospital level the administrative claims data, for all their limitations, can serve as a reasonable surrogate for the risk-standardized estimates from a model based on better data. The current cost of collecting medical record data precludes their collection as part of a national profiling effort. Therefore, the only current opportunity to develop a national hospital profiling effort can be based on administrative claims data. Our findings suggest that such an approach is possible.

We particularly sought an approach that could be released in the public domain. Many of the current publicly reported hospital profiling systems use proprietary approaches and do not provide information about model fit or validation against a gold standard.^{19,20} This model, including its methodology, covariates, and performance, can be posted and disseminated.

Our analytic approach used hierarchical modeling, which takes into account nesting of the data (ie, patients within hospitals). Patients within hospitals would be expected to

have a mortality risk that is more highly correlated than that of patients in different hospitals. This lack of independence of the observations can lead to underestimation of the SEs of risk factors and cause the appearance of statistically significant differences where none truly exist. In addition, sample sizes vary by hospital, and hierarchical modeling can take into account the differences in the amount of information provided by each hospital.^{12,21,22}

An important question is whether 30-day mortality is a suitable metric for the comparison of hospital performance. Outcomes are 1 of the 3 domains for quality measurement of Donabedian.²³ Some organizations are currently using this measure, although their methods are often obscure.²⁰ From the hospital perspective, there are concerns about whether the true profile of their patients' risk can be taken into account. Although the percentage of explained variability was low, the probability of discriminating survivors and nonsurvivors, the most common metric for assessing binary-valued outcomes, was 70% with the use of administrative data and 77% with the use of medical record data. In addition, the unexplained variation is a result of unmeasured risk factors, quality of care, and random variation. In the medical record model, we have included the risk factors that are considered most important for early mortality. It is possible that novel risk factors will be identified or that some other unmeasured risk factors might have added incrementally to the model, but it is

TABLE 3. AMI Medical Record Model Based on CCP Data Set (1994 to 1995)

Variable Definition	Estimate	<i>t</i>	Odds Ratio	95% CI
Intercept	-3.45	-77.30		
Demographics				
Male	-0.18	-12.61	0.84	0.81-0.86
Age, years over 65	0.05	50.29	1.05	1.05-1.05
Noncardiac history				
Diabetes (any type)	0.10	6.87	1.10	1.07-1.14
Cardiac history				
CVA/stroke	0.28	16.06	1.32	1.28-1.37
Myocardial infarction	-0.06	-3.72	0.95	0.92-0.97
Hypertension	-0.21	-15.25	0.81	0.79-0.83
COPD	0.05	3.25	1.05	1.02-1.09
Bypass surgery	0.12	5.53	1.12	1.08-1.17
Angioplasty	-0.35	-11.43	0.70	0.66-0.75
Cardiac symptoms (first 48 h of admission)				
Systolic blood pressure, mm Hg				
≤100	0.00		1.00	
>100	-0.99	-51.30	0.37	0.36-0.39
Missing	0.97	13.22	2.64	2.29-3.05
Shock	1.49	42.76	4.42	4.12-4.73
Heart failure/PE/CHF on x-ray/rales/gallop rhythm or S3	0.39	25.00	1.48	1.44-1.53
Time since chest pain started (relative to hospital arrival)				
<6 h	0.00		1.00	
6-12 h	0.16	6.19	1.17	1.11-1.23
>12 h	0.18	8.57	1.20	1.15-1.24
No chest pain	0.33	19.29	1.39	1.34-1.44
Unable to determine chest pain time	0.32	14.56	1.38	1.32-1.44
Initial laboratory results (first 24 h of admission)				
BUN, mg/dL*	0.01	14.16	1.01	1.01-1.01
BUN missing	0.27	4.63	1.31	1.17-1.47
Creatinine, mg/dL*	0.58	31.88	1.78	1.72-1.85
Creatinine missing	0.87	14.30	2.38	2.11-2.68
White blood cell count, $\mu\text{L} \times 1000$ (first 24 h of admission)				
<6			1.00	
6-12	0.29	9.23	1.34	1.26-1.43
>12	0.85	26.21	2.33	2.19-2.48
Missing	0.44	9.17	1.55	1.41-1.70
First ECG within 6 h before or after arrival				
ST-segment elevation	0.31	20.08	1.36	1.32-1.40
ECG unavailable	0.15	5.78	1.16	1.10-1.22
Left bundle-branch block	0.16	6.20	1.17	1.12-1.23
Right bundle-branch block	0.35	15.09	1.41	1.35-1.48
Second/third-degree heart block	0.28	5.55	1.32	1.20-1.46
Documented location(s) of AMI				
Anterior or lateral	0.56	37.64	1.75	1.70-1.81
No location determined	0.78	31.89	2.17	2.07-2.28

CVA indicates cerebrovascular accident; COPD, chronic obstructive pulmonary disease; PE, peripheral edema; CHF, congestive heart failure; and BUN, blood urea nitrogen. All $P < 0.0001$.

Estimate of between-hospital variance = 0.061 (SE = 0.0052).

*Mean, if not measured.

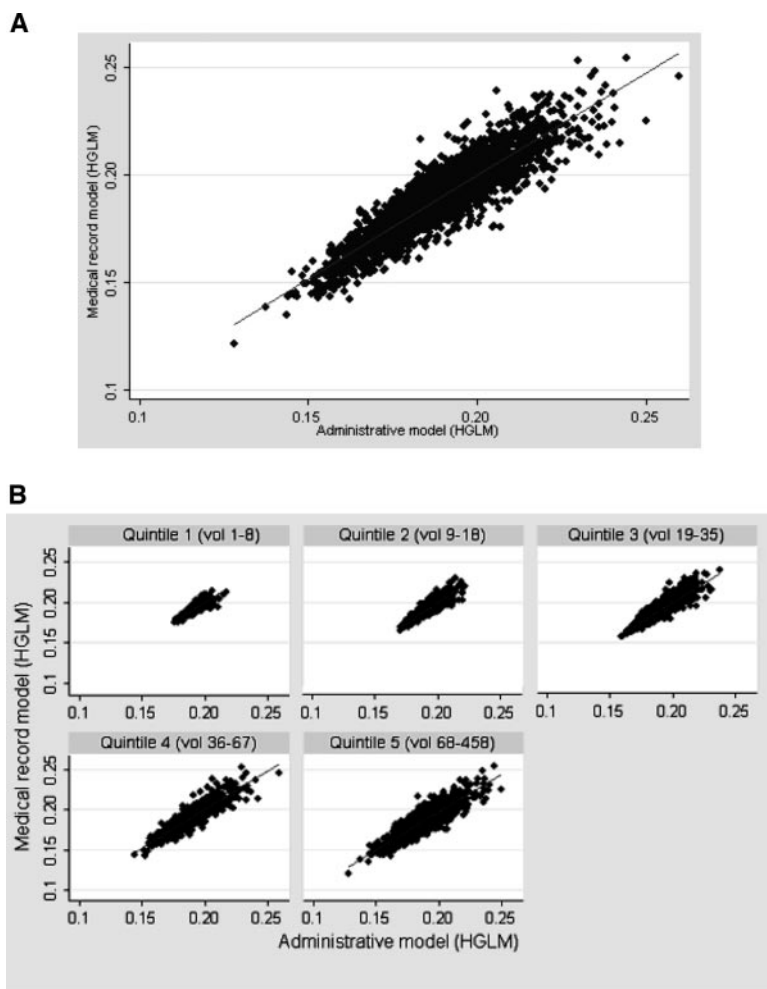


Figure 2. Comparison of the hospital-level risk-standardized mortality rates with the medical record model and the administrative model, overall (A) and stratified by volume (B).

unlikely that they would have markedly increased the explained variation. The value of outcome measurement over structure and process is that it represents what is most important to the patient. In addition, a focus on one process

can lead to ignoring other important processes that are not measured in an attempt to be viewed positively. Assuming that mortality is appropriate for profiling, we selected 30 days because it is a standardized outcome within a time frame that

TABLE 4. AMI Administrative Model and Medical Record Model Performance

Model	Over-Fitting Indices (Intercept, Slope)	Discrimination		
		Adjusted R^2 *	Predictive Ability† (Lowest Decile, Highest Decile)	ROC Curve Area
Administrative data derivation sample				
1998 (1st half) (n=140 120)	(0, 1)	0.12	4.0%–40.0%	0.71
Administrative data validation samples				
1998 (2nd half) (n=139 978)	(0.01, 0.99)	0.11	4.2%–40.1%	0.70
1995 (n=283 578)	(0.10, 0.99)	0.11	4.5%–39.3%	0.69
1996 (n=285 953)	(0.08, 1.00)	0.11	4.5%–39.1%	0.70
1997 (n=282 127)	(0.04, 1.001)	0.12	4.3%–39.0%	0.70
1999 (n=280 319)	(0.04, 1.01)	0.12	3.9%–40.6%	0.71
2000 (n=263 124)	(0.02, 1.00)	0.12	5.3%–40.6%	0.70
2001 (n=264 191)	(0.03, 1.00)	0.12	4.8%–41.1%	0.70
Linked administrative data model				
1994–1995 (n=181 032)	(0.02, 1.00)	0.10	5.3%–38.6%	0.69
Linked medical record data model				
1994–1995 (n=181 032)	(0, 1)	0.24	2.9%–59.0%	0.77

*Max-rescaled R^2 .

†Observed rates in deciles determined by estimated model.

hospital quality would expect to influence and has been commonly used in clinical trials.

We did not exclude hospitals with small volumes because the hierarchical model takes the sample size of each unit into account. In general, hospitals that provide little information (have small volumes) will have predicted risk-standardized mortality rates that are near the national average because these institutions do not provide sufficient information for an informed estimate of their performance. Because these small-volume hospitals would cluster at the average, they would not have a strong influence on the correlation between the 2 methods of calculating hospital performance. Moreover, no matter what analytic approach is adopted, it is difficult to calculate precise estimates of outcome rates for hospitals or other healthcare units with small volumes.

Our analysis has some issues to consider. This study is the first to validate a claims model in US hospitals using this approach. We required a large national sample of claims and medical record data. The data were only available in the CCP, a quality improvement initiative that abstracted records from 1994 to 1995. Nevertheless, even a chart-based model has limitations, and there is no true gold standard against which to compare the claims model. However, the chart-based model currently provides the best characterization of the demographic and clinical characteristics of the patients. Another issue to consider is that coding practices have changed since that time, as has the definition of AMI, although our evaluation of the patient-level performance of the model over time revealed that it did not change. Finally, although the output from the administrative data model was highly correlated with that from the medical record data, our findings should not deflect attention from the need to improve the quality of data available for profiling. We focused on the relationship between risk-standardized estimates derived from 2 data sources; if the primary goal is to identify high-quality hospitals, ie, those in the upper 10th percentile, then it will also be important to assess the sensitivity and specificity of the claims-based estimates for this activity.²⁴ Improving ICD codes holds great promise for enhancing our ability to track outcomes such as complications, to elevate risk-adjustment approaches, and to avoid manipulation of coding. Advances in electronic health records that would provide medical record data in digital format for use in profiling hospital performance will also advance our ability to adjust for hospital differences in case mix. The use of administrative data should only be a temporary solution as higher quality data become available through the adoption of electronic health records.

This model is specific to Medicare fee-for-service patients and may not be generalizable to other data sources and patient populations. Nevertheless, Medicare patients represent a majority of the patients hospitalized with an AMI. Within Medicare, the fee-for-service patients are the vast majority. This effort was limited by the availability of national claims and medical record data. However, the ability to profile institutions by the experience of the Medicare fee-for-service patients represents an advance. As better, more comprehensive databases become available for broader populations, it will be natural to extend this approach.

Growing interest in the public reporting of outcomes has focused attention on the need for models that can adjust for differences in case mix among hospitals. Only administrative claims data are widely available to perform these types of analyses. In this study we derived an administrative claims model that characterizes the performance of hospitals in a manner similar to that produced by the medical record model. Thus, this claims model can produce information about hospital performance that is comparable to that obtained from higher-quality data that are much more expensive to obtain.

Acknowledgments

The analyses on which this publication is based were performed under contract 500-05-CO01, entitled "Utilization and Quality Control Quality Improvement Organization for the State of Colorado," sponsored by CMS (formerly Health Care Financing Administration), Department of Health and Human Services. The content of the publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. The authors assume full responsibility for the accuracy and completeness of the ideas presented. This article is a direct result of the Health Care Quality Improvement Program initiated by CMS, which has encouraged identification of quality improvement projects derived from analysis of patterns of care, and therefore required no special funding on the part of this contractor. Ideas and contributions to the authors concerning experience in engaging with issues presented are welcomed. The authors thank Dr Jephtha Curtis, Dr JoAnne Foody, Dr Robert McNamara, Deron Galusha, and Amy Rich from the Yale University School of Medicine; Neil Gittings from CMS; and Debra Chromik from the Colorado Foundation for Medical Care for their contributions to this work. CMS reviewed and approved the use of its data for this work and approved submission of the manuscript.

Disclosures

Dr Krumholz is a consultant to United Healthcare. The other authors report no conflicts.

References

- Spertus JA, Radford MJ, Every NR, Ellerbeck EF, Peterson ED, Krumholz HM. Challenges and opportunities in quantifying the quality of care for acute myocardial infarction: summary from the Acute Myocardial Infarction Working Group of the American Heart Association/American College of Cardiology First Scientific Forum on Quality of Care and Outcomes Research in Cardiovascular Disease and Stroke. *Circulation*. 2003;107:1681-1691.
- Centers for Medicare and Medicaid Services. Available at: <http://www.cms.hhs.gov>. Accessed February 8, 2006.
- Joint Commission on Accreditation of Healthcare Organizations. Available at: <http://www.jcaho.org>. Accessed February 8, 2006.
- Normand ST, Glickman ME, Sharma RG, McNeil BJ. Using admission characteristics to predict short-term mortality from myocardial infarction in elderly patients: results from the Cooperative Cardiovascular Project. *JAMA*. 1996;275:1322-1328.
- Donabedian A. The role of outcomes in quality assessment and assurance. *QRB Qual Rev Bull*. 1992;11:356-360.
- Brook RH, Iezzoni LI, Jencks SF, Knaus WA, Krakauer H, Lohr KN, Moskowitz MA. Symposium: case-mix measurement and assessing quality of hospital care. *Health Care Financ Rev*. 1987;Spec No:39-48.
- Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. *Ann Intern Med*. 1993;119:844-850.
- Krumholz HM, Brindis RG, Brush JE, Cohen DJ, Epstein AJ, Furie K, Howard G, Peterson ED, Rathore SS, Smith SC, Spertus JA, Wang Y, Normand S-LT. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement

- from the Quality of Care and Outcomes Research Interdisciplinary Writing Group. *Circulation*. 2006;113:456–462.
9. Marciniak TA, Ellerbeck EF, Radford MJ, Kresowik TF, Gold JA, Krumholz HM, Kiefe CI, Allman RM, Vogel RA, Jencks SF. Improving the quality of care for Medicare patients with acute myocardial infarction: results from the Cooperative Cardiovascular Project. *JAMA*. 1998;279:1351–1357.
 10. Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Iezzoni LI, Ingber MJ, Levy JM, Robst J. Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev*. 2004;25:119–141.
 11. Normand SLT, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: issues and applications. *J Am Stat Assoc*. 1997;92:803–814.
 12. Shahian DM, Normand SL, Torchiana DF, Lewis SM, Pastore JO, Kuntz RE, Dreyer PI. Cardiac surgery report cards: comprehensive review and statistical critique. *Ann Thorac Surg*. 2001;72:2155–2168.
 13. Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical aspects of institutional performance. *J Royal Stat Soc*. 1996;159:385–444.
 14. Krumholz HM, Chen J, Wang Y, Radford MJ, Chen YT, Marciniak TA. Comparing AMI mortality among hospitals in patients 65 years of age and older: evaluating methods of risk adjustment. *Circulation*. 1999;99:2986–2992.
 15. Shahian DM, Torchiana DF, Shemin RJ, Rawn JD, Normand S-LT. The Massachusetts cardiac surgery report card: implications of statistical methodology. *Ann Thorac Surg*. 2005;80:2106–2113.
 16. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med*. 1997;127:764–768.
 17. Deleted in proof.
 18. Harrell FE. *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis*. 3rd ed. New York, NY: Springer; 2001.
 19. US News and World Report. Best hospitals 2005. Available at: <http://www.usnews.com/usnews/nycu/health/hosptl/tophosp.htm>. Accessed February 8, 2006.
 20. Healthgrades. Available at: www.healthgrades.com. Accessed February 8, 2006.
 21. Austin PC, Tu JV, Alter DA. Comparing hierarchical modeling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: should we be analyzing cardiovascular outcomes differently? *Am Heart J*. 2003;145:27–35.
 22. DeLong E. Hierarchical modeling: its time has come. *Am Heart J*. 2003;145:16–18.
 23. Donabedian A. The quality of care: how can it be assessed? *JAMA*. 1988;260:1743–1748.
 24. Normand S-LT, Wolf RE, Ayanian JZ, McNeil BJ. Assessing the accuracy of hospital clinical performance measures. *Med Decision Making*. In press.

CLINICAL PERSPECTIVE

Growing interest in the public reporting of outcomes has focused attention on the need for models that can adjust for differences in case mix among hospitals. Only administrative claims data are widely available to perform these types of analyses. A model using administrative claims data that is suitable for profiling hospital performance for acute myocardial infarction would be useful in quality assessment and improvement efforts. We developed a hierarchical regression model using Medicare claims data that produces hospital risk-standardized 30-day mortality rates that are similar to what can be derived from a medical record model. Thus, the results of this administrative model can be considered a surrogate for the results from the medical record model. This model has been endorsed by the National Quality Forum as a measure of hospital performance.

