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 #:** 0135 **NQF Project:** Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Evaluation of Left ventricular systolic function (LVS)
De.2 **Brief description of measure**: Percentage of heart failure patients with documentation in the hospital record that left ventricular systolic (LVS) function was evaluated before arrival, during hospitalization, or is planned for after 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. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	**A** **Y☐** **N☐**
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and	**B**

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

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	**Eval 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:** Heart failure (HF) is a major and growing public health problem in the United States that currently affects approximately 5.7 million Americans. More than 670,000 persons in the US are diagnosed with HF annually, and a person aged 40 years or older has a 1 in 5 chance of developing HF in their lifetime. HF is primarily a disease of the elderly, affecting more than 1 in 100 persons older than 65 years. HF is noted as the underlying cause of almost 59,000 deaths in the US annually, and the 5-year case fatality rate approaches 50%. HF was also responsible for more than 1 million hospitalizations and nearly 3.4 million ambulatory care visits in the US in 2006. Hospital discharges for HF increased by 126% between 1996 and 2006. It is the leading cause of hospitalization in persons older than 65 years. The estimated direct and indirect costs of HF in the United States for 2009, including inpatient and outpatient costs, were $37.2 billion. 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	**1a** C☐ P☐ M☐ N☐

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

Association. Circulation. 2010;121:e46–e215.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: Identification of patients with left ventricular systolic dysfunction and subsequent use of angiotensin converting enzyme inhibitors or angiotensin receptor blockers 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 measuring left ventricular function in their HF patients in order to determine the best course of treatment. 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.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
National performance rates:
2Q09: 97.2%
3Q09: 97.3%
4Q09: 97.6%
1Q10: 97.8%

1b.3 Citations for data on performance gap:
Clinical warehouse data:
2Q09: 199,878 HF patients, 4,061 hospitals
3Q09: 180,797 HF patients, 4,061 hospitals
4Q09: 198,429 HF patients, 4,101 hospitals
1Q10: 212,985 HF patients, 4,087 hospitals

1b.4 Summary of Data on disparities by population group:
At the univariate analysis level (unadjusted odds ratios) and consistent with findings in our other HF measures, one racial/ethnic group, namely Native American, had a lower rate in this measure (93.7%) compared to the other racial/ethnic groups (Caucasian 97.2%, African-American 97.8%, Hispanic 96.0%, and Asian/Pacific Islander 97.8%).

1b.5 Citations for data on Disparities:
2009 Clinical warehouse data (Total 773,293 patients with race not missing): 535,940 Caucasian patients, 163,219 African-American patients, 57,714 Hispanic patients, 13,004 Asian/Pacific Islander patients, and 3,416 Native American patients.

1b
C ☐
P ☐
M ☐
N ☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Evidence-based medical therapy to reduce morbidity and mortality in heart failure requires the identification of patients with impaired left ventricular systolic function. National guidelines advocate the evaluation of left ventricular systolic function as the single most important diagnostic test in the management of patients with heart failure. In addition to determining left ventricular systolic function, tests which evaluate LVF such as echocardiograms also provide an opportunity to assess for other structural abnormalities such as valvular, pericardial, or right ventricular abnormalities, which is important given that it is common for patients to have more than one cardiac abnormality that contributes to the development of HF. Furthermore, such studies serve as baselines for comparison for patients who have had a change in clinical status or who have experienced or recovered from a clinical event or received treatment that might have had a significant effect on cardiac function.

1c.2-3. Type of Evidence: Observational study, Expert opinion, 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*)**:**
A comprehensive 2-dimensional echocardiogram with Doppler flow studies is considered the single most

1c
C ☐
P ☐
M ☐
N ☐

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

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

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

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

useful diagnostic test in the evaluation of patients with heart failure. There is compelling evidence that ACE inhibitors and angiotensin receptor blockers reduce morbidity and mortality in HF; however, this benefit only accrues to patients with reduced LVEF. Evaluation of patients with HF to identify those patients with reduced LVEF is required to appropriately focus ACEI/ARB and other effective pharmacologic therapies.

1c.5 **Rating of strength/quality of evidence (***also provide narrative description of the rating and by whom***):** [ACCF/AHA]: Level of Evidence C (Consensus opinion of experts, case studies, or standard of care; Very limited populations evaluated). [HFSA]: Strength of Evidence C (Expert Opinion, Observational studies-epidemiologic findings, Safety Reporting from large-scale use in practice).