**2010 Measures Maintenance Technical Report:
Acute Myocardial Infarction, Heart Failure, and Pneumonia
30-Day Risk-Standardized Mortality Measures**

**Submitted By Yale New Haven Health Services Corporation/ Center for Outcomes
Research & Evaluation (YNHHSC/CORE)**

Susannah M. Bernheim, M.D., M.H.S.
Yun Wang, Ph.D.
Kanchana R. Bhat, M.P.H.
Shantal V. Savage, B.A.
Changqin Wang, M.D., M.S.**
Jacqueline N. Grady, M.S.
Yongfei Wang, M.S.**
Zhenqiu Lin, Ph.D.
Mayur M. Desai, Ph.D., M.P.H.*
Jinghong Gao, M.S.
Sharon-Lise T. Normand, Ph.D.***
Elizabeth E. Drye, M.D., S.M.**
Harlan M. Krumholz, M.D., S.M.**

*Yale School of Public Health

**Yale School of Medicine

***Harvard Medical School, Department of Health Care Policy and Harvard School of
Public Health, Department of Biostatistics

Under Contract # HHSM-500-2008-0025I-MIDS Task Order T0001
From the Centers for Medicare & Medicaid Services

Prepared For:

Centers for Medicare & Medicaid Services (CMS)

March 31, 2010

TABLE OF CONTENTS

LIST OF TABLES	3
LIST OF FIGURES.....	4
ACKNOWLEDGEMENTS.....	5
1. INTRODUCTION.....	6
1.1 Background on Mortality Measures	6
1.2 Overview of Measure Methodology	7
1.3 Goals of Measure Maintenance.....	10
2. UPDATES TO METHODS	12
2.1 Refinements to the Mortality Measures	12
2.1.1 Change in Collection of Claims History Data for Determination of Patient Comorbidities.....	12
2.1.2 Update to the CC Map	12
2.2 Changes to SAS Analytic Packages (SAS Packs).....	13
3. FINAL MODELS AND ASSESSMENT OF PERFORMANCE	14
3.1 Overview of Methodology and Results	14
3.2 2010 AMI Mortality Model.....	15
3.2.1 Index Cohort	15
3.2.2 Frequency of AMI Model Variables over Different Time Periods ..	17
3.2.3 Model Parameters	17
3.2.4 Distribution of Hospital RSMRs.....	17
3.3 2010 HF Mortality Model	22
3.3.1 Index Cohort	22
3.3.2 Frequency of HF Model Variables over Different Time Periods	24
3.3.3 Model Parameters	24
3.3.4 Distribution of Hospital RSMRs.....	24
3.4 2010 PN Mortality Model	29
3.4.1 Index Cohort	29
3.4.2 Frequency of PN Model Variables over Different Time Periods....	31
3.4.3 Model Parameters	31
3.4.4 Distribution of Hospital RSMRs.....	31
4. SAS PACK.....	36
4.1 Revision to SAS Pack.....	36
5. QUALITY ASSURANCE.....	37
5.1 QA for Input Data.....	37
6. REFERENCES.....	41
7. APPENDIX.....	43

LIST OF TABLES

Table 1 – Distribution of AMI Model Variables over Different Time Periods	18
Table 2 – Adjusted ORs and 95% CIs for AMI HGLM over Different Time Periods	19
Table 3 – Distribution of Hospital Volume and RSMR in AMI Cohort over Different Time Periods	20
Table 4 – Distribution of HF Model Variables over Different Time Periods.....	25
Table 5 – Adjusted ORs and 95% CIs for the HF HGLM over Different Time Periods ..	26
Table 6 – Distribution of Hospital Volume and RSMR in HF Cohort over Different Time Periods	27
Table 7 – Distribution of PN Model Variables over Different Time Periods	32
Table 8 – Adjusted ORs and 95% CIs for the PN HGLM over Different Time Periods ..	33
Table 9 – Distribution of Hospital Volume and RSMR in PN Cohort over Different Time Periods	34

LIST OF FIGURES

Figure 1 – Patient Sample for AMI in the 2006-2008 Calendar Year Dataset	16
Figure 2 – Distribution of Hospital 30-Day RSMRs for AMI in Three-Year Combined Data (2006-2008)	21
Figure 3 – Patient Sample for HF in the 2006-2008 Calendar Year Dataset.....	23
Figure 4– Distribution of Hospital 30-Day RSMRs for HF in Three-Year Combined Data (2006-2008)	28
Figure 5 – Patient Sample for PN in the 2006-2008 Calendar Year Dataset.....	30
Figure 6 – Distribution of Hospital 30-Day RSMRs for PN in Three-Year Combined Data (2006-2008)	35
Figure 7 – YNHHS-CORE QA Processes	39

ACKNOWLEDGEMENTS

This work is a collaborative effort, and the authors gratefully acknowledge and thank Angela Merrill, Sandi Nelson, Eric Schone, and Marian Wrobel from Mathematica Policy Research, Inc.; Jennifer Mattera and Lori Geary from Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE); and Lein Han and Michael Rapp at the Centers for Medicare & Medicaid Services for their contributions to this work.

1. INTRODUCTION

1.1 Background on Mortality Measures

In June 2007, the for Medicare & Medicaid Services (CMS) began publicly reporting hospital 30-day risk-standardized mortality rates (RSMRs) for acute myocardial infarction (AMI) and heart failure (HF) for the nation's acute care and critical access hospitals. CMS added the pneumonia (PN) mortality measure in August 2008. These measures are posted on *Hospital Compare* (<http://www.hospitalcompare.hhs.gov>) and CMS updates them annually. The measures are based on administrative claims and enrollment data for Fee-for-Service (FFS) Medicare beneficiaries.

These outcome measures were originally developed by a team of clinical and statistical experts from Yale New Haven Health Services Corporation / Center for Outcomes Research and Evaluation (YNHHSC/CORE), Yale University and Harvard University. All three measures are consistent with the American Heart Association standards for measures suitable for public reporting of outcomes¹ and have been endorsed by the National Quality Forum (NQF). CMS contracted with YNHHSC/CORE to prepare the 30-day mortality measures for AMI, HF and PN for 2010 public reporting through a process of measure maintenance.

This report summarizes our measure maintenance activities, describes the minor updates made to the model for this year, and presents the 2010 models. It is a supplement to and update of the prior methodology reports produced for each measure rather than a comprehensive description of measure methods. The reports that present the measure methodology in full for each measure are available at *QualityNet* (<http://www.qualitynet.org>):

- Risk-Adjustment Models for AMI and HF 30-Day Mortality: Methodology (2005)²;
- Risk-Adjustment Methodology for Hospital Monitoring/Surveillance and Public Reporting: Supplement #1: 30-Day Mortality Model for Pneumonia (2006)³;
- Acute Myocardial Infarction, Heart Failure, and Pneumonia Mortality Measures Maintenance Technical Report (2008)⁴;
- Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-Day Risk-Standardized Mortality Measures Maintenance Technical Report (2009)⁵.

The AMI and HF methodologies are also described in the medical literature.^{1, 6-9}

1.2 Overview of Measure Methodology

The 2010 mortality risk-adjustment models use the NQF approved methodology set forth in the original methods reports^{2,3} (with slight refinements to the measures as described in the two measures maintenance reports^{4,5}). Below we provide an overview of the methodology. Updates for 2010 are found in Section 2. The mortality measures use hierarchical generalized linear modeling (HGLM) to create a RSMR at the hospital level that reflects hospital quality. The measures incorporate claims data from one year prior to the hospital admission to adjust for case-mix differences at hospitals.

Cohort

Index Cohort

The AMI, HF, and PN measures include admissions for Medicare FFS beneficiaries aged ≥65 years discharged from non-federal acute care hospitals having a principal discharge diagnosis of AMI, HF, or PN, respectively, and with a complete claims history for the 12 months prior to the date of admission. For specific International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes used to define the inclusion cohort for each condition, refer to Sections 3.2.1, 3.3.1, and 3.4.1 for AMI, HF, and PN, respectively. For patients with more than one admission in a specific year for any given diagnosis, only one admission was randomly selected to keep in the cohort and others were excluded. An *index admission* is the hospitalization considered for the outcome.

Cohort Exclusions (Excluded Admissions)

The measures exclude admissions for patients:

- who were discharged on the day of admission or the following day and did not die or get transferred (because it is less likely they had a significant HF, diagnosis of AMI, or PN);
- who were transferred from another acute care hospital (because the death is attributed to the hospital where the patient was initially admitted);
- with inconsistent or unknown mortality status or other unreliable data (e.g. date of death precedes admission date);
- enrolled in the Medicare Hospice program any time in the 12 months prior to the index hospitalization including the first day of the index admission (since it is likely these patients are continuing to seek comfort measures only);
- who were discharged alive and against medical advice (AMA) (because providers did not have the opportunity to deliver full care and prepare the patient for discharge);
- that were not the first hospitalization in the 30 days prior to a patient's death. This exclusion criterion is applied after one admission per patient per year is randomly selected. It *only* applies when two randomly selected admissions

occur during the transition months (December and January for calendar-year data) and the patient subsequently dies. For example: a patient is admitted on December 18th, 2006 and readmitted on January 2nd, 2007; the patient dies on January 15th, 2007. If both of these admissions are randomly selected for inclusion (one for the 2006 calendar year time period and the other for the 2007 calendar year time period), the January 2, 2007 admission will be excluded to avoid assigning the death to two admissions (one in 2006 and one in 2007).

The number of patients excluded based on each criterion is available in Figure 1, 3, 5 for AMI, HF, and PN respectively.

Transferred Patients

The measures include patients who are admitted to a hospital with a diagnosis of AMI, HF, or PN and then transferred to another acute facility, if the primary discharge diagnosis at the second hospital matches the one at the first hospital. The model considers admission to the first hospital as the start of an acute episode of care and assigns the patient's outcome to the hospital that initially admitted them. The model does not assign these patients to the hospitals that receive them. For those seen in the emergency department only and transferred to another hospital, the model assigns them to the receiving hospital.

Outcome

30-Day Timeframe

The measures assess mortality within a 30-day period from admission for the index hospitalization. The standard period is necessary so that the outcome for each patient is measured consistently. Without a standard measurement period, variation in length of stay could have an undue influence on mortality rates, and institutions would have an incentive to adopt strategies to shift deaths out of the hospital without improving quality. The use of the 30-day timeframe also places an emphasis on transitions of care and the suitability of the patient for discharge.

All-Cause Mortality

The mortality measures assess deaths for all causes, regardless of the underlying principal diagnosis, because from a patient perspective death from any cause is an adverse event. In addition, it is also difficult to make inferences about quality issues and accountability based solely on the documented cause of death; in many cases, accurate attribution of the cause of death is challenging. For example, a patient with HF who develops a hospital-acquired infection may ultimately die of sepsis and multi-organ failure. It would be inappropriate to consider the death as unrelated to the care the patient received for HF during the hospitalization.

Risk-Adjustment Variables

The measures adjust for key variables that are clinically relevant and have strong relationships with the outcome (e.g. demographic factors, disease severity indicators, and indicators of frailty). For each patient, covariates are obtained from Medicare claims extending 12 months prior to and including the index admission.

The models seek to adjust for case differences based on the clinical status of the patient at the time of admission. Accordingly, only comorbidities that convey information about the patient at that time or in the 12-months prior, and not complications that arise during the course of the hospitalization are included in the risk-adjustment.

The models do not adjust for the patients' admission source and their discharge disposition (e.g. skilled nursing facility) because these factors are associated with structure of the health care system, not solely patients' clinical risk factors. Regional differences in resource availability and practice patterns may exert an undue influence on model results. Moreover, the validity of these admission and discharge disposition codes is not known. The measures also do not adjust for socioeconomic status (SES) because the association between SES and health outcomes can be due, in part, to the quality of health care. Risk-adjusting for patient SES would suggest that hospitals with low SES patients are held to different standards for the risk of mortality than hospitals treating higher SES patient populations. The intention is for the measures to adjust for patient demographic and clinical characteristics while illuminating important quality differences. This methodology is consistent with guidance from NQF.

Refer to Tables 1, 4, and 7 in this report for the list of risk-adjusted variables for AMI, HF, and PN respectively.

Calculating the RSMR

The measures estimate hospital-level 30-day all-cause RSMRs for each condition using HGLMs. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals.⁸ At the patient level, each model adjusts the log-odds of a hospital mortality within 30-days of discharge for age, sex, selected clinical covariates, and a hospital-specific intercept. Comorbidities from the index admission that could represent complications of care are not included in the risk adjustment unless they are also present in the 12 months prior to admission. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of mortality at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution in order to account for the clustering (non-independence) of

patients within the same hospital.⁸ If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

The RSMR is calculated as the ratio of the number of “predicted” to the number of “expected” deaths, multiplied by the national unadjusted mortality rate. For each hospital, the “numerator” of the ratio is the number of deaths within 30 days predicted on the basis of the hospital’s performance with its observed case mix, and the “denominator” is the number of deaths expected on the basis of the nation’s performance with that hospital’s case mix. This approach is analogous to a ratio of “observed” to “expected” used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital’s performance given its case-mix to an average hospital’s performance with the same case-mix. Thus a lower ratio indicates lower-than-expected mortality or better quality and a higher ratio indicates higher-than-expected mortality or worse quality.

The predicted hospital outcome (the numerator) is calculated by regressing the risk factors (see Tables 1, 5, and 9 for AMI, HF, and PN, respectively) and the hospital-specific intercept on the risk of mortality, multiplying the estimated regression coefficients by the patient characteristics in the hospital, transforming, and then summing over all patients attributed to the hospital to get a value. The expected number of deaths (the denominator) is obtained by regressing the risk factors and a common intercept on the mortality outcome using all hospitals in our sample, multiplying the subsequent estimated regression coefficients by the patient characteristics observed in the hospital, transforming, and then summing over all patients in the hospital to get a value. To assess hospital performance in any reporting period, we re-estimate the model coefficients using the years of data in that period.

The statistical models used are described fully in the original methodology reports.^{2,3}

1.3 Goals of Measure Maintenance

Measure maintenance is a process to continually improve the measures. Conducted annually, it is an opportunity to reflect on and respond to comments made in the last year of public reporting and to incorporate advances in the science and any changes in coding. It ensures that the risk-standardized mortality models are continually assessed and remain valid given possible changes in the data over time, and allows for model refinements. As described below, for 2010 public reporting, CMS undertook the following measure maintenance activities this year:

- making minor refinements to the model (see Section 2.1);

- validating the performance of each condition-specific model and its corresponding risk-adjustment variables in three recent one-year datasets (2006-2008);
- evaluating and validating model performance in this three-year combined dataset (2006-2008);
- updating the Quality Assurance (QA) process and SAS pack documentation.

2. UPDATES TO METHODS

2.1 Refinements to the Mortality Measures

For 2010 public reporting, we made the following minor refinements to the model:

- slightly refined how we gather claims history used to derive risk-adjustment variables to more fully capture patient risk factors;
- updated the Condition Categories (CC) map to incorporate ICD-9-CM coding updates.

We assessed the effects of these changes using admissions with discharge dates from January 1, 2006, through December 31, 2008 (this dataset is referred to in this report as the 2006-2008 calendar year dataset). These changes are discussed in more detail below.

2.1.1 Change in Collection of Claims History Data for Determination of Patient Comorbidities

Modification: In order to obtain clinical covariates from Medicare claims for risk-adjustment, a small change was made this year. Previously, claims from visits that occurred within 365 days prior to the admission date of the index admission were included, based on the *admission or begin* date of the prior visit. This year, we included all the claims from prior visits for which the *discharge or end* date occurred within 365 days prior to the admission date of the index admission.

Rationale: This change allows for inclusion of clinical covariates from visits that overlap with the 365-day period prior to the index admission. This approach is slightly more inclusive of covariates for risk-adjustment, as it includes covariates from prior visits that began more than 365 days before the index admission but ended fewer than 365 days earlier.

Effects on patient cohort: The result of this change is a slight increase in the frequency of a number of covariates.

2.1.2 Update to the CC Map

Modification: A second CMS contractor, RTI International, updated the map linking ICD-9-CM codes to CCs – clinically related groups of conditions used for measure risk-adjustment variables – to reflect ICD-9-

CM codes in use during the full reporting period. RTI International, contracted by CMS to maintain the CC system, assigned new ICD-9-CM codes to the existing CCs based on their clinical expertise and the historical assignment of related ICD-9-CM codes to the CCs.

Rationale: CMS revises the ICD-9-CM CC map annually to reflect changes in ICD-9-CM codes so that the models will capture all relevant comorbidities coded in patient claims data.

Effects on Model Variables: The assignment of new codes and the removal of retired codes had little impact on the model variables since RTI assigned the majority of new codes, which were more specific versions of retired codes, to the same CCs as retired codes. For more details on the CC changes, see Appendix for RTI's memo to CMS detailing the map changes.

2.2 Changes to SAS Analytic Packages (SAS Packs)

We revised the SAS pack to reflect any changes to the admission cohorts and models as needed, this includes any ad hoc patches that address data issues. There were two additions to the SAS Macro program this year. One provides a more thorough clean-up of all intermediate files after the SAS pack has completed and the other creates a graphical representation of the distribution of 30-Day RSMRs for the time period and condition that is being run. The new SAS pack is named "Final Mortality 2010 SAS Pack" (see Section 4 for details).

3. FINAL MODELS AND ASSESSMENT OF PERFORMANCE

3.1 Overview of Methodology and Results

The 2010 mortality models estimate hospital-specific, 30-day all-cause RSMRs using HGLMs. To adjust for differences in hospital case mix, the models adjust for patient risk factors, including age and comorbidities present at the time of admission. A brief description of the measure methodology and model risk-adjustment variables is in Section 1.2 of this report and in detail in prior technical reports.²⁻⁵

The measure links admissions for patients who are transferred between acute care hospitals and have the same qualifying diagnosis at both hospitals into a single acute episode of care. The outcome for the patient is assigned to the first hospital in the sequence of transfers for the purposes of evaluating 30-day mortality.

To evaluate the performance of the models used for 2010 reporting, we fit the revised models to three single, calendar-year datasets (2006, 2007, and 2008) and to the combined three-year 2006-2008 calendar-year dataset. We re-estimated the model variable coefficients, examined their trends across time periods, and examined the model performance in each of these datasets. We also examined trends in the frequency of patient risk factors. Although we made the minor model changes described in Section 2, we otherwise preserved the original methodology and did not, for example, reselect variables for inclusion into the models.

For each of the three measures, we assessed HGLM performance in terms of discriminant ability and overall fit for each calendar year of data (2006, 2007, 2008) and for the three year combined period (2006-2008). We computed two summary statistics for assessing model performance: the adjusted R^2 , which indicates the percentage of the patient-level variation in the outcome explained by the model variables, and the area under the receiver operating characteristic (ROC) curve (c-statistic), which is an indicator of the model's discriminant ability or ability to correctly classify those who die and do not die within 30 days (values range from 0.5 meaning no better than chance to 1.0 meaning perfect discrimination).