1c.6 **Method for rating evidence:** [ACCF/AHA]
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.
[HFSA]
· Strength of Evidence A - Randomized, Controlled, Clinical Trials; May be assigned based on results of a single trial: Randomized controlled clinical trials provide what is considered the most valid form of guideline evidence. Some guidelines require at least 2 positive randomized clinical trials before the evidence for a recommendation can be designated level A. The HFSA guideline committee has occasionally accepted a single randomized, controlled, outcome-based clinical trial as sufficient for level A evidence when the single trial is large with a substantial number of endpoints and has consistent and robust outcomes. However, randomized clinical trial data, whether derived from one or multiple trials, have not been taken simply at face value. They have been evaluated for: (1) endpoints studied, (2) level of significance, (3) reproducibility of findings, (4) generalizability of study results, and (5) sample size and number of events on which outcome results are based.
· Strength of Evidence B - Cohort and Case-Control Studies; Post hoc, subgroup analysis, and meta-analysis; Prospective observational studies or registries: The HFSA guideline process also considers evidence arising from cohort studies or smaller clinical trials with physiologic or surrogate endpoints. This level B evidence is derived from studies that are diverse in design and may be prospective or retrospective in nature. They may involve subgroup analyses of clinical trials or have a case control or propensity design using a matched subset of trial populations. Dose-response studies, when available, may involve all or a portion of the clinical trial population. Evidence generated from these studies has well-recognized, inherent limitations. Nevertheless, their value is enhanced through attention to factors such as pre-specification of hypotheses, biologic rationale, and consistency of findings between studies and across different populations.
· Strength of Evidence C - Expert Opinion; Observational studies-epidemiologic findings; Safety Reporting from large-scale use in practice: The present HFSA guideline makes extensive use of expert opinion, or C-level evidence. The need to formulate recommendations based on level C evidence is driven primarily by a paucity of scientific evidence in many areas critical to a comprehensive guideline. For example, the diagnostic process and the steps used to evaluate and monitor patients with established HF have not been the subject of clinical studies that formally test the validity of one approach versus another. In areas such as these, recommendations must be based on expert opinion or go unaddressed.

1c.7 **Summary of Controversy/Contradictory Evidence:** There is no direct evidence that measuring LV systolic function with echocardiography or other testing by itself improves patient outcomes. However, there is no means of identifying those patients who will benefit from evidence based therapies such as ACE/ARB, beta blockade, or implantable cardioverter defibrillators if this assessment is not performed. Without measuring this process of care, it would be more challenging to ensure that providers are doing what is necessary to identify the appropriate evidence-based therapy for their patients with HF.
Like other process measures targeting under-use, this measure does not provide the capacity to characterize over-use of imaging procedures to assess LV systolic function. However, measurement development groups have suggested that this issue would be best approached with a separate measure.

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.8 Citations for Evidence (*other than guidelines*): There is little direct evidence linking the assessment of LV systolic function to patient outcomes; however, all the landmark studies that have shown benefits of ACE-inhibitors or ARB in patients with HF have been restricted to patients with left ventricular systolic dysfunction. The studies of patients with preserved systolic function have not shown such benefits. The guidelines reflect this evidence base by reserving class I recommendations for ACE or ARB to those patients with LV systolic dysfunction. Thus determining LV systolic function is central to tailoring evidence-based HF therapy.

· Packer M, Cohn J. Consensus recommendations for the management of chronic heart failure. On behalf of the membership of the advisory council to improve outcomes nationwide in heart failure. Am J Cardiol 1999;83:1A-38A.
· The CONSENSUS Trial Study Group. Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). N Engl J Med 1987;316:1429-35.
· The SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. N Engl J Med 1991;325:293-302.
· Cohn JN, Johnson G, Ziesche S, et al. A comparison of enalapril with hydralazine-isosorbide dinitrate in the treatment of chronic congestive heart failure. N Engl J Med 1991;325:303-10.
· Yusuf S, Pfeffer MA, Swedberg K, et al. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial. Lancet 2003;362:777-81.
· Cohn JN, Tognoni G. A randomized trial of the angiotensin-receptor blocker valsartan in chronic heart failure. N Engl J Med 2001;345:1667-75.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number*):
[ACCF/AHA]
7. Two-dimensional echocardiography with Doppler should be performed during initial evaluation of patients presenting with HF to assess LVEF, left ventricular size, wall thickness, and valve function. Radionuclide ventriculography can be performed to assess LVEF and volumes. [p. 1348]
[HFSA]
4.8 It is recommended that patients with a diagnosis of HF undergo evaluation … Assess cardiac structure and function [p. 482]

1c.10 Clinical Practice Guideline Citation:

· Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, Givertz MM, Klapholz M, Moser DK, Rogers JG, Starling RC, Stevenson WG, Tang WHW, Teerlink JR, Walsh MN. Executive Summary: HFSA 2010 Comprehensive Heart Failure Practice Guideline. J Card Fail 2010;16:475e539.
· Jessup M, Abraham WT, Casey DE, Feldman AM, Francis GS, Ganiats TG, et al, writing on behalf of the 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult Writing Committee. 2009 focused update: ACCF/AHA guidelines for the diagnosis and management of heart failure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2009;53:1343–82.