We also assessed model performance for each measure using preliminary public reporting data for 2010 (admissions with discharges between July 1, 2006, and June 30, 2009). The results (data not shown) were substantively similar to those for the 2006-2008 calendar year dataset.

The administrative data sources for the measure maintenance analyses are Medicare administrative datasets that contain claims and enrollment information for FFS hospitalizations for calendar years 2006–2008. The datasets also contain associated inpatient and outpatient claims information in the prior 12 months for patients admitted in each of these years. Please see the methodology reports²⁻⁵ for complete descriptions of these data sources.

The results of these analyses for each of the three measures (AMI, HF, and PN) are presented below in Sections 3.2, 3.3, and 3.4, respectively.

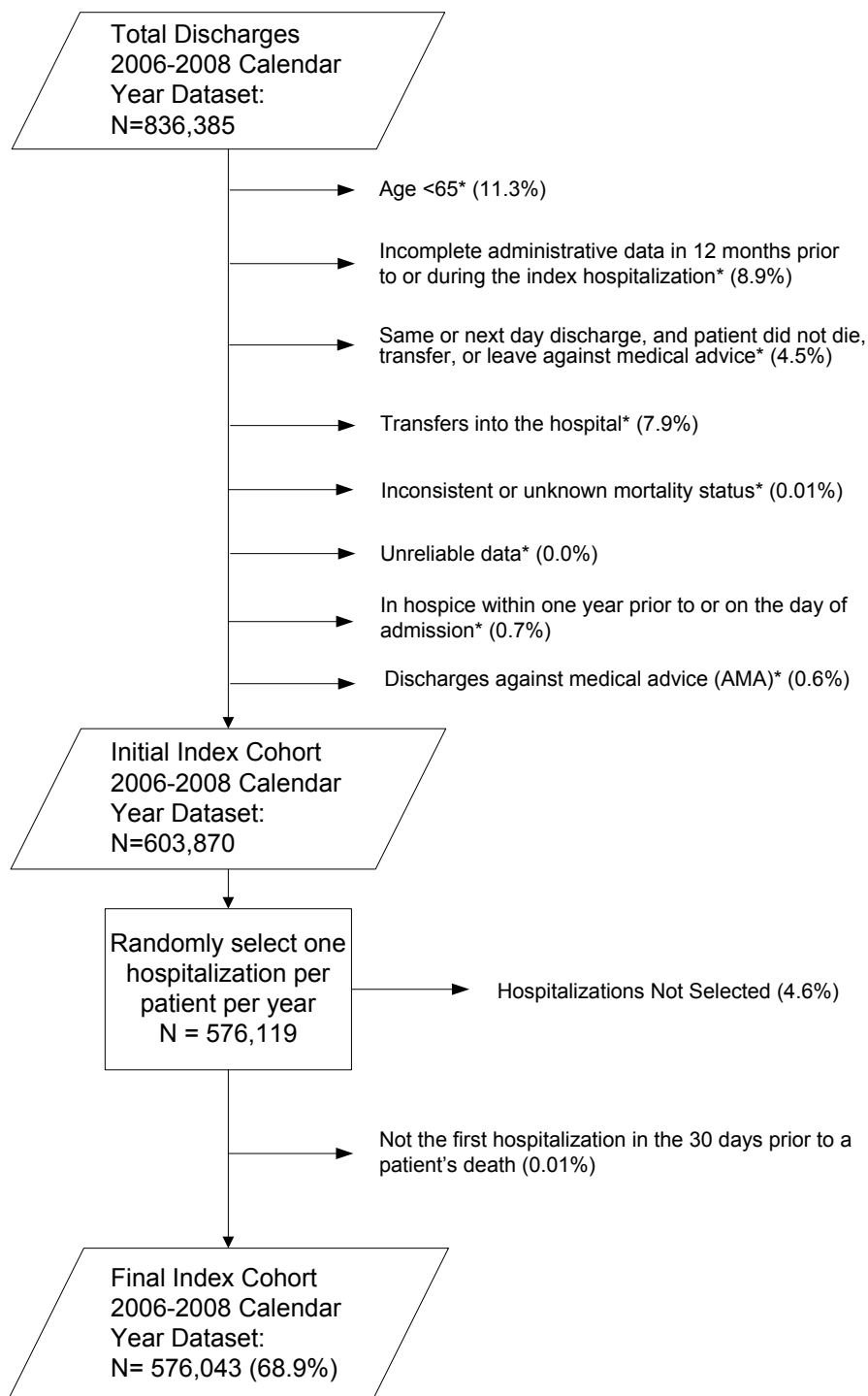
3.2 2010 AMI Mortality Model

3.2.1 Index Cohort

The cohort includes admissions for Medicare FFS beneficiaries aged ≥ 65 years discharged from the hospital with a principal discharge diagnosis of ICD-9-CM code 410.xx, excluding those with 410.x2 (AMI, subsequent episode of care) and with a complete claims history for the 12 months prior to admission.

The exclusion criteria for the measures are presented in Section 1.2, and the percentage of AMI patients meeting each exclusion criterion in the 2006-2008 calendar year dataset is presented in Figure 1.

Figure 1 – Patient Sample for AMI in the 2006-2008 Calendar Year Dataset



*Categories are not mutually exclusive

3.2.2 Frequency of AMI Model Variables over Different Time Periods

We examined the temporal variation in frequency of clinical and demographic variables. The crude mortality rate across the cohorts decreased slightly from 16.5% in 2006 to 16.2% in 2008 (Table 1). The only notable changes among comorbid conditions were in: renal failure, the percentage of AMI patients with renal failure increased from 16.5% in 2006 to 20.3% in 2008 and hypertension, which also increased from 79.1% in 2006 to 83.1% in 2008.

3.2.3 Model Parameters

Table 2 conveys the risk-adjusted odds ratios (ORs) and 95% confidence intervals (CIs) for the AMI model by individual year and for the 2006-2008 calendar year dataset. The parameters are consistent across all time periods. Age was strongly associated with risk of death (OR per year over age 65: 1.06; 95% CI 1.06-1.06). History of percutaneous transluminal coronary angioplasty (PTCA) and coronary artery bypass graft (CABG), chronic atherosclerosis, and hypertension were inversely associated with risk of death. All other variables were associated with an increased risk of death. Model performance was stable over all time periods; the area under the ROC curve was 0.72 across all time periods.

3.2.4 Distribution of Hospital RSMRs

Table 3 shows the distributions of hospital volume, hospital RSMR and between-hospital variance over different time periods. Mean AMI volume slightly decreased from 46 admissions (SD: 63) per hospital in 2006, to 44 admissions (SD: 59) per hospital in 2008. RSMR also decreased over the three year period, from 16.3% in 2006 to 15.9% in 2008. The mean hospital RSMR for the combined three-year data was 15.8 (SD: 1.8; range 10.6 – 24.6). The 25th and 75th percentiles were 14.6% and 16.9% respectively in the combined three-year dataset. Between-hospital variance remained stable across all cohort years ranging from 0.042 (SE: 0.004) – 0.054 (SE: 0.005). Between-hospital variance in the combined, three-year dataset was 0.047 (SE: 0.003). If there were no systematic differences between hospitals, the between-hospital variance would be 0.

Figure 2 shows the overall distribution of the hospital RSMRs for the three year calendar year dataset. The odds of all-cause mortality for a hospital that was one standard deviation above average were 1.55 times that of a hospital that was one standard deviation below average. If there were no systematic differences between hospitals, the OR would be 1.0.¹⁰

Table 1 – Distribution of AMI Model Variables over Different Time Periods

Description	2006	2007	2008	2006-2008
Total Number of Admissions	201,479	189,599	185,041	576,043
Crude Mortality Rate (%)	16.5	16.4	16.2	16.4
Demographics				
Mean Age (SD)	79.2 (8.0)	79.3 (8.1)	79.5 (8.2)	79.3 (8.1)
Male (%)	48.98	48.97	48.93	48.96
Cardiovascular (%)				
History of PTCA	7.55	7.63	7.71	7.63
History of CABG	6.42	6.23	6.04	6.24
Congestive heart failure (CC 80)	31.10	31.31	31.26	31.22
Acute myocardial infarction (CC 81)	13.34	13.35	13.69	13.45
Unstable angina (CC 82)	14.03	13.67	13.51	13.74
Anterior myocardial infarction (ICD9 410.00-410.19)	11.47	10.67	9.78	10.66
Other location of myocardial infarction (ICD9 410.20-410.69)	15.16	14.45	13.18	14.29
Chronic atherosclerosis (CC 83, 84)	75.86	76.25	76.95	76.34
Cardio-respiratory failure and shock (CC 79)	7.81	8.71	9.21	8.56
Valvular and rheumatic heart disease (CC 86)	30.94	31.13	27.31	29.83
Comorbidity (%)				
Hypertension (CC 89, 91)	79.10	81.65	83.09	81.22
Stroke (CC 95 or 96)	8.36	8.25	8.23	8.28
Cerebrovascular disease (CC 97-99, 103)	18.68	19.21	19.74	19.20
Renal failure (CC 131)	16.54	19.25	20.33	18.65
Chronic obstructive pulmonary disease (COPD) (CC 108)	30.41	30.40	28.64	29.84
Pneumonia (CC 111-113)	23.02	23.54	25.01	23.83
Diabetes or DM complications (CC 15-20, 120)	40.38	41.23	41.94	41.16
Protein-calorie malnutrition (CC 21)	3.61	4.02	4.91	4.16
Dementia or senility (CC 49, 50)	17.33	17.72	18.40	17.80
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177, 178)	5.31	5.41	5.83	5.51
Peripheral vascular disease (CC 104, 105)	24.00	24.95	25.50	24.79
Metastatic cancer, acute leukemia and other major cancers (CC 7, 8)	3.59	3.82	3.90	3.76
Trauma in last year (CC 154-156, 158-162)	27.60	27.91	28.32	27.93
Major psych disorders (CC 54-56)	6.26	6.55	6.84	6.54
Chronic liver disease (CC 25-27)	0.93	1.01	0.96	0.96

Table 2 – Adjusted ORs and 95% CIs for AMI HGLM over Different Time Periods

Variable	2006 OR	2006 95% CI	2007 OR	2007 95% CI	2008 OR	2008 95% CI	2006-2008 OR	2006-2008 95% CI
Demographics								
Age-65 ^a	1.06	(1.05-1.06)	1.06	(1.05-1.06)	1.06	(1.05-1.06)	1.06	(1.06-1.06)
Male	1.15	(1.12-1.18)	1.13	(1.1-1.16)	1.14	(1.11-1.17)	1.14	(1.12-1.16)
Cardiovascular								
History of PTCA	0.62	(0.59-0.66)	0.66	(0.62-0.70)	0.63	(0.6-0.67)	0.64	(0.62-0.66)
History of CABG	0.84	(0.79-0.89)	0.83	(0.78-0.88)	0.85	(0.79-0.90)	0.83	(0.81-0.86)
Congestive heart failure (CC 80)	1.44	(1.40-1.49)	1.42	(1.38-1.46)	1.43	(1.39-1.48)	1.43	(1.40-1.45)
History of AMI (CC 81)	1.01	(0.98-1.05)	1.06	(1.02-1.11)	1.01	(0.97-1.05)	1.02	(1.00-1.05)
Unstable angina (CC 82)	0.99	(0.95-1.03)	0.94	(0.90-0.98)	1.01	(0.97-1.05)	0.98	(0.96-1.00)
Anterior myocardial infarction (ICD9 410.00-410.19)	1.77	(1.70-1.83)	1.89	(1.82-1.97)	1.81	(1.74-1.89)	1.83	(1.79-1.87)
Other location of myocardial infarction (ICD9 410.20-410.69)	1.51	(1.46-1.57)	1.52	(1.46-1.58)	1.53	(1.47-1.59)	1.53	(1.49-1.56)
Chronic atherosclerosis (CC 83, 84)	0.51	(0.50-0.53)	0.54	(0.52-0.55)	0.53	(0.52-0.55)	0.53	(0.52-0.54)
Cardio-respiratory failure and shock (CC 79)	1.23	(1.18-1.28)	1.23	(1.18-1.29)	1.23	(1.18-1.29)	1.23	(1.20-1.26)
Valvular and rheumatic heart disease (CC 86)	1.00	(0.98-1.03)	1.02	(0.99-1.05)	1.01	(0.98-1.04)	1.01	(1.00-1.03)
Comorbid Conditions								
Hypertension (CC 89, 91)	0.71	(0.69-0.73)	0.72	(0.70-0.74)	0.69	(0.67-0.72)	0.71	(0.69-0.72)
Stroke (CC 95, 96)	1.13	(1.08-1.18)	1.14	(1.09-1.19)	1.09	(1.04-1.14)	1.12	(1.09-1.15)
Cerebrovascular disease (CC 97-99, 103)	0.99	(0.95-1.02)	1.00	(0.96-1.03)	0.97	(0.94-1.00)	0.99	(0.97-1.00)
Renal failure (CC 131)	1.30	(1.26-1.34)	1.30	(1.26-1.35)	1.25	(1.21-1.29)	1.28	(1.26-1.31)
COPD (CC 108)	1.12	(1.09-1.15)	1.10	(1.07-1.13)	1.04	(1.01-1.07)	1.09	(1.07-1.11)
Pneumonia (CC 111-113)	1.33	(1.29-1.37)	1.37	(1.33-1.41)	1.42	(1.38-1.46)	1.37	(1.35-1.39)
Diabetes and DM complications (CC 15-20, 120)	1.06	(1.04-1.09)	1.05	(1.02-1.08)	1.05	(1.02-1.08)	1.05	(1.04-1.07)
Protein-calorie malnutrition (CC 21)	1.53	(1.45-1.61)	1.65	(1.57-1.74)	1.73	(1.65-1.82)	1.64	(1.59-1.69)
Dementia and senility (CC 49, 50)	1.42	(1.38-1.47)	1.37	(1.33-1.42)	1.41	(1.37-1.45)	1.40	(1.37-1.42)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177, 178)	1.27	(1.21-1.34)	1.24	(1.17-1.30)	1.23	(1.16-1.29)	1.24	(1.20-1.28)
Peripheral vascular disease (CC 104, 105)	1.12	(1.08-1.15)	1.12	(1.09-1.16)	1.14	(1.11-1.18)	1.13	(1.11-1.15)
Metastatic cancer, acute leukemia, and other severe cancers (CC 7, 8)	1.98	(1.87-2.09)	2.11	(2.00-2.23)	1.99	(1.88-2.10)	2.03	(1.96-2.09)
Trauma in last year (CC 154-156, 158-162)	1.01	(0.98-1.03)	1.00	(0.97-1.03)	1.00	(0.97-1.03)	1.00	(0.98-1.02)
Major psych disorders (CC 54-56)	1.18	(1.12-1.23)	1.14	(1.09-1.20)	1.08	(1.03-1.13)	1.13	(1.10-1.16)
Chronic liver disease (CC 25-27)	1.42	(1.27-1.59)	1.52	(1.36-1.70)	1.51	(1.35-1.70)	1.48	(1.39-1.58)
Model Performance^b								
Adjusted R ²		0.13		0.13		0.13		0.13
c- statistic		0.72		0.72		0.72		0.72

^a Patients' age – 65

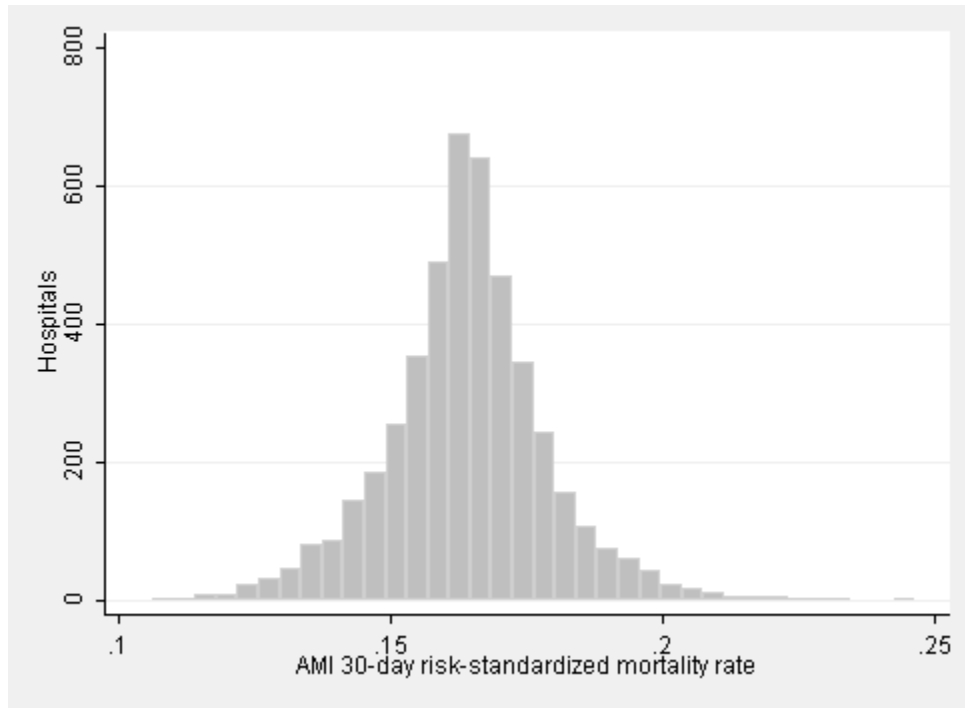
^b Obtained from GLM

Table 3 – Distribution of Hospital Volume and RSMR in AMI Cohort over Different Time Periods

Characteristic	2006	2007	2008	2006-2008
Number of Hospitals	4,344	4,287	4,251	4,590
Hospital Volume				
Mean Number of Admissions (SD)	46.4 (62.7)	44.2 (59.9)	43.5 (58.7)	125.5 (176.8)
Range (min. – max.)	1 - 509	1-509	1 - 474	1 – 1,421
25 th Percentile	6	6	5	13
50 th Percentile	21	19	19	50
75 th Percentile	64	61	60	174
RSMR (%) (Percentiles below weighted by hospital volume)				
Mean (SD)	16.3 (1.7)	16.1 (1.4)	15.9 (1.6)	15.8 (1.8)
Range (min. – max.)	11.7 - 22.1	11.7 - 21.3	11.1 - 25.2	10.6 - 24.6
25 th Percentile ⁺	15.1	15.2	14.8	14.6
50 th Percentile ⁺	16.2	16.1	15.8	15.7
75 th Percentile ⁺	17.3	17.0	17.0	16.9
Between Hospital Variance* (SE)	0.051 (0.005)	0.042 (0.004)	0.054 (0.005)	0.047 (0.003)

* Results from hierarchical model

Figure 2 – Distribution of Hospital 30-Day RSMRs for AMI in Three-Year Combined Data (2006-2008)



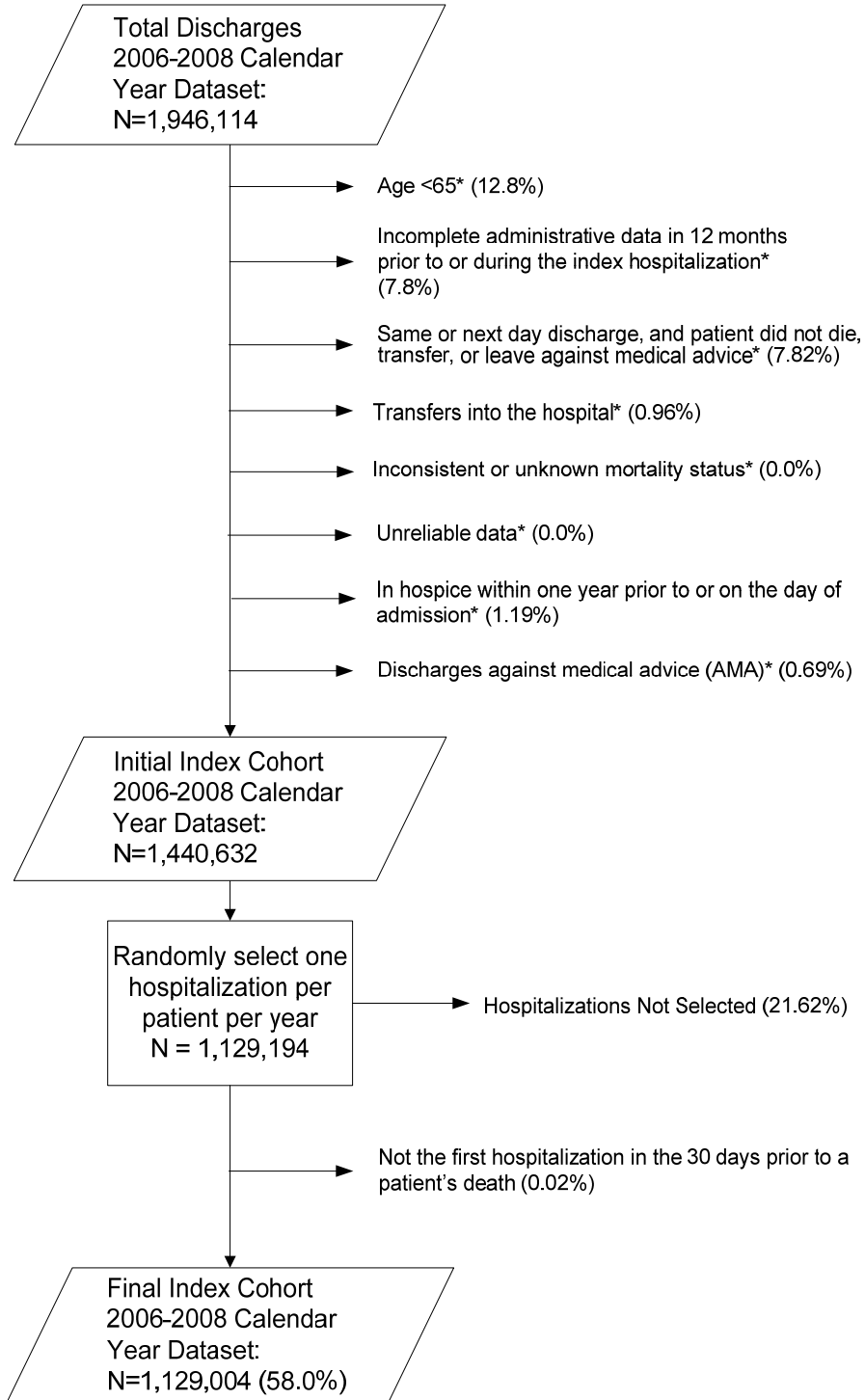
3.3 2010 HF Mortality Model

3.3.1 Index Cohort

The cohort includes admissions for Medicare FFS beneficiaries aged ≥ 65 years discharged from the hospital with a principal discharge diagnosis of HF (ICD-9-CM codes 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, and 428.xx) and with a complete claims history for 12 months prior to admission.

The exclusion criteria for the measures are presented in Section 1.2, and the percentage of HF patients meeting each exclusion criteria in the 2006-2008 calendar year dataset is presented in Figure 3.

Figure 3 – Patient Sample for HF in the 2006-2008 Calendar Year Dataset



*Categories are not mutually exclusive

3.3.2 Frequency of HF Model Variables over Different Time Periods

We examined the temporal variation in frequency of clinical and demographic variables by time period (Table 4). The crude mortality rate across cohorts remained constant at about 11%. The only notable changes among comorbid conditions were in: renal failure, the percentage of HF patients with renal failure increased from 33.2% in 2006 to 38.8% in 2008 and hypertension, which also increased from 84.1% in 2006 to 88.2% in 2008.

3.3.3 Model Parameters

Table 5 conveys the adjusted ORs and 95% CIs for the HF model by time period. The coefficients are consistent across all cohort years. Age was most strongly associated with risk of death (OR 1.05; 95% CI 1.05 – 1.05). History of PTCA and CABG, unstable angina, chronic atherosclerosis, hypertension and diabetes were inversely associated with risk of death. All other variables were associated with an increased risk of death. Model performance was stable over all time periods; the area under the ROC curve was 0.69 across all time periods.

3.3.4 Distribution of Hospital RSMRs

Table 6 shows the distributions of hospital volume, RSMR and between-hospital variance by time period. Mean HF volume decreased from 87 admissions (SD: 98) in 2006, to 77 admissions (SD: 88) in 2008. RSMRs remained stable over the three year period. The mean RSMR for the combined three-year data was 10.9 (SD: 1.6; range 6.4 – 19.4). The 25th and 75th percentiles were 9.8% and 11.9% respectively in the combined three-year dataset. Between-hospital variance remained stable across all time periods ranging from 0.054 (SE: 0.004) – 0.056 (SE: 0.004). Between-hospital variance in the combined three-year dataset was 0.053 (SE: 0.002). If there were no systematic differences between hospitals, the between-hospital variance would be 0.

Figure 4 shows the distribution of the hospital RSMRs based on the three year combined data. The odds of all-cause mortality for a patient treated at a hospital that was one standard deviation above average, was 1.59 times that of a hospital that was one standard deviation below average. If there were no systematic differences between hospitals, the OR would be 1.0.¹⁰

Table 4 – Distribution of HF Model Variables over Different Time Periods

Description	2006	2007	2008	2006-2008
Total Number of Admissions	403,462	371,347	354,385	1,129,004
Crude Mortality Rate (%)	11.04	11.15	11.36	11.16
Demographics				
Mean Age (SD)	80.7 (7.93)	80.9 (7.96)	81.1 (8.03)	80.9 (7.97)
Male (%)	42.21	42.50	43.10	42.59
Cardiovascular (%)				
History of PTCA	6.71	6.62	6.40	6.58
History of CABG	10.96	10.32	9.40	10.26
Congestive heart failure (CC 80)	73.08	73.39	73.41	73.28
Acute myocardial infarction (CC 81)	9.35	9.36	9.72	9.47
Unstable angina (CC 82)	14.33	13.65	13.29	13.78
Chronic atherosclerosis (CC 83, 84)	70.73	70.74	70.74	70.73
Cardio-respiratory failure and shock (CC 79)	18.13	19.88	20.64	19.49
Valvular and rheumatic heart disease (CC 86)	48.48	49.41	45.53	47.86
Comorbidity (%)				
Hypertension (CC 89, 91)	84.06	86.76	88.21	86.25
Stroke (CC 95, 96)	10.46	10.47	10.30	10.41
Renal failure (CC 131)	33.24	37.12	38.80	36.26
COPD (CC 108)	47.38	47.31	45.04	46.62
Pneumonia (CC 111-113)	40.61	41.88	43.28	41.86
Diabetes and DM complications (CC 15-20, 120)	49.35	49.90	50.11	49.77
Protein-calorie malnutrition (CC 21)	5.60	6.02	7.22	6.24
Dementia and senility (CC 49, 50)	20.66	21.17	21.64	21.13
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177, 178)	6.74	6.79	7.20	6.90
Peripheral vascular disease (CC 104, 105)	33.32	34.19	34.84	34.08
Metastatic cancer, acute leukemia, and other severe cancers (CC 7, 8)	4.08	4.16	4.25	4.16
Trauma in last year (CC 154-156, 158-162)	34.81	35.47	35.60	35.27
Major psych disorders (CC 54-56)	8.62	8.73	9.05	8.79
Chronic liver disease (CC 25-27)	1.96	2.03	2.02	2.00

Table 5 – Adjusted ORs and 95% CIs for the HF HGLM over Different Time Periods

Variable	2006 OR	2006 95% CI	2007 OR	2007 95% CI	2008 OR	2008 95% CI	2006-2008 OR	2006-2008 95% CI
Demographics								
Age-65 ^a	1.05	(1.05-1.05)	1.05	(1.05-1.05)	1.06	(1.05-1.06)	1.05	(1.05-1.05)
Male	1.27	(1.24-1.29)	1.27	(1.24-1.29)	1.30	(1.27-1.33)	1.28	(1.26-1.29)
Cardiovascular								
History of PTCA	0.61	(0.58-0.64)	0.60	(0.57-0.64)	0.61	(0.58-0.65)	0.61	(0.59-0.63)
History of CABG	0.63	(0.60-0.65)	0.64	(0.61-0.67)	0.64	(0.61-0.67)	0.63	(0.62-0.65)
Congestive heart failure (CC 80)	1.26	(1.23-1.29)	1.26	(1.23-1.30)	1.29	(1.25-1.32)	1.27	(1.25-1.29)
Acute myocardial infarction (CC 81)	1.36	(1.31-1.41)	1.37	(1.32-1.42)	1.35	(1.30-1.41)	1.36	(1.33-1.39)
Unstable angina (CC 82)	0.96	(0.92-0.99)	0.93	(0.90-0.96)	0.95	(0.91-0.98)	0.95	(0.93-0.97)
Chronic atherosclerosis (CC 83, 84)	0.85	(0.83-0.87)	0.87	(0.85-0.90)	0.87	(0.85-0.89)	0.87	(0.86-0.88)
Cardio-respiratory failure and shock (CC 79)	1.19	(1.16-1.22)	1.22	(1.19-1.25)	1.19	(1.16-1.22)	1.20	(1.18-1.22)
Valvular and rheumatic heart disease (CC 86)	1.03	(1.01-1.06)	1.02	(1.00-1.05)	1.00	(0.98-1.02)	1.02	(1.01-1.04)
Comorbidity								
Hypertension (CC 89, 91)	0.65	(0.63-0.66)	0.65	(0.63-0.67)	0.65	(0.63-0.67)	0.65	(0.64-0.66)
Stroke (CC 95, 96)	1.01	(0.98-1.04)	1.02	(0.99-1.06)	1.03	(0.99-1.06)	1.02	(1.00-1.04)
Renal failure (CC 131)	1.31	(1.28-1.34)	1.32	(1.29-1.35)	1.27	(1.24-1.30)	1.30	(1.29-1.32)
COPD (CC 108)	1.10	(1.07-1.12)	1.06	(1.04-1.08)	1.04	(1.02-1.07)	1.07	(1.06-1.08)
Pneumonia (CC 111-113)	1.37	(1.35-1.40)	1.33	(1.30-1.36)	1.38	(1.35-1.42)	1.36	(1.34-1.38)
Diabetes and DM complications (CC 15-20, 120)	0.94	(0.92-0.96)	0.95	(0.93-0.97)	0.94	(0.92-0.97)	0.94	(0.93-0.96)
Protein-calorie malnutrition (CC 21)	1.98	(1.91-2.04)	1.97	(1.91-2.04)	1.97	(1.90-2.03)	1.98	(1.94-2.02)
Dementia and senility (CC 49, 50)	1.37	(1.34-1.41)	1.32	(1.28-1.35)	1.29	(1.26-1.32)	1.33	(1.31-1.35)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177, 178)	1.19	(1.14-1.23)	1.16	(1.11-1.21)	1.17	(1.12-1.21)	1.17	(1.14-1.20)
Peripheral vascular disease (CC 104, 105)	1.04	(1.02-1.07)	1.02	(1.00-1.04)	1.03	(1.01-1.06)	1.04	(1.02-1.05)
Metastatic cancer, acute leukemia, and other severe cancers (CC 7, 8)	2.10	(2.02-2.19)	2.05	(1.97-2.14)	1.91	(1.83-2.00)	2.03	(1.99-2.08)
Trauma in last year (CC 154-156, 158-162)	1.05	(1.03-1.08)	1.07	(1.05-1.09)	1.07	(1.05-1.10)	1.06	(1.05-1.08)
Major psych disorders (CC 54-56)	1.10	(1.07-1.14)	1.15	(1.11-1.19)	1.08	(1.04-1.12)	1.11	(1.09-1.13)
Chronic liver disease (CC 25-27)	1.52	(1.43-1.62)	1.51	(1.41-1.61)	1.55	(1.45-1.66)	1.54	(1.48-1.60)
Model Performance ^b								
Adjusted R ²		0.09		0.09		0.09		0.09
c-statistic		0.69		0.69		0.69		0.69

^a Patients' age – 65

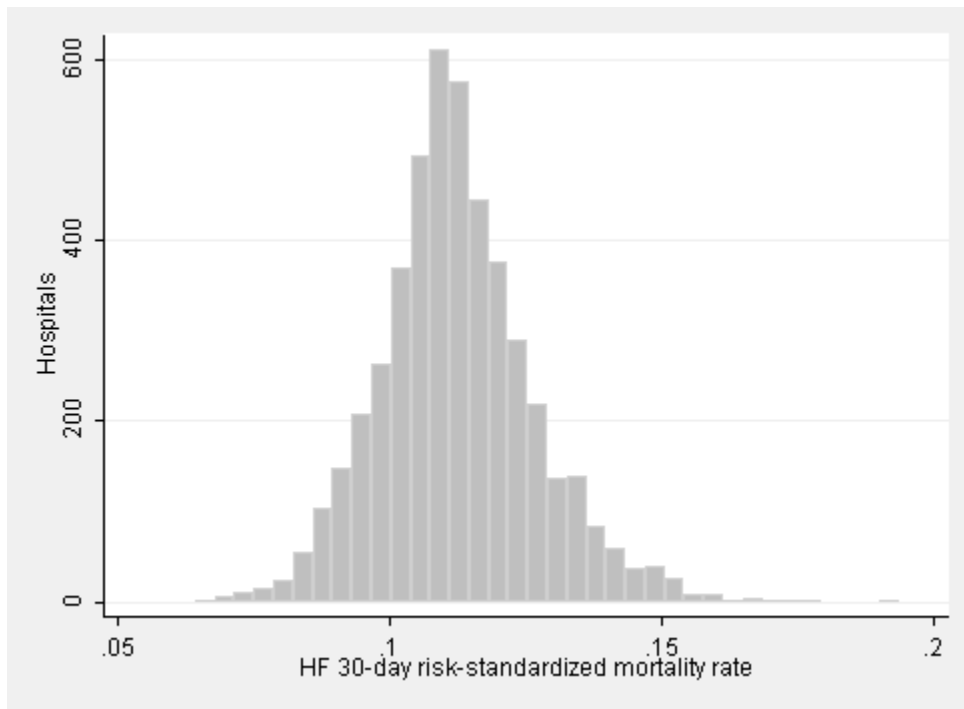
^b Obtained from GLM

Table 6 – Distribution of Hospital Volume and RSMR in HF Cohort over Different Time Periods

Characteristic	2006	2007	2008	2006-2008
Number of Hospitals	4,666	4,631	4,614	4,760
Hospital Volume				
Mean Number of Admissions (SD)	86.5 (98.0)	80.2 (91.4)	76.8 (88.2)	237.2 (274.3)
Range (min. – max.)	1 – 1,043	1 - 900	1 - 863	1 – 2,806
25 th Percentile	20	18	17	51
50 th Percentile	52	47	44	138
75 th Percentile	119	113	107	332
RSMR (%) (Percentiles below weighted by hospital volume)				
Mean (SD)	10.9 (1.4)	11 (1.3)	11.2 (1.4)	10.9 (1.6)
Range (min. – max.)	7 - 17.4	7.5 - 17	7.8 - 17	6.4 - 19.4
25 th Percentile	10	10.1	10.3	9.8
50 th Percentile	10.8	10.9	11.1	10.8
75 th Percentile	11.7	11.8	12	11.9
Between Hospital Variance* (SE)	0.056 (0.0038)	0.054 (0.0040)	0.054 (0.0041)	0.053 (0.0024)

* Results from hierarchical model

Figure 4– Distribution of Hospital 30-Day RSMRs for HF in Three-Year Combined Data (2006-2008)