1c.11 National Guideline Clearinghouse or other URL:
http://www.scpcp.org/dnn/WebDocs/HFSA%202010%20HF%20Guidelines.pdf,
http://content.onlinejacc.org/cgi/reprint/53/15/1343.pdf

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom*):
[ACCF/AHA]: 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. [HFSA]: Strength of recommendation - "Is recommended": The recommended therapy or management process should be followed as often as possible in individual patients (part of routine care).

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

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.

Recommendations are assigned strength by the Task Force based upon evidence, benefit vs. risk vs. harm, and patient preference. [HFSA} There are several degrees of favorable recommendations and a single category for therapies felt to be not effective. · "Is recommended": The recommended therapy or management process should be followed as often as possible in individual patients (part of routine care). Exceptions are carefully delineated and should be minimized. · "Should be considered": A majority of patients should receive the intervention, with some discretion involving individual patients. · "May be considered": Individualization of therapy is indicated. · "Is not recommended": Therapeutic intervention should not be used. 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. The HFSA guidelines also characterize their recommendations according to both the weight of evidence (on an A, B, C scale) as well as the strength of the recommendation (categorized as "is recommended," "should be considered," "may be considered," and "is not recommended"). 1c.14 **Rationale for using this guideline over others:** The ACCF/AHA and HFSA guidelines are the only national guidelines that address the therapy of patients with HF; they use an explicit and transparent methodology; and have thus served as the foundation of national quality metrics.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Eval Rating**
2a. MEASURE SPECIFICATIONS	
S.1 **Do you have a web page where current detailed measure specifications can be obtained?** S.2 **If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 **Numerator Statement (***Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** HF patients with documentation in the hospital record that LVS function was evaluated before arrival, during hospitalization, or is planned for after discharge 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 1-254 through 1-256.	**2a-specs** **C☐** **P☐** **M☐** **N☐**

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

· Section 2 - Measurement Information | Section 2.2 – Heart Failure (HF) – pages HF-2-1 through HF-2-5.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
HF patients (ICD-9-CM principal diagnosis of HF: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9)

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:
402.01: Hypertensive heart disease, malignant, with heart failure
402.11: Hypertensive heart disease, benign, with heart failure
402.91: Hypertensive heart disease, unspecified, with heart failure
404.01: Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.03: Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease
404.11: Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.13: Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease stage V or end stage renal disease
404.91: Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.93: Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease
428.0: Congestive heart failure, unspecified
428.1: Left heart failure
428.20: Unspecified systolic heart failure
428.21: Acute systolic heart failure
428.22: Chronic systolic heart failure
428.23: Acute on chronic systolic heart failure
428.30: Unspecified diastolic heart failure
428.31: Acute diastolic heart failure
428.32: Chronic diastolic heart failure
428.33: Acute on chronic diastolic heart failure
428.40: Unspecified combined systolic and diastolic heart failure
428.41: Acute combined systolic and diastolic heart failure
428.42: Chronic combined systolic and diastolic heart failure
428.43: Acute on chronic combined systolic and diastolic heart failure
428.9: Heart failure, unspecified

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 enrolled in clinical trials

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.

•Patients with comfort measures only documented
•Reasons for no LVS function evaluation documented by a physician, advanced practice nurse, or physician assistant
•Patients who had a left ventricular assistive device (LVAD) or heart transplant procedure during hospital stay (ICD-9-CM procedure code of LVAD or Heart Transplant: 33.6, 37.51, 37.52, 37.53, 37.54, 37.60, 37.62, 37.63, 37.65, 37.66, 37.68)

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, 1-201, 1-204 through 1-205, and 1-254 through 1-256.
· Section 2 - Measurement Information | Section 2.2 – Heart Failure (HF) – pages HF-5 plus HF-2-1 through HF-2-5

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 2.2 – Heart Failure (HF) – pages HF-5 plus HF-2-4 through HF-2-5.

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 HF as defined in section 2a.8, no ICD-9-CM Principal or Other Procedure Code of Left Ventricular Assistive Device (LVAD) or Heart Transplant as defined in section 2a.9, 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):
>= 506: 102
131-505: 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)*:
Clinical Trial – 98.9%
Comfort Measures Only – 94.3%
LVF Assessment – 94.5%

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 HF 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 lack of LVSF evaluation is often clinically appropriate. In relation to the exclusion of LVAD and heart transplant cases, there is no clinical data to support the use of ACE-inhibitors in this specific population, therefore it makes clinical sense

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 ... [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.

to automatically exclude these cases from this measure where the intention is primarily to identify appropriate ACEI candidates. 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 cases where a physician decides against assessing left ventricular function and documents his/her reasons. Reasons vary, from patient refusal, to clinical conditions such as ESRD, where the physician believes EF measurement is not indicated. In these types of cases, not doing an LVSF evaluation should not count against the provider if the clinical reason for not assessing LVSF is documented. Exclusions in this measure are concordant with the 2010 ACC/AHA/PCPI Heart Failure Performance Measure Set.