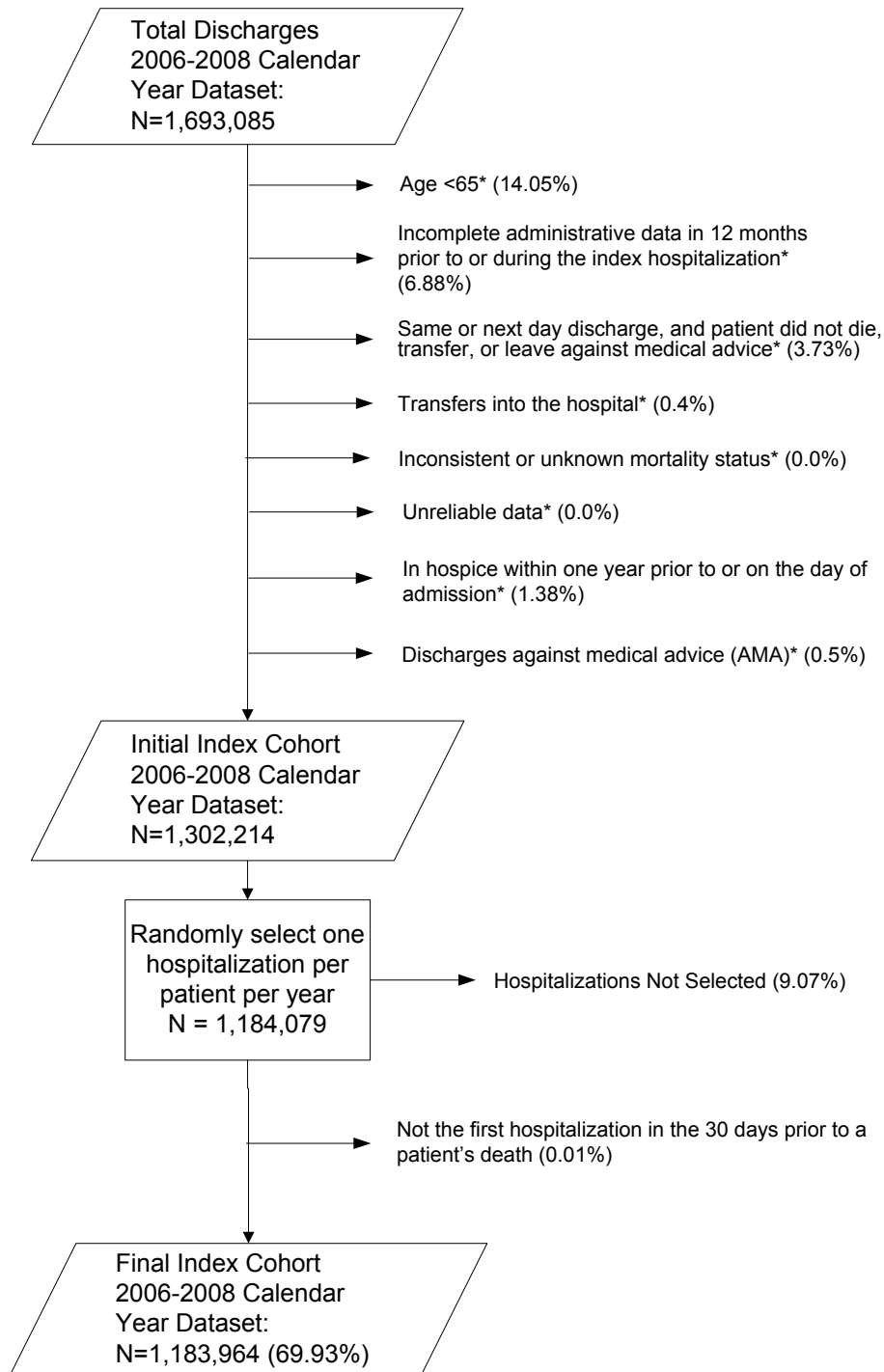
3.4 2010 PN Mortality Model

3.4.1 Index Cohort

The cohort includes admissions for Medicare FFS beneficiaries aged ≥ 65 years discharged from the hospital with a principal discharge diagnosis of PN (ICD-9-CM codes 480.0, 480.1, 480.2, 480.3, 480.8, 480.9, 481, 482.0, 482.1, 482.2, 482.30, 482.31, 482.32, 482.39, 482.40, 482.41, 482.49, 482.81, 482.82, 482.83, 482.84, 482.89, 482.9, 483.0, 483.1, 483.8, 485, 486, and 487.0) and with a complete claims history for 12 months prior to admission.

The exclusion criteria for the measures are presented in Section 1.2, and the percentage of PN patients meeting each exclusion criteria in the 2006-2008 calendar year dataset is presented in Figure 5.

Figure 5 – Patient Sample for PN in the 2006-2008 Calendar Year Dataset



*Categories are not mutually exclusive

3.4.2 Frequency of PN Model Variables over Different Time Periods

We examined the temporal variation in frequency of clinical and demographic variables. The crude mortality rate increased slightly from 11.4% in 2006 to 11.9% in 2008 (Table 7). The only notable changes among comorbid conditions were in: hypertension, the percentage of PN patients with hypertension increased from 77.4% in 2006 to 81.2% in 2008; iron deficiency, which also increased from 45.9% in 2006 to 59.1% in 2008; and COPD which decreased from 56.4% in 2006 to 52.2% in 2008.

3.4.3 Model Parameters

Table 8 conveys the risk-adjusted ORs and 95% CIs for the PN model by time period. The parameters are consistent across all time periods. Age was most strongly correlated with risk of death (OR 1.05; 95% CI 1.05-1.06). History of PTCA and CABG, unstable angina, chronic atherosclerosis, hypertension, cerebrovascular disease, COPD, depression and asthma were inversely associated with risk of death. All other variables were associated with an increased risk of death. Model performance was stable over all time periods; the area under the ROC curve was 0.72 across all time periods.

3.4.4 Distribution of Hospital RSMRs

Table 9 shows the distributions of hospital volume, RSMR and between-hospital variance by time period. Mean PN volume across hospitals decreased from 91 admissions (SD: 85) in 2006, to 79 admissions (SD: 74) in 2008. RSMR increased slightly from 11.3% in 2006 to 11.8% in 2008. The mean hospital RSMR for the combined three-year data was 11.4 (SD: 2.0; range 6.8 – 20.7. The 25th and 75th percentiles were 10% and 12.5% respectively in the combined three-year dataset. Between-hospital variance remained stable across years ranging from 0.070 (SE: 0.004) -0.074 (SE: 0.004). Between-hospital variance in the combined, three-year dataset was 0.078 (SE: 0.003). If there were no systematic differences between hospitals, the between-hospital variance would be 0.

Figure 6 shows the overall distribution of the RSMRs based on three-year combined data. The odds of all-cause mortality for a patient treated at a hospital that was one standard deviation above the national average was 1.68 times higher than that of a patient treated at a hospital that was one standard deviation below the national average. If there were no systematic differences between hospitals, the odds of all-cause mortality would be 1.0.¹⁰

Table 7 – Distribution of PN Model Variables over Different Time Periods

Description	2006	2007	2008	2006-2008
Total Number of Admissions	425,823	386,946	371,310	1,183,964
Crude Mortality Rate (%)	11.35	11.37	11.83	11.5
Demographics				
Mean Age (SD)	80.2 (8.01)	80.3 (8.08)	80.6 (8.12)	80.3 (8.07)
Male (%)	44.11	44.27	44.07	44.15
Cardiovascular (%)				
History of PTCA	3.36	3.34	3.23	3.31
History of CABG	5.01	4.75	4.42	4.74
Congestive heart failure (CC 80)	38.28	38.42	37.84	38.18
Acute myocardial infarction (CC 81)	3.50	3.48	3.73	3.57
Unstable angina (CC 82)	6.68	6.38	6.14	6.41
Chronic atherosclerosis (CC 83, 84)	46.36	46.72	46.61	46.56
Cardio-respiratory failure and shock (CC 79)	14.89	16.57	16.59	15.97
Comorbidity (%)				
Hypertension (CC 89, 91)	77.37	79.52	81.18	79.27
Stroke (CC 95, 96)	10.54	10.37	10.15	10.36
Cerebrovascular disease (CC 97-99, 103)	19.97	20.68	20.95	20.51
Renal failure (CC 131)	16.99	19.66	21.37	19.24
COPD (CC 108)	56.38	56.43	52.18	55.08
Pneumonia (CC 111-113)	41.65	42.14	41.96	41.91
Protein-calorie malnutrition (CC 21)	8.31	8.79	10.21	9.06
Dementia and senility (CC 49, 50)	27.78	28.08	28.88	28.22
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177, 178)	7.51	7.52	7.98	7.66
Peripheral vascular disease (CC 104, 105)	27.06	28.21	28.59	27.91
Metastatic cancer, acute leukemia, and other severe cancers (CC 7,8)	8.13	8.54	8.62	8.41
Trauma in last year (CC 154-156, 158-162)	35.20	35.65	36.03	35.61
Major psych disorders (CC 54-56)	11.38	11.68	12.09	11.70
Chronic liver disease (CC 25-27)	1.41	1.44	1.46	1.43
Severe hematological disorders (CC 44)	3.64	3.63	4.13	3.79
Iron deficiency/anemias/blood disease (CC 47)	45.87	47.51	49.07	47.41
Depression (CC 58)	16.37	16.71	16.90	16.64
Parkinson's/Huntington's diseases (CC 73)	4.26	4.16	4.24	4.22
Seizure disorders and convulsions (CC 74)	5.79	5.67	5.41	5.63
Fibrosis of lung and other chronic lung disorders (CC 109)	16.52	16.58	15.80	16.32
Asthma (CC 110)	11.51	11.31	10.91	11.26
Vertebral fractures (CC 157)	5.02	5.08	5.17	5.09

Table 8 – Adjusted ORs and 95% CIs for the PN HGLM over Different Time Periods

Variable	2006 OR	2006 95% CI	2007 OR	2007 95% CI	2008 OR	2008 95% CI	2006-2008 OR	2006-2008 95% CI
Demographics								
Age-65 ^a	1.05	(1.05-1.06)	1.05	(1.05-1.06)	1.05	(1.05-1.05)	1.05	(1.05-1.05)
Male	1.24	(1.22-1.27)	1.24	(1.21-1.26)	1.18	(1.16-1.21)	1.22	(1.21-1.24)
Cardiovascular								
History of PTCA	0.60	(0.56-0.65)	0.56	(0.52-0.61)	0.65	(0.60-0.70)	0.61	(0.58-0.63)
History of CABG	0.64	(0.60-0.67)	0.61	(0.58-0.65)	0.65	(0.61-0.69)	0.63	(0.61-0.66)
Congestive heart failure (CC 80)	1.30	(1.27-1.32)	1.29	(1.26-1.32)	1.33	(1.30-1.36)	1.30	(1.29-1.32)
Acute myocardial infarction (CC 81)	1.36	(1.29-1.43)	1.31	(1.24-1.38)	1.27	(1.21-1.34)	1.31	(1.27-1.35)
Unstable angina (CC 82)	0.94	(0.90-0.98)	1.00	(0.95-1.04)	0.95	(0.91-0.99)	0.96	(0.94-0.99)
Chronic atherosclerosis (CC 83, 84)	0.90	(0.88-0.92)	0.90	(0.88-0.92)	0.90	(0.88-0.92)	0.90	(0.89-0.91)
Cardio-respiratory failure and shock (CC 79)	1.29	(1.25-1.32)	1.28	(1.24-1.31)	1.25	(1.22-1.29)	1.27	(1.25-1.29)
Comorbidity								
Hypertension (CC 89, 91)	0.79	(0.77-0.80)	0.78	(0.76-0.80)	0.79	(0.77-0.81)	0.78	(0.77-0.79)
Stroke (CC 95, 96)	1.11	(1.08-1.15)	1.10	(1.07-1.14)	1.09	(1.05-1.12)	1.10	(1.08-1.12)
Cerebrovascular disease (CC 97-99, 103)	0.92	(0.89-0.94)	0.95	(0.92-0.97)	0.93	(0.90-0.95)	0.93	(0.92-0.94)
Renal failure (CC 131)	1.23	(1.20-1.27)	1.21	(1.18-1.24)	1.20	(1.17-1.23)	1.22	(1.20-1.24)
COPD (CC 108)	0.98	(0.96-1.00)	0.94	(0.92-0.96)	0.97	(0.95-0.99)	0.97	(0.95-0.98)
Pneumonia (CC 111-113)	1.11	(1.09-1.13)	1.09	(1.07-1.12)	1.10	(1.08-1.13)	1.10	(1.08-1.11)
Protein-calorie malnutrition (CC 21)	2.18	(2.12-2.24)	2.18	(2.12-2.25)	2.14	(2.09-2.20)	2.18	(2.15-2.22)
Dementia and senility (CC 49, 50)	1.53	(1.50-1.56)	1.50	(1.47-1.54)	1.52	(1.48-1.55)	1.51	(1.50-1.53)
Hemiplegia, paraplegia, paralysis, functional disability (CC 67-69, 100-102, 177, 178)	1.21	(1.17-1.25)	1.20	(1.16-1.25)	1.24	(1.19-1.28)	1.21	(1.19-1.24)
Peripheral vascular disease (CC 104, 105)	1.07	(1.05-1.09)	1.05	(1.03-1.08)	1.05	(1.02-1.07)	1.06	(1.05-1.07)
Metastatic cancer, acute leukemia, and other severe cancers (CC 7,8)	3.20	(3.11-3.29)	3.26	(3.16-3.36)	3.17	(3.08-3.27)	3.22	(3.17-3.28)
Trauma in last year (CC 154-156, 158-162)	1.10	(1.08-1.12)	1.08	(1.06-1.10)	1.06	(1.04-1.08)	1.08	(1.07-1.09)
Major psych disorders (CC 54-56)	1.10	(1.07-1.14)	1.11	(1.07-1.14)	1.12	(1.09-1.16)	1.11	(1.09-1.13)
Chronic liver disease (CC 25-27)	1.47	(1.37-1.58)	1.46	(1.35-1.57)	1.44	(1.34-1.56)	1.46	(1.40-1.53)
Severe hematological disorders (CC 44)	1.30	(1.25-1.36)	1.29	(1.23-1.35)	1.30	(1.24-1.36)	1.30	(1.27-1.34)
Iron deficiency/anemias/blood disease (CC 47)	1.07	(1.05-1.10)	1.06	(1.04-1.08)	1.06	(1.04-1.08)	1.07	(1.05-1.08)
Depression (CC 58)	0.96	(0.94-0.99)	0.97	(0.94-1.00)	0.94	(0.92-0.97)	0.96	(0.94-0.97)
Parkinson's/Huntington's diseases (CC 73)	1.19	(1.14-1.24)	1.11	(1.06-1.16)	1.14	(1.09-1.19)	1.15	(1.12-1.18)
Seizure disorders and convulsions (CC 74)	1.03	(0.99-1.07)	1.04	(1.00-1.08)	0.97	(0.93-1.01)	1.01	(0.99-1.04)
Fibrosis of lung and other chronic lung disorders (CC 109)	1.07	(1.04-1.10)	1.06	(1.03-1.09)	1.08	(1.05-1.11)	1.08	(1.06-1.09)
Asthma (CC 110)	0.66	(0.63-0.68)	0.66	(0.64-0.69)	0.66	(0.64-0.69)	0.66	(0.65-0.68)
Vertebral fractures (CC 157)	1.22	(1.17-1.27)	1.21	(1.16-1.26)	1.24	(1.19-1.29)	1.22	(1.19-1.25)
Model Performance^b								
Adjusted R ²		0.12		0.12		0.11		0.12
c- statistic		0.72		0.72		0.72		0.72

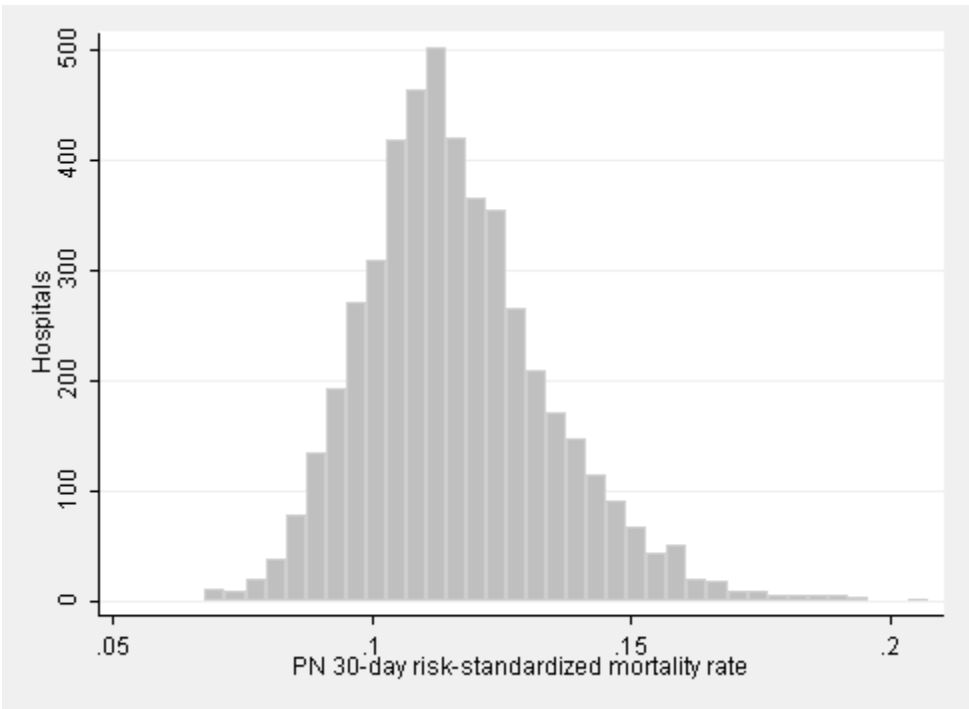
^a Patients' age – 65^b Obtained from GLM

Table 9 – Distribution of Hospital Volume and RSMR in PN Cohort over Different Time Periods

Characteristic	2006	2007	2008	2006-2008
Number of Hospitals	4,699	4,695	4,677	4,804
Hospital Volume				
Mean Number of Admissions (SD)	90.6 (84.8)	82.4 (77.3)	79.4 (73.5)	246.5 (232.97)
Range (min. – max.)	1 ~ 1,009	1 - 838	1 - 633	1 ~ 2,478
25 th Percentile	31	27	28	81
50 th Percentile	65	59	58	178
75 th Percentile	126	113	108	343
RSMR (%) (Percentiles below weighted by hospital volume)				
Mean (SD)	11.3 (1.7)	11.3(1.7)	11.8 (1.7)	11.4 (1.9)
Range (min. – max.)	7.2 - 18.2	6.7 - 18.8	6.7 - 18.6	6.8 - 20.7
25 th Percentile	10.1	10.2	10.7	10.0
50 th Percentile	11.1	11.1	11.7	11.2
75 th Percentile	12.3	12.3	12.7	12.5
Between Hospital Variance* (SE)	0.074 (0.0041)	0.073 (0.0043)	0.0698 (0.0043)	0.0678 (0.0026)

* Results from hierarchical model

Figure 6 – Distribution of Hospital 30-Day RSMRs for PN in Three-Year Combined Data (2006-2008)



4. SAS PACK

4.1 Revision to SAS Pack

The basic structure of the SAS pack remains the same from the previous maintenance year. The Mortality Measure SAS pack is comprised of a master program (Mortality-2010-three-year-01-19-2010.sas) and a program of SAS macros (Mymacro-2010-BP-v1.sas) designed by following the fundamental concept based on the object-oriented programming paradigm. This allows for easier use for creating the current measures and for a more efficient means of making any future modifications. The revised SAS pack and its corresponding documentation are available from CMS upon request at mortalitymeasures@mathematica-mpr.com.

5. QUALITY ASSURANCE

5.1 QA for Input Data

In measure maintenance, we have a three-phase approach to internal QA for the mortality measures (see Figure 7 below).

Phase I

The first step in the QA process is to ensure the validity of the input data files. Although there were no substantive changes in the formatting of the input data source this year, there were a number of improvements to the data extraction process that needed to be targeted for QA purposes. In general, all condition-specific files for each reporting year are evaluated by comparing to prior year QA results for the same condition/year. We conduct data validity checks including cross checking of death information, distributions of ICD-9 codes, and frequencies of key variables. We employ both manual scan and descriptive analysis to carry out these tasks. The results are reviewed for accuracy and trend over time as compared to prior datasets. Any new variable constructs and other changes in formatting to the input files are also verified as part of this process. We share our QA findings with our data extraction contractor as needed.

To assure accuracy in SAS coding, two analysts independently write SAS code for all steps in calculating the mortality measures: data preparation, sample selection and hierarchical modeling. This process will highlight any programming errors in syntax or logic. Once the parallel programming process are complete, the analysts cross-check their codes by analyzing datasets in parallel and checking for consistency of output and reconciling any inconsistencies.

Phase II

A third analyst reviews the finalized SAS code and recommends changes to the coding and readability of the SAS pack, where appropriate. The primary analyst receives the suggested changes for possible re-coding or program documentation.

This phase also includes a comparison of prior year risk-adjustment coefficients and variable frequencies. This enables us to check for potential inconsistencies in the data as well as any changes to the SAS pack.

Phase III

This phase involves a manual check of the results produced from the SAS code. For our test dataset, we randomly select 100 patients in the study sample for each condition. Two research assistants check the inclusion and exclusion algorithms to be sure they correctly identified the cohort to be included in the final measures. To test our algorithms, we check the raw data and the inclusion/exclusion status of each admission of the patients in our test dataset reported by the final measure against the inclusion and exclusion criteria. We also verify that the risk factor evaluation by the models was consistent with the input data.

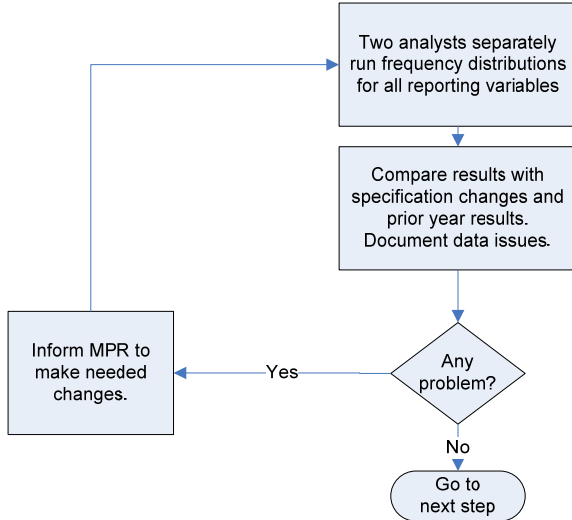
For risk factor evaluation, we investigate and verify the mapping of CCs, grouping of CC variables, and coding of complications for each patient in our test dataset. Two research assistants and one analyst working independently, manually check the coding of risk factor comorbidities in the SAS master file using the history of coding for the patient in the SAS diagnosis and procedure history file. The research assistants check 25 randomly chosen patients in each condition, checking for 'false positives' in the coding. False positives are instances where comorbidities coded positively in the master file are not supported by codes in the SAS diagnosis and procedure history files. Instances of false positives are reported to the analyst, who then studies the codes further. By manually checking the model results against the raw data, we further target any inconsistent programming logic for revision.

(This section represents QA for the subset of the work conducted by YNHHS/CORE to maintain and report these mortality measures. It does not describe the QA to process data and create the input files nor does it include QA for the final processing of production data for public reporting done by other CMS contractors.)

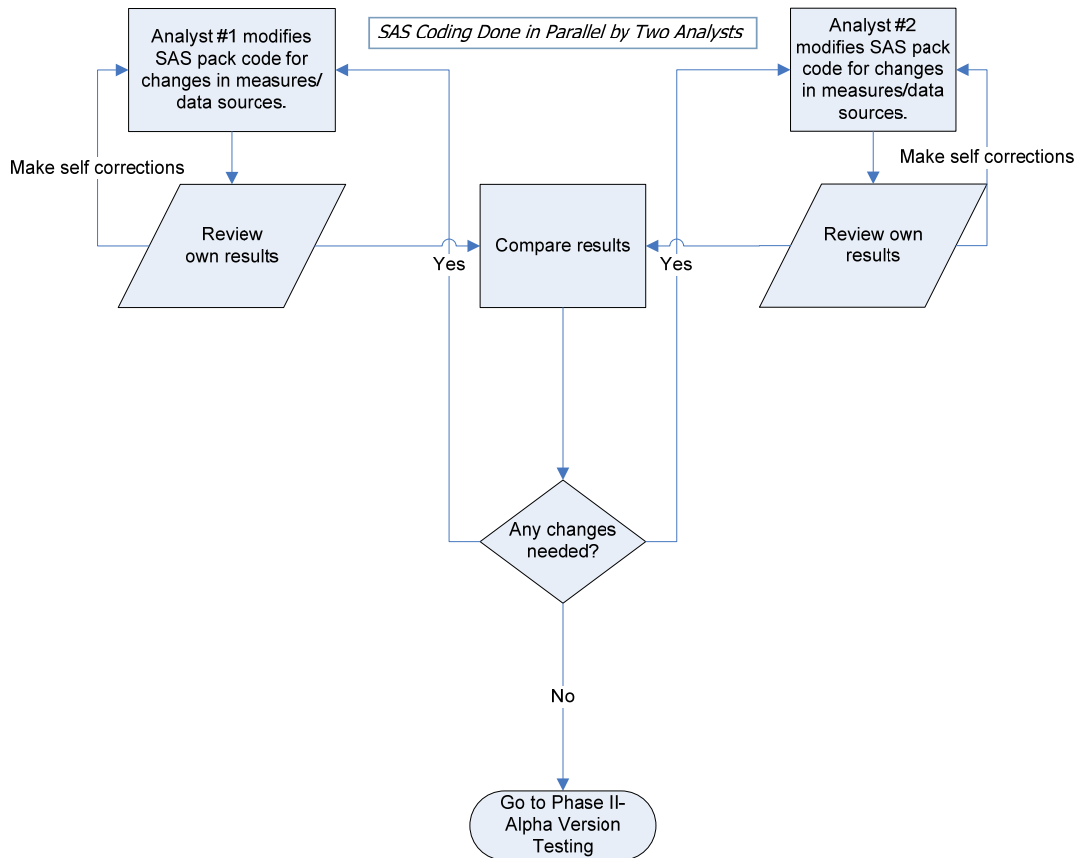
Figure 7 – YNHHS-CORE QA Processes

Phase I

Pre SAS Package Processing Input Data QA

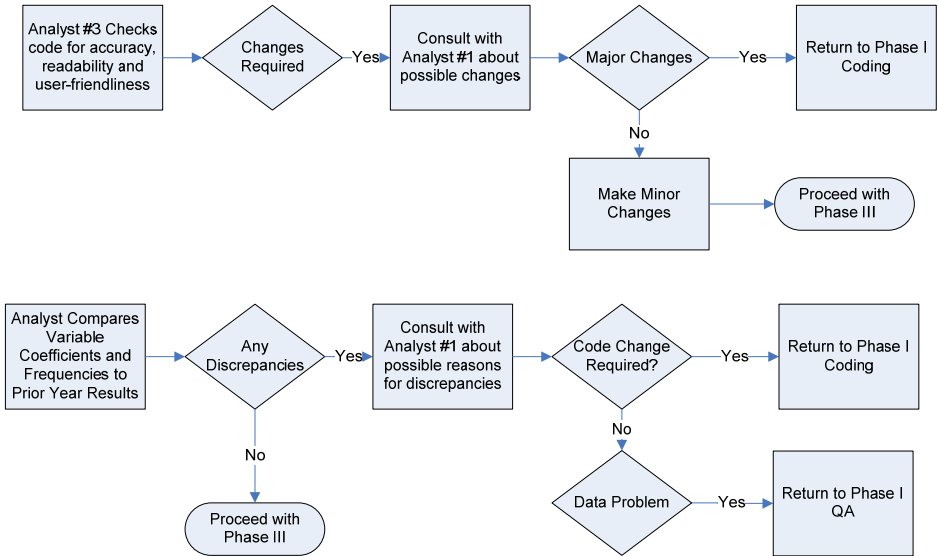


SAS Package QA



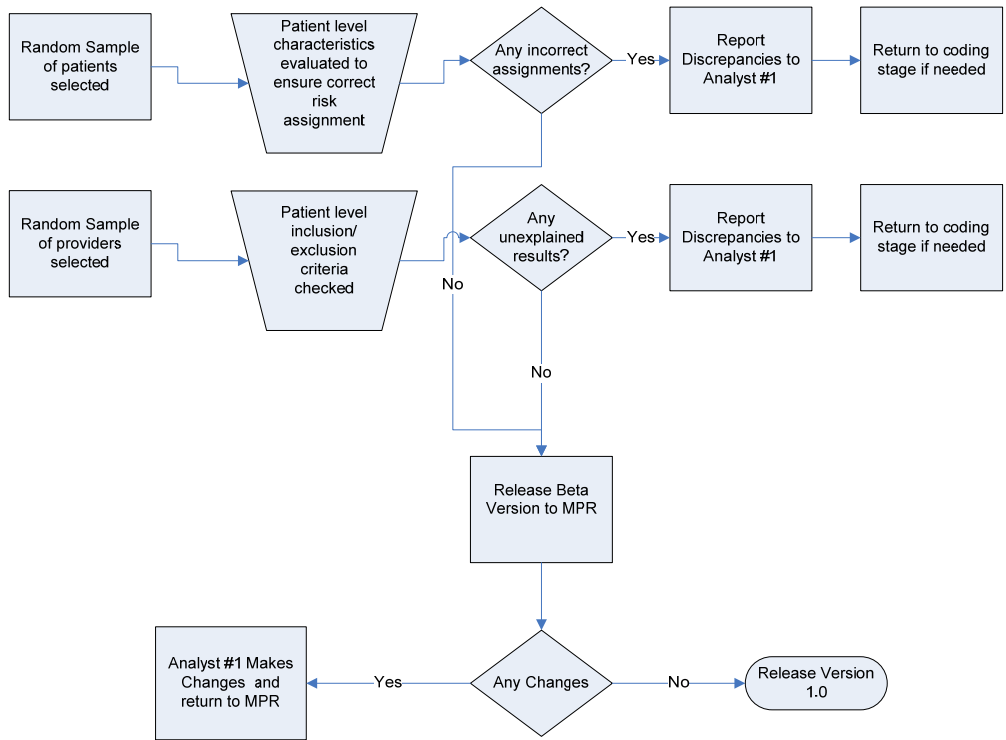
Phase II

Results Testing – Alpha Version



Phase III

Results Testing – Individual Patient Level



6. REFERENCES

1. Krumholz HM, Brindis RG, Brush JE, Cohen DJ, Epstein AJ, Furie K, Howard G, Peterson ED, Rathore SS, Smith SC Jr, Spertus JA, Wang Y, Normand S-LT. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation* 113: 456-462. 2006.
2. Krumholz HM, Normand S-LT, Galusha DH, Mattera JA, Rich AS, Wang YF, Wang Y. et al. Risk-Adjustment Models for AMI and HF: 30-Day Mortality: Report prepared for the Centers for Medicare & Medicaid Services; 2005. Available at: <http://www.qualitynet.org/>
3. Krumholz HM, Normand S-LT, Bratzler DW, Mattera JA, Rich AS, Wang YF, Wang Y. Risk-Adjustment Methodology for Hospital Monitoring/Surveillance and Public Reporting Supplement #1: 30-Day Mortality Model for Pneumonia: Report prepared for the Centers for Medicare & Medicaid Services; 2006. Available at: <http://www.qualitynet.org/>
4. Krumholz HM, Normand S-LT, Bhat KR, Drye EE, Schreiner GC, Wang Y, Wang Y. Acute Myocardial Infarction, Heart Failure, and Pneumonia Mortality Measures Maintenance Technical Report: Report prepared for the Centers for Medicare & Medicaid Services; 2008. Available at: <http://www.qualitynet.org/>
5. Krumholz HM, Normand S-LT, Drye EE, Bhat KR, Schreiner GC, Wang YF, Wang Y. Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-Day Risk-Standardized Mortality Measures Maintenance Technical Report: Report prepared for the Centers for Medicare & Medicaid Services; 2009. Available at: <http://www.qualitynet.org/>
6. Krumholz HM, Wang Y, Mattera JA, Wang YF, Han LF, Ingber MJ, Roman S, Normand S-LT. An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With an Acute Myocardial Infarction. *Circulation* 113(13): 1683-1692. 2006.
7. Krumholz HM, Wang Y, Mattera JA, Wang YF, Han LF, Ingber MJ, Roman S, Normand S-LT. An Administrative Claims Model Suitable for Profiling Hospital Performance Based on 30-Day Mortality Rates Among Patients With Heart Failure. *Circulation* 113(13): 1693-1701. 2006.

8. Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. *Stat Sci* 22 (2): 206-226.
9. Normand S-LT, Wang Y, Krumholz HM. Assessing surrogacy of data sources for institutional comparisons. *Health Services and Outcomes Research Methodology* 7: 79-96. 2007.
10. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Chichester, England: John Wiley & Sons Ltd.; 2004.

7. APPENDIX

MEMORANDUM

From: RTI International
To: CMS/OCSQ
Date: November 25, 2009

Subject: Overview of update of mappings of ICD-9-CM codes to CC groups for risk adjustment of hospital mortality and readmission models, changes related to FY2009 codes. This is in relation to creating a mapping covering FY2005 – FY2009.

Overview

Each year the CDC National Center for Health Statistics and the Centers for Medicare & Medicaid Services oversee the changes and modifications to the ICD-9-CM system made through the Coordination and Maintenance Committee. The committee is a joint public-private effort to update and improve the coding system.

RTI has developed and supported a classification system that uses these codes as the basis for risk adjustment systems. The Hierarchical Condition Category (HCC) system groups the ICD-9-CM codes into larger groups that are used in a model to predict medical care utilization, spending, mortality or other related measures. The condition categories (CCs) may also be used without the hierarchies that are used to characterize a person's medical conditions into the highest severity category of a set of related conditions. For this project the full set of 189 CCs were updated for FY2009 changes and the changes documented.

New ICD-9 codes generally become effective October 1 of each year, though there is a round of changes that may be made in an April announcement. Each calendar year of diagnosis data encompasses 2 years of codes. In the new mappings codes valid in FY2006 through FY2009 are all mapped to CCs. This allows the mapping to fully cover data from October 1, 2005 through September 30, 2009. These codes span CY2006 through CY2008 and the first nine months of 2009. The last three months of 2009 fall into FY2010.

Method

Additions and deletions

When the code changes are announce each year there may be both additions, deletions and changes to the descriptions of codes. We map only the valid codes, those of highest specificity, each year. ICD-9-CM codes have a minimum of three characters, mostly digits, and a maximum of five characters. The form is NNN, NNN.N or NNN.NN.² Code numbers after the decimal point are subclasses of the 3-digit main classes. An addition of new codes may be at

² In the Medicare data and our mappings the decimal points are omitted and all codes are left justified to remove ambiguity.

any level from a new 3-digit class to new 4th and 5th digit subclasses. Deletions from ICD-9 may be explicit, the removal of a code from the code book. But deletions from our mapping may occur because new more specific subcodes are introduced. Introduction of a new code of higher specificity than the old code it sprang from does not remove the original 3- or 4- digit code from the ICD-9 book, but since coding is supposed to be done to the highest specificity, we remove the more general code from the mappings for that year. If the new high specificity code is just an addition to an existing subset of codes of similar specificity, the new code is added. There would be no change in the status of the more general code. That code would have already have been superseded by higher specificity codes.

As an example, in 2009, code 046.1 *Jakob-Creutzfeldt disease*, was split into 046.11, *Variant Creutzfeldt-Jakob disease*, and 046.19, *Other and unspecified Creutzfeldt-Jakob disease*. The two new 5-digit codes were added to our mapping and were assigned to the same CC that *Jakob-Creutzfeldt disease* was assigned to. The old 4-digit code would have been removed, except that 046.1 was valid in 2006, 2007 and 2008. Since our mapping is intended to allow valid codes from those years, 046.1 was retained.

In 2009 there were many 5-digit codes added to ICD-9. Twenty-seven were added to the cancer codes in the 203.XX – 208.XX series with the last digit 2. The last digit 2 signifies *in relapse*. The codes in this series ending with a fifth digit 0 were all relabeled to indicate *without mention of having achieved remission*. A new series of neuroendocrine tumors was added, 209.XX. There are 43 5-digit codes in this series, malignant and benign, specific to anatomical sites. Another example of new codes is a set of 20 5-digit codes indicating Secondary Diabetes Mellitus similar to the existing set of diabetes codes in the inclusion or exclusion of complications.

There were 374 new codes added at the highest level of specificity. In some cases new code clusters at the 3- or 4-digit level were added, containing 4- or 5-digit splits. There are 25 codes that are no longer at the highest specificity and invalid for FY2009. However, they are retained in the mapping because they were valid in the three years prior to FY2009.

Mapping

Mapping of the new codes is done by review of the changes by RTI staff and clinical consultants. In most cases the codes of higher specificity are mapped to the same CC as the more general code that was split. This does not always occur. These are some examples:

In the case of Secondary Diabetes there were many 5-digit codes added. These were all assigned to the same CCs as the Primary Diabetes, uncomplicated secondary with uncomplicated primary and each secondary with a complication assigned to the same CC as the primary with the same complication. It was deemed that the implications of the diabetes itself are similar, whether primary or secondary.