2d.2 Citations for Evidence:

· Bonow RO, Ganiats TG, Beam CT, Blake K, Casey DE, Goodlin SJ, et al. December 2010. American College of Cardiology Foundation/American Heart Association/Physician Consortium for Performance Improvement Heart Failure Performance Measurement Set (voting draft). In American Medical Association. Retrieved December 2010, from http://www.ama-assn.org/ama1/pub/upload/mm/370/heart-failure-measures.pdf.

2d.3 Data/sample *(description of data/sample and size)***:** Clinical warehouse data: 245,776 HF patients, 4,116 hospitals, 1Q10.

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.

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)***:**
Rates of Exclusion:

· Patients with comfort measures only documented: 2.7%
· Patients enrolled in clinical trials: 0.2%
· 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: 10.1%
· Patients with a documented reason for no LVS function evaluation documented by a physician, advanced practice nurse, or physician assistant: 0.4%

2e. Risk Adjustment for Outcomes/ Resource Use Measures

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

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

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

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

2e
C☐ P☐ M☐ N☐ NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)***:** Clinical warehouse data:
2Q09: 199,878 HF patients, 4,061 hospitals
3Q09: 180,797 HF patients, 4,061 hospitals
4Q09: 198,429 HF patients, 4,101 hospitals
1Q10: 212,985 HF patients, 4,087 hospitals

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

2f
C☐ P☐ M☐ N☐

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.

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.

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: 97.2% (benchmark 100.0%)
3Q09: 97.3% (benchmark 100.0%)
4Q09: 97.6% (benchmark 100.0%)
1Q10: 97.8% (benchmark 100.0%)

2g. Comparability of Multiple Data Sources/Methods

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.

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).

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

2g C☐ P☐ M☐ N☐ NA☐

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

2h. Disparities in Care

2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)***:** Not stratified, but results according to race, sex, etc can be determined.

2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**
Although preliminary univariate analyses suggested a possible disparity (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.

2h C☐ P☐ M☐ N☐ NA☐

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

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

2

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

2
C☐ P☐ M☐ N☐

3. USABILITY

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

Eval Rating

3a. Meaningful, Understandable, and Useful Information

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

3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:**
Hospital Inpatient Quality Reporting Program:

3a C☐ P☐ M☐ N☐

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

· 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). 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): · 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.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:**	
(for NQF staff use) **Notes on similar/related endorsed or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): 3b.2 **Are the measure specifications harmonized? If not, why?**	**3b** C☐ P☐ M☐ N☐ NA ☐
3c. Distinctive or Additive Value 3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** 5.1 **If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:** No NQF-endorsed measures with same topic and target population.	**3c** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability?*	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability***, met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be	**Eval**

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).

implemented for performance measurement. (evaluation criteria)	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?** 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** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.** Retooling work with HHS is expected to be completed in the near future.	**4b** C☐ P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA☐
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. Because the denominator exclusion "Patients with a documented reason for no LVS function evaluation" allows for any physician/advance practice nurse/physician assistant/pharmacist-documented "other reason" for not assessing LVSF, 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. 2. 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 that an echo after discharge is being considered vs. a definitive plan, what constitutes acceptable physician documentation of a reason for not assessing LVSF). 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 one data element unique to this measure, LVF Assessment, amounted to 18, only 4.6% of the total 390 Quest questions received for 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.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:**	**4e** C☐ P☐ M☐ 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.

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).

The decision points relating to exclusions comfort measures only, clinical trial, and discharge disposition in the algorithms were rearranged for April 2008+ discharges. The new order enabled tool developers to program tools in such a way that the abstractor could skip abstraction of Comfort Measures Only (challenging data to abstract from some medical records) if the patient was transferred to another acute care hospital, left AMA, expired, or was discharged to hospice, saving valuable abstraction time. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** 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. 4e.3 **Evidence for costs:** N/A 4e.4 **Business case documentation:** N/A	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** **C☐** **P☐** **M☐** **N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐** **N☐** **A☐**

CONTACT INFORMATION

Co.1 **Measure Steward (Intellectual Property Owner)**
Co.1 **Organization**
Centers for Medicare & Medicaid Services, 7500 Security Boulevard , 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 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 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

Ad.2 If adapted, provide name of original measure: N/A
Ad.3-5 If adapted, provide original specifications URL or attachment

Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 **Year the measure was first released:** 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
Ad.10 **Copyright statement/disclaimers:**
Ad.11 -13 **Additional Information web page URL or attachment:**
Date of Submission (*MM/DD/YY*)**:** 12/14/2010

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

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

Page 9: [2] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

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

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

AND

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

AND

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

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

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 #:** 0162 **NQF Project:** Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** ACEI or ARB for left ventricular systolic dysfunction - Heart Failure (HF) Patients
De.2 **Brief description of measure**: Percentage of heart failure (HF) 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. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	**A** **Y**☐ **N**☐