The code 337.0 relates to *Idiopathic peripheral autonomic neuropathy*. It was split into 5-digit codes that did not all get mapped to one place. Code 337.00 is *Idiopathic peripheral autonomic neuropathy, unspecified*. It was placed in the CC for Polyneuropathy, the CC in which the parent code had been mapped. Code 337.01, *Carotid sinus syndrome*, was placed in

the CC for Other Heart Rhythm and Conduction Disorders. It is distinct from the others in the group. The last new code in the series, 337.09, *Other idiopathic peripheral autonomic neuropathy* was grouped with Polyneuropathy. Code 337.0 was retained because it was still valid in earlier years included in the data.

New codes for Dural tear were added as a new group, 349.3X. In this case 349.31, *Accidental puncture or laceration of dura during a procedure*, was assigned to the CC for Other Complications of Medical Care. The other code in the group, 349.39, *Other dural tear*, has no indication of being a complication and was assigned to the CC for Nonpsychotic Organic Brain Syndromes/Conditions. The root code 349.3 is not mapped, as it did not exist in a prior year and is not the highest specificity code for the group.

There are cases in which it is clear where conditions may have been mapped prior to the new code and cases in which it is not. A new set of codes 209.XX, neuroendocrine tumors were assigned to a cancer CC if malignant or an other neoplasm CC, if benign. It is not clear how all these conditions were coded in the past. Some might have been coded in the cancer group and some in the CC for significant endocrine disorders. We have annotated both the CC assignments of the new codes and the CCs in which these conditions may have been coded in prior years.

The general practice in maintaining the mappings for this work has been to maintain the existing structure of the CCs and to map the new codes to the location they would have gone to in prior years. However, sometimes the new specificity makes clear enough distinctions that new related codes do not all logically go to one place. And some new codes create puzzles of their own that require judgment calls to be made. Our decision committee brings together both the people who maintain the integrity of the system and the people who provide the clinical expertise. The changes for FY2009 did not create a need for major changes.

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 #: 0282	NQF Project: Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION	
De.1 Measure Title: Angina without procedure (PQI 13)	
De.2 Brief description of measure: All non-maternal discharges of age 18 years and older with ICD-9-CM principal diagnosis code for angina.	
1.1-2 Type of Measure: Access	
De.3 If included in a composite or paired with another measure, please identify composite or paired measure Prevention Quality Indicator Composite (overall and for chronic conditions)	
De.4 National Priority Partners Priority Area: Population health, Safety	
De.5 IOM Quality Domain: Effectiveness	
De.6 Consumer Care Need: Staying healthy	

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
<p>A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. <i>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.</i></p> <p>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</p> <p>A.2 Indicate if Proprietary Measure (as defined in measure steward agreement):</p> <p>A.3 Measure Steward Agreement: Government entity and in the public domain - no agreement necessary</p> <p>A.4 Measure Steward Agreement attached:</p>	<p>A</p> <p>Y <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
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	<p>B</p> <p>Y <input type="checkbox"/></p>

every 3 years. Yes, information provided in contact section	N <input type="checkbox"/>
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ► Purpose: Public reporting , Internal quality improvement	C Y <input type="checkbox"/> N <input type="checkbox"/>
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 <input type="checkbox"/> N <input type="checkbox"/>
(for NQF staff use) Have all conditions for consideration been met? Staff Notes to Steward (if submission returned):	Met Y <input type="checkbox"/> N <input type="checkbox"/>
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. <i>Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)</i> 1a. High Impact	Eval Ratin g
(for NQF staff use) Specific NPP goal :	
1a.1 Demonstrated High Impact Aspect of Healthcare: Patient/societal consequences of poor quality 1a.2 1a.3 Summary of Evidence of High Impact: Admissions for angina are common and there increasing evidence that the rate of angina admissions is partially a function of the quality of care in a community. Stable angina can be managed in an outpatient setting using drugs such as aspirin and beta blockers, as well as advice to change diet and exercise habits. Effective treatments for coronary artery disease reduce admissions for serious complications of ischemic heart disease, including unstable angina. 1a.4 Citations for Evidence of High Impact: Gibbons RJ, Chatterjee K, Daley J, et al. ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task force on Practice Guidelines (Committee on Management of Patients with Chronic Stable Angina) [published erratum appears in J Am Coll Cardiol 1999 Jul;34(1):314]. J Am Coll Cardiol 1999;33(7):2092-197. Blustein J, Hanson K, Shea S. Preventable hospitalizations and socioeconomic status. Health Aff (Millwood) 1998;17(2):177-89. Brunwald E, Antman EM, Beasley JW et al. ACC/AHA guidelines for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients with Unstable Angina). J Am Coll Cardiol 2000;36(3):970-1062.	1a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>

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).

<p>Billings J, Zeital L, Lukomnik J, et al. Analysis of variation in hospital admission rates associated with area income in New York City. Unpublished report.</p> <p>Millman M, editor. Committee on Monitoring Access to Personal Health Care Services. Washington, DC: National Academy Press. 1993.</p>	
<p>1b. Opportunity for Improvement</p> <p>1b.1 Benefits (improvements in quality) envisioned by use of this measure: Providers can implement processes of care to reduce the likelihood of a hospital admission or the health system can implement system processes to improve access to high quality outpatient care</p> <p>1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers: 5th 25th Median 75th 95th 0.000000 0.000003 0.000043 0.000260 0.001208</p> <p>1b.3 Citations for data on performance gap: 2007 AHRQ State Inpatient Databases with 4,000 counties 57,000 numerator discharges</p> <p>1b.4 Summary of Data on disparities by population group: Based on the 2008 national statistics for angina without procedure (http://hcupnet.ahrq.gov) the 2008 rates are as follows: Overall rate per 100,000: 24.93; Risk adjusted rate: 24.05 Male: 24.42 Female: 25.42 Age groups: 18-39: 2.80; 40-64: 30.37; 65-74: 53.90; 75+: 74.27</p> <p>1b.5 Citations for data on Disparities: Nationwide Inpatient Sample, 2008</p>	<p>1b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>1c. Outcome or Evidence to Support Measure Focus</p> <p>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): Hospital admission is a proxy outcome for a decrease in health status</p> <p>1c.2-3. Type of Evidence: Systematic synthesis of research</p> <p>1c.4 Summary of Evidence (as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome): Hospital admission for angina is a PQI that would be of most interest to comprehensive health care delivery systems. Admission for angina is relatively common, suggesting that the indicator will be measured with good precision. The observed variation likely reflects true differences in area performance. Age-sex adjustment has a moderate impact. Other risk factors for consideration include smoking, hyperlipidemia, hypertension, diabetes, and socioeconomic status. The patient populations served by hospitals that contribute the most to the overall area rate for angina may be a starting point for interventions. As a PQI, angina without procedure is not a measure of hospital quality, but rather one measure of outpatient and other health care. This indicator has unclear construct validity, because it has not been validated except as part of a set of indicators. Providers may reduce admission rates without actually improving quality of care by shifting care to an outpatient setting. Some angina care takes place in emergency rooms. Combining inpatient and emergency room data may give a more accurate picture.</p>	<p>1c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

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])

Face validity: Stable angina can be managed in an outpatient setting using drugs such as aspirin and beta blockers, as well as advice to change diet and exercise habits.⁸⁴ Effective treatments for coronary artery disease reduce admissions for serious complications of ischemic heart disease, including unstable angina.

Precision: Reasonably precise estimates of area angina rates should be feasible, as one study shows that unstable angina accounts for 16.3% of total admissions for ACSCs.⁸⁵ Based on empirical evidence, this indicator is adequately precise, with a raw area level rate of 166.0 per 100,000 population and a standard deviation of 135.7.

The signal ratio (i.e., the proportion of the total variation across areas that is truly related to systematic differences in area performance rather than random variation) is very high, at 91.6%, indicating that the observed differences in age-sex adjusted rates likely represent true differences across areas. Using multivariate signal extraction techniques appears to have little additional impact on estimating true differences across areas.

Minimum bias: No evidence exists in the literature on the potential bias of this indicator. The incidence of angina is related to age structure and risk factors (smoking, hyperlipidemia, hypertension, diabetes) in a population. Elderly age (over 70), diabetes, and hypertension have also been associated with being at higher risk for angina.⁸⁶

Construct validity: Billings et al. found that low-income ZIP codes in New York City had 2.3 times more angina hospitalizations than high-income ZIP codes.⁸⁷ Household income explained 13% of this variation. In addition, Millman et al.⁸⁸ reported that low-income ZIP codes had 2.7 times more angina hospitalizations per capita than high-income ZIP codes. Based on empirical study, areas with high rates of angina admissions tend to have higher rates of other ACSC admissions.

Fosters true quality improvement: Use of this quality indicator might raise the threshold for admission of angina patients. Because some angina can be managed on an outpatient basis, a shift to outpatient care may occur but is unlikely for severe angina.

Prior use: This indicator was originally developed by Billings et al. in conjunction with the United Hospital Fund of New York.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom*): Not applicable

1c.6 Method for rating evidence: Not applicable

1c.7 Summary of Controversy/Contradictory Evidence: In a study of approximately 124,000 cancer-free Medicare beneficiaries/year, with subjects contributing data for 1-8 years, angina PQI hospital discharges declined 75% between 1992 and 1999. CAD hospital discharges rose in a reciprocal pattern, while angina discharges with revascularization declined and discharges for myocardial infarction and ischemic heart disease remained relatively constant. The authors conclude "The marked decline in angina PQI hospital discharges during 1992-1999 does not appear to represent improvements in access to care or prevention of heart disease, but rather increased coding of more specific discharge diagnoses for CAD. Our findings suggest that angina hospitalization is not a valid measure for monitoring access to care and, more generally, demonstrate the need for careful, periodic reevaluation of quality measures." [1]
 [1] Barry G. Saver; Sharon A. Dobie; Pamela K. Green; Ching-Yun Wang; Laura-Mae Baldwin. No Pain, but No Gain? The Disappearance of Angina Hospitalizations, 1992-1999. *Med Care*. 2009 October ; 47(10): 1106-1110. doi:10.1097/MLR.0b013e31819e1f53.

1c.8 Citations for Evidence (other than guidelines): [1] Gibbons RJ, Chatterjee K, Daley J, et al. ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task force on Practice Guidelines (Committee on Management of Patients with Chronic Stable Angina) [published erratum appears in *J Am Coll Cardiol* 1999 Jul;34(1):314]. *J Am Coll Cardiol* 1999;33(7):2092-197.
 [2] Blustein J, Hanson K, Shea S. Preventable hospitalizations and socioeconomic status. *Health Aff (Millwood)* 1998;17(2):177-89.

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.

<p>[3] Brunwald E, Antman EM, Beasley JW et al. ACC/AHA guidelines for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients with Unstable Angina). J Am Coll Cardiol 2000;36(3):970-1062.</p> <p>[4] Billings J, Zeital L, Lukomnik J, et al. Analysis of variation in hospital admission rates associated with area income in New York City. Unpublished report.</p> <p>[5] Millman M, editor. Committee on Monitoring Access to Personal Health Care Services. Washington, DC: National Academy Press. 1993.</p> <p>1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number): Patients and their close associates should be informed of the nature of angina pectoris, and the implications of the diagnosis and the treatments that may be recommended. The patient can be reassured that, in most cases, both the symptoms of angina and prognosis can be improved with proper management. Comprehensive risk stratification should be conducted as outlined above, and particular attention should be paid to the elements of lifestyle that could have contributed to the condition and which may influence prognosis, including physical activity, smoking, and dietary habits. The recommendations of the Third Joint European Societies' Task Force on Cardiovascular Disease Prevention in Clinical Practice should be followed.</p> <p>1c.10 Clinical Practice Guideline Citation: Guidelines on the management of stable angina pectoris. Fox K, Alonso Garcia MA, Ardissino D, Buszman P, Camici PG, Crea F, Daly C, DeBacker G, Hjendahl P, Lopez-Sendon J, Marco J, Morais J, Pepper J, Sechtem U, Simoons M, Thygesen K. Guidelines on the management of stable angina pectoris. Sophia Antipolis, France: European Society of Cardiology; 2006. 63 p. [683 references]</p> <p>1c.11 National Guideline Clearinghouse or other URL: http://www.guidelines.gov/content.aspx?id=9421</p> <p>1c.12 Rating of strength of recommendation (also provide narrative description of the rating and by whom): Not applicable</p> <p>1c.13 Method for rating strength of recommendation (If different from USPSTF system, also describe rating and how it relates to USPSTF): Not applicable</p> <p>1c.14 Rationale for using this guideline over others: Not applicable</p>	
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Importance to Measure and Report</i>?</p>	1
<p>Steering Committee: Was the threshold criterion, <i>Importance to Measure and Report</i>, met? Rationale:</p>	1 Y <input type="checkbox"/> N <input type="checkbox"/>
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
<p>Extent to which the measure, <u>as specified</u>, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)</p>	Eval Rating
2a. MEASURE SPECIFICATIONS	
<p>S.1 Do you have a web page where current detailed measure specifications can be obtained? S.2 If yes, provide web page URL:</p>	
<p>2a. Precisely Specified</p> <p>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): All discharges of age 18 years and older with ICD-9-CM principal diagnosis code for angina.</p>	2a-specs C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>

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.

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).

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator*):
Time window can be determined by user, but is generally a calendar year.

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions*):

Include ICD-9-CM diagnosis codes:

4111 INTERMED CORONARY SYND
41181 CORONARY OCCLSN W/O MI
41189 AC ISCHEMIC HRT DIS NEC
4130 ANGINA DECUBITUS
4131 PRINZMETAL ANGINA
4139 ANGINA PECTORIS NEC/NOS

Exclude cases:

- transfer from a hospital (different facility)
- transfer from a skilled Nursing Facility (SNF) or Intermediate Care Facility (ICF)
- transfer from another health care facility
- MDC 14 (pregnancy, childbirth, and puerperium)
- with a code for cardiac procedure

ICD-9-CM Cardiac procedure codes:

0050 IMPL CRT PACEMAKER SYS OCT02-
0051 IMPL CRT DEFIBRILLAT OCT02-
0052 IMP/REP LEAD LF VEN SYS OCT02-
0053 IMP/REP CRT PACEMKR GEN OCT02-
0054 IMP/REP CRT DEFIB GENAT OCT02-
0056 INS/REP IMPL SENSOR LEAD OCT06-
0057 IMP/REP SUBCUE CARD DEV OCT06-
0066 PTCA OCT06-1751 IMPLANTATION OF RECHARGEABLE CARDIAC CONTRACTILITY MODULATION [CCM],
TOTAL SYSTEM OCT09-
1752 IMPLANTATION OR REPLACEMENT OF CARDIAC CONTRACTILITY MODULATION [CCM] RECHARGEABLE
PULSE GENERATOR ONLY OCT09-
3500 CLOSED VALVOTOMY NOS
3501 CLOSED AORTIC VALVOTOMY
3502 CLOSED MITRAL VALVOTOMY
3503 CLOSED PULMON VALVOTOMY
3504 CLOSED TRICUSP VALVOTOMY
3510 OPEN VALVULOPLASTY NOS
3511 OPN AORTIC VALVULOPLASTY
3512 OPN MITRAL VALVULOPLASTY
3513 OPN PULMON VALVULOPLASTY
3514 OPN TRICUS VALVULOPLASTY
3520 REPLACE HEART VALVE NOS
3521 REPLACE AORT VALV-TISSUE
3522 REPLACE AORTIC VALVE NEC
3523 REPLACE MITR VALV-TISSUE
3524 REPLACE MITRAL VALVE NEC
3525 REPLACE PULM VALV-TISSUE
3526 REPLACE PULMON VALVE NEC
3527 REPLACE TRIC VALV-TISSUE
3528 REPLACE TRICUSP VALV NEC
3531 PAPILLARY MUSCLE OPS
3532 CHORDAE TENDINEAE OPS
3533 ANNULOPLASTY
3534 INFUNDIBULECTOMY
3535 TRABECUL CARNEAE CORD OP
3539 TISS ADJ TO VALV OPS NEC
3541 ENLARGE EXISTING SEP DEF

3542 CREATE SEPTAL DEFECT
 3550 PROSTH REP HRT SEPTA NOS
 3551 PROS REP ATRIAL DEF-OPN
 3552 PROS REPAIR ATRIA DEF-CL
 3553 PROST REPAIR VENTRIC DEF
 3554 PROS REP ENDOCAR CUSHION
 3555 PROS REP VENTRC DEF-CLOS OCT06-
 3560 GRFT REPAIR HRT SEPT NOS
 3561 GRAFT REPAIR ATRIAL DEF
 3562 GRAFT REPAIR VENTRIC DEF
 3563 GRFT REP ENDOCAR CUSHION
 3570 HEART SEPTA REPAIR NOS
 3571 ATRIA SEPTA DEF REP NEC
 3572 VENTR SEPTA DEF REP NEC
 3573 ENDOCAR CUSHION REP NEC
 3581 TOT REPAIR TETRAL FALLOT
 3582 TOTAL REPAIR OF TAPVC
 3583 TOT REP TRUNCUS ARTERIOS
 3584 TOT COR TRANSPOS GRT VES
 3591 INTERAT VEN RETRN TRANSP
 3592 CONDUIT RT VENT-PUL ART
 3593 CONDUIT LEFT VENTR-AORTA
 3594 CONDUIT ARTIUM-PULM ART
 3595 HEART REPAIR REVISION
 3596 PERC HEART VALVULOPLASTY
 3598 OTHER HEART SEPTA OPS
 3599 OTHER HEART VALVE OPS
 3601 PTCA-1 VESSEL W/O AGENT
 3602 PTCA-1 VESSEL WITH AGNT
 3603 OPEN CORONRY ANGIOPLASTY
 3604 INTRCORONRY THROMB INFUS
 3605 PTCA-MULTIPLE VESSEL
 3606 INSERT OF COR ART STENT OCT95-
 3607 INS DRUG-ELUT CORONRY ST OCT02-
 3609 REM OF COR ART OBSTR NEC
 3610 AORTOCORONARY BYPASS NOS
 3611 AORTOCOR BYPAS-1 COR ART
 3612 AORTOCOR BYPAS-2 COR ART
 3613 AORTOCOR BYPAS-3 COR ART
 3614 AORTOCOR BYPAS-4+ COR ART
 3615 1 INT MAM-COR ART BYPASS
 3616 2 INT MAM-COR ART BYPASS
 3617 ABD-CORON ART BYPASS OCT96-
 3619 HRT REVAS BYPS ANAS NEC
 362 ARTERIAL IMPLANT REVAS
 363 OTH HEART REVASCULAR
 3631 OPEN CHEST TRANS REVAS
 3632 OTH TRANSMYO REVASCULAR
 3633 ENDO TRANSMYO REVASCULAR OCT06-
 3634 PERC TRANSMYO REVASCULAR OCT06-
 3639 OTH HEART REVASULAR
 3691 CORON VESS ANEURYSM REP
 3699 HEART VESSLE OP NEC
 3731 PERICARDIECTOMY
 3732 HEART ANEURYSM EXCISION
 3733 EXC/DEST HRT LESION OPEN
 3734 EXC/DEST HRT LES OTHER
 3735 PARTIAL VENTRICULECTOMY

3736 EXCISION OR DESTRUCTION OF LEFT ATRIAL APPENDAGE (LAA) OCT08-
 3741 IMPLANT PROSTH CARD SUPPORT DEV OCT06
 375 HEART TRANSPLANTATION (NOT VALID AFTER OCT 03)
 375 HEART TRANPLANTATION OCT03-
 3752 IMPLANT TOT REP HRT SYS OCT03-
 3753 REPL/REP THORAC UNIT HRT OCT03-
 3754 REPL/REP OTH TOT HRT SYS OCT03-
 3755 REMOVAL OF INTERNAL BIVENTRICULAR HEART REPLACEMENT SYSTEM OCT08-
 3760 IMPLANTATION OR INSERTION OF BIVENTRICULAR EXTERNAL HEART ASSIST SYSTEM OCT08-
 3761 IMPLANT OF PULSATION BALLOON
 3762 INSERTION OF NON-IMPLANTABLE HEART ASSIST SYSTEM
 3763 REPAIR OF HEART ASSIST SYSTEM
 3764 REMOVAL OF HEART ASSIST SYSTEM
 3765 IMPLANT OF EXTERNAL HEART ASSIST SYSTEM
 3766 INSERTION OF IMPLANTABLE HEART ASSIST SYSTEM
 3770 INT INSERT PACEMAK LEAD
 3771 INT INSERT LEAD IN VENT
 3772 INT INSERT LEAD ATRI-VENT
 3773 INT INSER LEAD IN ATRIUM
 3774 INT OR REPL LEAD EPICAR
 3775 REVISION OF LEAD
 3776 REPL TV ATRI-VENT LEAD
 3777 REMOVAL OF LEAD W/O REPL
 3778 INSER TEAM PACEMAKER SYS
 3779 REVIS OR RELOCATE POCKET
 3780 INT OR REPL PERM PACEMKR
 3781 INT INSERT 1-CHAM, NON
 3782 INT INSERT 1-CHAM, RATE
 3783 INT INSERT DUAL-CHAM DEV
 3785 REPL PACEM W 1-CHAM, NON
 3786 REPL PACEM 1-CHAM, RATE
 3787 REPL PACEM W DUAL-CHAM
 3789 REVISE OR REMOVE PACEMAK
 3794 IMPLT/REPL CARDDEFIB TOT
 3795 IMPLT CARDIODEFIB LEADS
 3796 IMPLT CARDIODEFIB GENATR
 3797 REPL CARDIODEFIB LEADS
 3798 REPL CARDIODEFIB GENRATR

Exclude cases:

- transfer from a hospital (different facility)
- transfer from a skilled Nursing Facility (SNF) or Intermediate Care Facility (ICF)
- transfer from another health care facility
- MDC 14 (pregnancy, childbirth, and puerperium)
- with a code for cardiac procedure in any field

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured*):

Population in Metro Area or county, age 18 years and older.

2a.5 Target population gender: Female, Male

2a.6 Target population age range: age 18 and over

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

Time window can be determined by user, but is generally a calendar year.

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target*)

<p>population being measured - including all codes, logic, and definitions): Population in Metro Area or county, age 18 years and older.</p>
<p>2a.9 Denominator Exclusions (Brief text description of exclusions from the target population): none</p>
<p>2a.10 Denominator Exclusion Details (All information required to collect exclusions to the denominator, including all codes, logic, and definitions): none</p>
<p>2a.11 Stratification Details/Variables (All information required to stratify the measure including the stratification variables, all codes, logic, and definitions): Observed rates may be stratified by age sex.</p>
<p>2a.12-13 Risk Adjustment Type: Risk adjustment method widely or commercially available</p>
<p>2a.14 Risk Adjustment Methodology/Variables (List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method): The predicted value for each case is computed using standard logistic regression and covariates for gender and age (in 5-year age groups). The reference population used in the regression is the universe of discharges for states that participate in the HCUP State Inpatient Databases (SID) for the year 2007, a database consisting of approximately 35 million discharges from 43 states. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., county or state). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.</p>
<p>2a.15-17 Detailed risk model available Web page URL or attachment: URL http://www.qualityindicators.ahrq.gov/downloads/pqi/PQI_Risk_Adjustment_Tables_(Version_4_2).pdf</p>
<p>2a.18-19 Type of Score: Rate/proportion 2a.20 Interpretation of Score: Better quality = Lower score</p>
<p>2a.21 Calculation Algorithm (Describe the calculation of the measure as a flowchart or series of steps): Each Prevention Quality Indicator (PQI) expressed as a rate, is defined as outcome of interest/population at risk or numerator/denominator. The Quality Indicators software performs five steps to produce the PQI rates. 1) Discharge-level data is used to mark inpatient records containing outcomes of interest. 2) Identify populations at risk. For provider PQIs, such as short-term complications from diabetes, populations at risk are derived from hospital discharge records. 3) Calculate observed rates. Using output data from steps 1 and 2, PQI rates are calculated for user-specified combinations of stratifiers. 4) Risk adjust the PQI rates. Regression coefficients from a reference population database are applied to the observed rates in the risk-adjustment process. The risk-adjusted rates will then reflect the age and sex distribution of data in the reference population. 5) Create multivariate signal extraction (MSX) smoothed rates. Shrinkage factors are applied to the risk-adjusted rates for each PQI in the MSX process. For each PQI, the shrinkage estimate reflects a reliability adjustment unique to each indicator. Full information on PQI algorithms and specification can be found at http://qualityindicators.ahrq.gov/pqi_download.htm.</p>
<p>2a.22 Describe the method for discriminating performance (e.g., significance testing): Significance testing is not prescribed by the software. Users may define their methods of discriminating performance according to their application. Although all cases are measured, the rate is considered a sample in time, given the variations in case mix over time. Confidence intervals can be calculated, but again are not prescribed.</p>
<p>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): Not applicable</p>
<p>2a.24 Data Source (Check the source(s) for which the measure is specified and tested) Electronic administrative data/claims</p>
<p>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.): Hospital administrative discharge data. See data requirements in the AHRQ QI Windows Application Documentation: http://www.qualityindicators.ahrq.gov/software.htm</p>

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.

<p>2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL http://www.qualityindicators.ahrq.gov/software.htm</p> <p>2a.29-31 Data dictionary/code table web page URL or attachment: URL http://www.qualityindicators.ahrq.gov/downloads/winqi/AHRQ_QI_Windows_Software_Documentation_V41a.pdf,</p> <p>2a.32-35 Level of Measurement/Analysis (Check the level(s) for which the measure is specified and tested) Population: states, Population: counties or cities</p> <p>2a.36-37 Care Settings (Check the setting(s) for which the measure is specified and tested) Ambulatory Care: Office</p> <p>2a.38-41 Clinical Services (Healthcare services being measured, check all that apply) Clinicians: Physicians (MD/DO)</p>	
TESTING/ANALYSIS	
<p>2b. Reliability testing</p> <p>2b.1 Data/sample (description of data/sample and size): 2007 AHRQ State Inpatient Databases</p> <p>2b.2 Analytic Method (type of reliability & rationale, method for testing): Annual review of ICD-9-CM coding updates for numerator specifications</p> <p>2b.3 Testing Results (reliability statistics, assessment of adequacy in the context of norms for the test conducted): Not applicable</p>	<p>2b</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>2c. Validity testing</p> <p>2c.1 Data/sample (description of data/sample and size): 2007 AHRQ State Inpatient Databases</p> <p>2c.2 Analytic Method (type of validity & rationale, method for testing): Annual update of risk-adjustment covariates and comparative data</p> <p>2c.3 Testing Results (statistical results, assessment of adequacy in the context of norms for the test conducted): Signal variance of 0.000000249270; Average signal ratio of 0.99</p>	<p>2c</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
<p>2d. Exclusions Justified</p> <p>2d.1 Summary of Evidence supporting exclusion(s): Not applicable</p> <p>2d.2 Citations for Evidence: Not applicable</p> <p>2d.3 Data/sample (description of data/sample and size): Not applicable</p> <p>2d.4 Analytic Method (type analysis & rationale): Not applicable</p> <p>2d.5 Testing Results (e.g., frequency, variability, sensitivity analyses): Not applicable</p>	<p>2d</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>
<p>2e. Risk Adjustment for Outcomes/ Resource Use Measures</p> <p>2e.1 Data/sample (description of data/sample and size): 2007 AHRQ State Inpatient Databases (SID)</p>	<p>2e</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></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.

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) 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.

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
 •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 ca... [2])

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 outcome (but not disparities in care) and are present at start of care; Error! Bookmark not defined. OR ... [3]

<p>2e.2 Analytic Method (<i>type of risk adjustment, analysis, & rationale</i>): The predicted value for each case is computed using standard logistic regression and covariates for gender and age (in 5-year age groups). The reference population used in the regression is the universe of discharges for states that participate in the HCUP State Inpatient Databases (SID) for the year 2007, a database consisting of approximately 35 million discharges from 43 states. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., county or state). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.</p> <p>2e.3 Testing Results (<i>risk model performance metrics</i>): c-statistic not calculated</p> <p>2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: Not applicable</p>	N <input type="checkbox"/> NA <input type="checkbox"/>
<p>2f. Identification of Meaningful Differences in Performance</p> <p>2f.1 Data/sample from Testing or Current Use (<i>description of data/sample and size</i>): 2007 AHRQ State Inpatient Databases (SID)</p> <p>2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (<i>type of analysis & rationale</i>): Posterior probability distribution (gamma) and 95% probability intervals</p> <p>2f.3 Provide Measure Scores from Testing or Current Use (<i>description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance</i>): 5th 25th Median 75th 95th 0.000000 0.000003 0.000043 0.000260 0.001208</p>	2f C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
<p>2g. Comparability of Multiple Data Sources/Methods</p> <p>2g.1 Data/sample (<i>description of data/sample and size</i>): Not applicable</p> <p>2g.2 Analytic Method (<i>type of analysis & rationale</i>): Not applicable</p> <p>2g.3 Testing Results (<i>e.g., correlation statistics, comparison of rankings</i>): Not applicable</p>	2g C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/>
<p>2h. Disparities in Care</p> <p>2h.1 If measure is stratified, provide stratified results (<i>scores by stratified categories/cohorts</i>): Based on the 2008 national statistics for angina without procedure (http://hcupnet.ahrq.gov) the 2008 rates are as follows: Overall rate per 100,000: 24.93; Risk adjusted rate: 24.05 Male: 24.42 Female: 25.42 Age groups: 18-39: 2.80; 40-64: 30.37; 65-74: 53.90; 75+: 74.27</p> <p>2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: Rates may be reported by age, gender and race/ethnicity</p>	2h C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Scientific Acceptability of Measure Properties?</p> <p>Steering Committee: Overall, to what extent was the criterion, Scientific Acceptability of Measure Properties, met? Rationale:</p>	2 C <input type="checkbox"/> P <input type="checkbox"/>

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 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.

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 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.

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

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.

	M <input type="checkbox"/> N <input type="checkbox"/>
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 Ratin g
<p>3a. Meaningful, Understandable, and Useful Information</p> <p>3a.1 Current Use: In use</p> <p>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): 1) State of California: Hospital Inpatient Mortality Indicators for California, http://www.oshpd.ca.gov/hid/products/preventable_hospitalizations/pdfs/PH_REPORT_WEB.pdf 2) State of Kentucky, http://chfs.ky.gov/ohp/healthdata 3) State of New Jersey: Find and Compare Quality Care in New Jersey Hospitals, http://www.nj.gov/health/healthcarequality/ 4) Niagara Health Quality Coalition and Alliance for Quality Health Care: New York State Hospital Report Card, http://www.myhealthfinder.com/ 5) State of Texas: Reports on Hospital Performance, http://www.dshs.state.tx.us/thcic/ 6) State of Nevada: Nevada Compare Care, http://nevadacomparecare.net/Monahrq/home.html</p> <p>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): he software is publically available and free of charge (http://www.qualityindicators.ahrq.gov/). Users apply the software to their own administrative data (UB-04 or claims) that is readily available. Hundreds of users have downloaded AHRQ Quality Indicators software.</p> <p>Testing of Interpretability (Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement)</p> <p>3a.4 Data/sample (description of data/sample and size): The AHRQ State Inpatient Databases (SID) consist of approximatley 4,500 counties and 38 million discharges</p> <p>3a.5 Methods (e.g., focus group, survey, QI project): A research team from the School of Public Affairs, Baruch College, under contracts with the Department of Public Health, Weill Medical College and Battelle, Inc., has developed a pair of Hospital Quality Model Reports at the request of the Agency for Healthcare Research & Quality (AHRQ). The AHRQ hip fracture mortality measure is included in the reports. These reports are designed specifically to report comparative information on hospital performance based on the AHRQ Quality Indicators (QIs). The work was done in close collaboration with AHRQ staff and the AHRQ Quality Indicators team.</p> <p>The Model Reports (discussed immediately above) are based on:</p> <ul style="list-style-type: none"> • Extensive search and analysis of the literature on hospital quality measurement and reporting, as well as public reporting on health care quality more broadly; • Interviews with quality measurement and reporting experts, purchasers, staff of purchasing coalitions, and executives of integrated health care delivery systems who are responsible for quality in their facilities; • Two focus groups with chief medical officers of hospitals and/or systems and two focus groups with quality managers from a broad mix of hospitals; • Four focus groups with members of the public who had recently experienced a hospital admission; and • Four rounds of cognitive interviews (a total of 62 interviews) to test draft versions of the two Model Reports with members of the public with recent hospital experience, basic computer literacy but widely varying levels of education. <p>3a.6 Results (qualitative and/or quantitative results and conclusions):</p>	3a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>

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.

Given the above review of the literature and original research that was conducted, a Model report was the result that could help sponsors use the best evidence on public reports so they are most likely to have the desired effects on quality.	
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 <u>or</u> different topic but same target population): 3b.2 Are the measure specifications harmonized ? If not, why?	3b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/>
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:	3c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/>
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, <i>Usability</i> , met? Rationale:	3 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
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 <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
4b. Electronic Sources 4b.1 Are all the data elements available electronically? (<i>elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims</i>) Yes 4b.2 If not, specify the near-term path to achieve electronic capture by most providers.	4b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
4c. Exclusions 4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications? No 4c.2 If yes, provide justification.	4c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/>

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.

<p>4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences</p> <p>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. In a study of approximately 124,000 cancer-free Medicare beneficiaries/year, with subjects contributing data for 1-8 years, angina PQI hospital discharges declined 75% between 1992 and 1999. CAD hospital discharges rose in a reciprocal pattern, while angina discharges with revascularization declined and discharges for myocardial infarction and ischemic heart disease remained relatively constant. The authors conclude "The marked decline in angina PQI hospital discharges during 1992-1999 does not appear to represent improvements in access to care or prevention of heart disease, but rather increased coding of more specific discharge diagnoses for CAD. Our findings suggest that angina hospitalization is not a valid measure for monitoring access to care and, more generally, demonstrate the need for careful, periodic reevaluation of quality measures." [1]</p> <p>[1] Barry G. Saver; Sharon A. Dobie; Pamela K. Green; Ching-Yun Wang; Laura-Mae Baldwin. No Pain, but No Gain? The Disappearance of Angina Hospitalizations, 1992-1999. Med Care. 2009 October ; 47(10): 1106-1110. doi:10.1097/MLR.0b013e31819e1f53.</p>	<p>4d C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>4e. Data Collection Strategy/Implementation</p> <p>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: None</p> <p>4e.2 Costs to implement the measure (<i>costs of data collection, fees associated with proprietary measures</i>): Administrative and census data are collected as part of routine operations. Some staff time is required to download and execute the software</p> <p>4e.3 Evidence for costs: User reports</p> <p>4e.4 Business case documentation: None</p>	<p>4e C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Feasibility</i>?</p>	<p>4</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Feasibility</i>, met? Rationale:</p>	<p>4 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p style="text-align: center;">RECOMMENDATION</p>	
<p>(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.</p>	<p>Time-limited <input type="checkbox"/></p>
<p>Steering Committee: Do you recommend for endorsement? Comments:</p>	<p>Y <input type="checkbox"/> N <input type="checkbox"/> A <input type="checkbox"/></p>
<p style="text-align: center;">CONTACT INFORMATION</p>	
<p>Co.1 Measure Steward (Intellectual Property Owner) Co.1 <u>Organization</u> Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850</p> <p>Co.2 <u>Point of Contact</u></p>	

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

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).

John, Bott, MSSW, MBA, john.bott@ahrq.hhs.gov , 301-427-1317-
<p>Measure Developer If different from Measure Steward</p> <p>Co.3 Organization Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850</p> <p>Co.4 Point of Contact John, Bott, MSSW, MBA, john.bott@ahrq.hhs.gov, 301-427-1317-</p>
<p>Co.5 Submitter If different from Measure Steward POC John, Bott, MSSW, MBA, john.bott@ahrq.hhs.gov, 301-427-1317-, Agency for Healthcare Research and Quality</p>
<p>Co.6 Additional organizations that sponsored/participated in measure development UC Davis Stanford University Battelle Memorial Institute</p>
ADDITIONAL INFORMATION
<p>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. None</p>
<p>Ad.2 If adapted, provide name of original measure: Not applicable Ad.3-5 If adapted, provide original specifications URL or attachment</p>
<p>Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 Year the measure was first released: 2001 Ad.7 Month and Year of most recent revision: 10, 2010 Ad.8 What is your frequency for review/update of this measure? annually Ad.9 When is the next scheduled review/update for this measure? 05, 2011</p>
<p>Ad.10 Copyright statement/disclaimers: The AHRQ QI software is publicly available. We have no copyright disclaimers.</p>
<p>Ad.11 -13 Additional Information web page URL or attachment:</p>
<p>Date of Submission (MM/DD/YY): 12/31/2010</p>

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.

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).

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.