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Accountability, Payment incentive	**C** Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☐ N☐
(for NQF staff use**) Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	**Eval Rating**
(for NQF staff use) **Specific NPP goal**:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Leading cause of morbidity/mortality, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Heart failure (HF) is a major and growing public health problem in the United States that currently affects approximately 5.7 million Americans. More than 670,000 persons in the US are diagnosed with HF annually, and a person aged 40 years or older has a 1 in 5 chance of developing HF in their lifetime. HF is primarily a disease of the elderly, affecting more than 1 in 100 persons older than 65 years. HF is noted as the underlying cause of almost 59,000 deaths in the US annually, and the 5-year case fatality rate approaches 50%. HF was also responsible for more than 1 million hospitalizations and nearly 3.4 million ambulatory care visits in the US in 2006. Hospital discharges for HF increased by 126% between 1996 and 2006. It is the leading cause of hospitalization in persons older than 65 years. The estimated direct and indirect costs of HF in the United States for 2009, including inpatient and outpatient costs, were $37.2 billion. 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	**1a** C☐ P☐ M☐ N☐

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

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).

Subcommittee. Heart disease and stroke statistics—2010 update: a report from the American Heart Association. Circulation. 2010;121:e46–e215.

1b. Opportunity for Improvement

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 HF 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.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
National performance rates:
2Q09: 93.8%
3Q09: 93.6%
4Q09: 94.3%
1Q10: 94.7%

1b.3 Citations for data on performance gap:
Clinical warehouse data:
2Q09: 66,437 HF patients, 3,709 hospitals
3Q09: 59,825 HF patients, 3,622 hospitals
4Q09: 64,433 HF patients, 3,689 hospitals
1Q10: 67,827 HF patients, 3,724 hospitals

1b.4 Summary of Data on disparities by population group:
At the univariate analysis level (unadjusted odds ratios) and consistent with findings in our other HF measures, one racial/ethnic group, namely Native American, had a lower rate in this measure (91.8%) compared to the other racial/ethnic groups (Caucasian 93.1%, African-American 95.1%, Hispanic 93.5%, and Asian/Pacific Islander 95.3%).

1b.5 Citations for data on Disparities:
2009 Clinical warehouse data (Total 250,713 patients with race not missing): 155,808 Caucasian patients, 69,597 African-American patients, 20,068 Hispanic patients, 3,962 Asian/Pacific Islander patients, and 1,278 Native American patients.

1b
C☐
P☐
M☐
N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** ACE inhibitors reduce mortality and morbidity in patients with heart failure and left ventricular systolic dysfunction and are effective in a wide range of patients. 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 heart failure. Guideline committees have also supported the inclusion of ARBs in performance measures for heart failure.

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*)**:**
There is compelling evidence that ACE inhibitors should be used to inhibit the renin-angiotensin-aldosterone system (RAAS) in all HF patients with reduced LVEF. Several large clinical trials have demonstrated in the benefits of ACE-inhibitors on morbidity and mortality in HF patients with reduced LVEF, both chronically and post-MI. Benefits of ACE inhibition were seen in patients with mild, moderate, or severe symptoms and in patients with or without coronary artery disease. Angiotensin converting enzyme inhibitors remain the first

1c
C☐
P☐
M☐
N☐

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

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

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

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

choice for inhibition of the renin-angiotensin system in chronic HF, but ARBs should be considered a reasonable alternative for patients unable to tolerate ACEIs because of cough, The ARBs valsartan and candesartan have demonstrated the benefit of reducing hospitalizations and mortality in patients with LVSD. Additionally, 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]: Level of Evidence A (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.). [HFSA]: Strength of Evidence A (Randomized, controlled, clinical trials; May be assigned based on results of a single trial).

1c.6 Method for rating evidence: [ACCF/AHA]
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.
[HFSA]
· Strength of Evidence A - Randomized, Controlled, Clinical Trials; May be assigned based on results of a single trial: Randomized controlled clinical trials provide what is considered the most valid form of guideline evidence. Some guidelines require at least 2 positive randomized clinical trials before the evidence for a recommendation can be designated level A. The HFSA guideline committee has occasionally accepted a single randomized, controlled, outcome-based clinical trial as sufficient for level A evidence when the single trial is large with a substantial number of endpoints and has consistent and robust outcomes. However, randomized clinical trial data, whether derived from one or multiple trials, have not been taken simply at face value. They have been evaluated for: (1) endpoints studied, (2) level of significance, (3) reproducibility of findings, (4) generalizability of study results, and (5) sample size and number of events on which outcome results are based.
· Strength of Evidence B - Cohort and Case-Control Studies; Post hoc, subgroup analysis, and meta-analysis; Prospective observational studies or registries: The HFSA guideline process also considers evidence arising from cohort studies or smaller clinical trials with physiologic or surrogate endpoints. This level B evidence is derived from studies that are diverse in design and may be prospective or retrospective in nature. They may involve subgroup analyses of clinical trials or have a case control or propensity design using a matched subset of trial populations. Dose-response studies, when available, may involve all or a portion of the clinical trial population. Evidence generated from these studies has well-recognized, inherent limitations. Nevertheless, their value is enhanced through attention to factors such as pre-specification of hypotheses, biologic rationale, and consistency of findings between studies and across different populations.
· Strength of Evidence C - Expert Opinion; Observational studies-epidemiologic findings; Safety Reporting from large-scale use in practice: The present HFSA guideline makes extensive use of expert opinion, or C-level evidence. The need to formulate recommendations based on level C evidence is driven primarily by a paucity of scientific evidence in many areas critical to a comprehensive guideline. For example, the diagnostic process and the steps used to evaluate and monitor patients with established HF have not been the subject of clinical studies that formally test the validity of one approach versus another. In areas such as these, recommendations must be based on expert opinion or go unaddressed.

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***):** · Packer M, Cohn J. Consensus recommendations for the management of chronic heart failure. On behalf of the membership of the advisory council to improve outcomes nationwide in heart failure. Am J Cardiol 1999;83:1A-38A.

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.

· The CONSENSUS Trial Study Group. Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). N Engl J Med 1987;316:1429-35.
· The SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. N Engl J Med 1991;325:293-302.
· 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.
· Cohn JN, Johnson G, Ziesche S, et al. A comparison of enalapril with hydralazine-isosorbide dinitrate in the treatment of chronic congestive heart failure. N Engl J Med 1991;325:303-10.
· Yusuf S, Pfeffer MA, Swedberg K, et al. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial. Lancet 2003;362:777-81.
· Cohn JN, Tognoni G. A randomized trial of the angiotensin-receptor blocker valsartan in chronic heart failure. N Engl J Med 2001;345:1667-75.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
[ACCF/AHA]
3 (under class I). Angiotensin-converting enzyme inhibitors are recommended for all patients with current or prior symptoms of HF and reduced LVEF, unless contraindicated. [p. 1353]
5. Angiotensin II receptor blockers are recommended in patients with current or prior symptoms of HF and reduced LVEF who are ACE inhibitor-intolerant. [p. 1353]
3 (under class IIa). Angiotensin II receptor blockers are reasonable to use as alternatives to ACE inhibitors as first-line therapy for patients with mild to moderate HF and reduced LVEF, especially for patients already taking ARBs for other indications. [p. 1355]
[HFSA]
5.5 ACE inhibitor therapy is recommended for asymptomatic patients with reduced LVEF (<40%). [p. 485]
7.1 ACE inhibitors are recommended for routine administration to symptomatic and asymptomatic patients with LVEF < 40%. [p. 487]
7.3 ARBs are recommended for routine administration to symptomatic and asymptomatic patients with an LVEF < 40% who are intolerant to ACE inhibitors for reasons other than hyperkalemia or renal insufficiency. [p. 487]

1c.10 Clinical Practice Guideline Citation: · Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, Givertz MM, Klapholz M, Moser DK, Rogers JG, Starling RC, Stevenson WG, Tang WHW, Teerlink JR, Walsh MN. Executive Summary: HFSA 2010 Comprehensive Heart Failure Practice Guideline. J Card Fail 2010;16:475e539.
· Jessup M, Abraham WT, Casey DE, Feldman AM, Francis GS, Ganiats TG, et al, writing on behalf of the 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult Writing Committee. 2009 focused update: ACCF/AHA guidelines for the diagnosis and management of heart failure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2009;53:1343–82.
1c.11 National Guideline Clearinghouse or other URL:
http://www.scpcp.org/dnn/WebDocs/HFSA%202010%20HF%20Guidelines.pdf,
http://content.onlinejacc.org/cgi/reprint/53/15/1343.pdf

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
[ACCF/AHA]: [3. and 5.] Class I recommendations - 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.; [3.] 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. It is reasonable to perform procedure/administer treatment. [HFSA]: Strength of recommendation - "Is recommended": The recommended therapy or management process should be followed as often as possible in individual patients (part of routine care).

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.13 Method for rating strength of recommendation (*If different from USPSTF system, also describe rating and how it relates to USPSTF*)**:** [ACCF/AHA] 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. [HFSA} There are several degrees of favorable recommendations and a single category for therapies felt to be not effective. · "Is recommended": The recommended therapy or management process should be followed as often as possible in individual patients (part of routine care). Exceptions are carefully delineated and should be minimized. · "Should be considered": A majority of patients should receive the intervention, with some discretion involving individual patients. · "May be considered": Individualization of therapy is indicated. · "Is not recommended": Therapeutic intervention should not be used. 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. The HFSA guidelines also characterize their recommendations according to both the weight of evidence (on an A, B, C scale) as well as the strength of the recommendation (categorized as "is recommended," "should be considered," "may be considered," and "is not recommended"). **1c.14 Rationale for using this guideline over others:** The ACCF/AHA and HFSA guidelines are the only national guidelines that address the therapy of patients with HF; they use an explicit and transparent methodology; and have thus served as the foundation of national quality metrics.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Eval Rating**
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** HF patients who are prescribed an ACEI or ARB at hospital discharge **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***):**	**2a-specs** **C☐** **P☐** **M☐** **N☐**

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

Refer to
http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036:
· Section 1 - Data Dictionary | Alphabetical Data Dictionary – pages 1-18 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 2.2 – Heart Failure (HF) – pages HF-3-1 through HF-3-5.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
HF patients (International Classification of Diseases, 9th revision, Clinical Modification [ICD-9-CM] principal diagnosis code of HF: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9); 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

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:
402.01: Hypertensive heart disease, malignant, with heart failure
402.11: Hypertensive heart disease, benign, with heart failure
402.91: Hypertensive heart disease, unspecified, with heart failure
404.01: Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.03: Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease
404.11: Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.13: Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease stage V or end stage renal disease
404.91: Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.93: Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease
428.0: Congestive heart failure, unspecified
428.1: Left heart failure
428.20: Unspecified systolic heart failure
428.21: Acute systolic heart failure
428.22: Chronic systolic heart failure
428.23: Acute on chronic systolic heart failure
428.30: Unspecified diastolic heart failure
428.31: Acute diastolic heart failure
428.32: Chronic diastolic heart failure
428.33: Acute on chronic diastolic heart failure
428.40: Unspecified combined systolic and diastolic heart failure
428.41: Acute combined systolic and diastolic heart failure
428.42: Chronic combined systolic and diastolic heart failure
428.43: Acute on chronic combined systolic and diastolic heart failure
428.9: Heart failure, unspecified
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:
•Patients who had a left ventricular assistive device (LVAD) or heart transplant procedure during hospital stay (ICD-9-CM procedure code of LVAD or Heart Transplant: 33.6, 37.51, 37.52, 37.53, 37.54, 37.60, 37.62, 37.63, 37.65, 37.66, 37.68)
•<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 enrolled in clinical trials
•Patients with comfort measures only documented
•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, 1-201, 1-204 through 1-205, 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 2.2 – Heart Failure (HF) – pages HF-5 plus HF-3-1 through HF-3-5

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 2.2 – Heart Failure (HF) – pages HF-5 plus HF-3-4 through HF-3-5.

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 HF as defined in section 2a.8, no ICD-9-CM Principal or Other Procedure Code of Left Ventricular Assistive Device

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.

(LVAD) or Heart Transplant as defined in section 2a.9, 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):
>= 506: 102
131-505: 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%
LVSD – 94.7%
Reason for No ACEI and No ARB at Discharge – 77.5%

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c
C☐
P☐

2c.1 **Data/sample** *(description of data/sample and size)***:** Face validity is regularly assessed with the

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.

Technical Expert Panel responsible for reviewing and supporting the measure topic.

M☐ N☐

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

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 HF 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. In relation to the exclusion of LVAD and heart transplant cases, there is no clinical data to support the use of ACE-inhibitors in this specific population. 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 both the 2005 ACC/AHA Clinical Performance Measures for Adults With Chronic Heart Failure and the 2010 ACC/AHA/PCPI Heart Failure Performance Measure Set.

2d.2 Citations for Evidence:

- Bonow RO, Bennett S, Casey DE, Ganiats TG, Hlatky MA, Konstam MA, et al. ACC/AHA Clinical Performance Measures for Adults With Chronic Heart Failure: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures (Writing Committee to Develop Heart Failure Clinical Performance Measures). J Am Coll Cardiol. 2005;46:1144–78.
- Bonow RO, Ganiats TG, Beam CT, Blake K, Casey DE, Goodlin SJ, et al. December 2010. American College of Cardiology Foundation/American Heart Association/Physician Consortium for Performance Improvement Heart Failure Performance Measurement Set (voting draft). In American Medical Association. Retrieved December 2010, from http://www.ama-assn.org/ama1/pub/upload/mm/370/heart-failure-measures.pdf.

2d.3 Data/sample *(description of data/sample and size)*: Clinical warehouse data: 245,779 HF patients, 4,116 hospitals, 1Q10.

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.

2d.5 Testing Results *(e.g., frequency, variability, sensitivity analyses)*:
Rates of Exclusion:
- Patients with comfort measures only documented: 2.7%
- Patients enrolled in clinical trials: 0.2%
- 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: 10.1%
- LVSD not documented as either EF < 40% or a narrative description consistent with moderate or severe systolic dysfunction: 51.1%
- Patients with a documented reason for no ACEI and no ARB at discharge: 8.3%

2d C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e C☐

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;[Error! Bookmark not defined.] OR rationale/data support no risk adjustment.

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

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

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

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

P☐ M☐ N☐ NA☐

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.

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: Clinical warehouse data:
2Q09: 66,437 HF patients, 3,709 hospitals
3Q09: 59,825 HF patients, 3,622 hospitals
4Q09: 64,433 HF patients, 3,689 hospitals
1Q10: 67,827 HF patients, 3,724 hospitals

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.

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.

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.

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: 93.8% (benchmark 99.8%)
3Q09: 93.6% (benchmark 99.8%)
4Q09: 94.3% (benchmark 99.8%)
1Q10: 94.7% (benchmark 99.8%)

2f C☐ P☐ M☐ N☐

2g. Comparability of Multiple Data Sources/Methods

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

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.

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).

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

2g C☐ P☐ M☐ N☐ NA☐

2h. Disparities in Care

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.

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.

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:
Although preliminary univariate analyses suggested a possible disparity (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

2h C☐ P☐ M☐ N☐ NA☐

discharged from the same hospitals.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* **C☐ P☐ M☐ N☐**
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** In use 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** 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). 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): · 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.6 **Results** *(qualitative and/or quantitative results and conclusions)***:** Not available.	**3a** **C☐ P☐ M☐ N☐**
3b/3c. Relation to other NQF-endorsed measures 3b.1 **NQF # and Title of similar or related measures:** NQF #0610: Heart Failure - Use of ACE Inhibitor (ACEI) or Angiotensin Receptor Blocker (ARB) Therapy	
(for NQF staff use) **Notes on similar/related endorsed or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): 3b.2 **Are the measure specifications harmonized? If not, why?**	**3b** **C☐ P☐ M☐**

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.

No, this measure's specifications are not harmonized with NQF #0610 measure specifications. NQF #0610 is an outpatient measure which uses a three year time window and is based on administrative data. In contrast, this measure is concentrated on care of the HF patient who is admitted for inpatient care; a completely different focus in terms of setting and care. NQF #0092 appears to use the same ICD-9-CM codes to identify HF patients as this measure, and like this measure, excludes patients with aortic stenosis and ACEI/ARB allergies, but it automatically excludes many other types of patients, including but not limited to those with hyperpotassemia, secondary renovascular hypertension, chronic kidney disease, multiple myeloma, hypertrophic cardiomyopathy, pregnancy, pulmonary hypertension treatment, hydralazine after prior ACEI/ARB use, and evidence of metastatic disease or active treatment of malignancy in the last 6 months – 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).	N☐ NA ☐
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. 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.	**3c** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, met? **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	**Eval Rating**
4a. Data Generated as a Byproduct of Care Processes 4a.1-2 **How are the data elements that are needed to compute measure scores generated?** 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** C☐ P☐ M☐ N☐
4b. Electronic Sources 4b.1 **Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) No 4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.** Retooling work with HHS is expected to be completed in the near future.	**4b** C☐ P☐ M☐ N☐
4c. Exclusions 4c.1 **Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No 4c.2 **If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA ☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences 4d.1 **Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and**	**4d** C☐ 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.

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

describe how these potential problems could be audited. If audited, provide results. 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. 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. 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. 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 AMI 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.	**M**☐ **N**☐
4e. Data Collection Strategy/Implementation **4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:** 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+	**4e** **C**☐ **P**☐ **M**☐ **N**☐

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

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. Additionally, the decision points relating to exclusions comfort measures only, clinical trial, and discharge disposition in the algorithms were rearranged for April 2008+ discharges. The new order enabled tool developers to program tools in such a way that the abstractor could skip abstraction of Comfort Measures Only (challenging data to abstract from some medical records) if the patient was transferred to another acute care hospital, left AMA, expired, or was discharged to hospice, saving important abstraction time as well. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** 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. 4e.3 **Evidence for costs:** N/A 4e.4 **Business case documentation:** N/A	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** **C☐** **P☐** **M☐** **N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐** **N☐** **A☐**
CONTACT INFORMATION	
Co.1 **Measure Steward (Intellectual Property Owner)** Co.1 **Organization** Centers for Medicare & Medicaid Services, 7500 Security Boulevard , 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 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 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
Ad.2 If adapted, provide name of original measure: N/A **Ad.3-5 If adapted, provide original specifications URL or attachment**
Measure Developer/Steward Updates and Ongoing Maintenance **Ad.6 Year the measure was first released:** 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
Ad.10 Copyright statement/disclaimers:
Ad.11 -13 Additional Information web page URL or attachment:
Date of Submission (*MM/DD/YY*)**:** 12/14/2010

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

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

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 #:** 0136 **NQF Project:** Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Heart Failure (HF): Detailed discharge instructions
De.2 **Brief description of measure**: Percentage of heart failure patients discharged home with written instructions or educational material given to patient or caregiver at discharge or during the hospital stay addressing all of the following: activity level, diet, discharge medications, follow-up appointment, weight monitoring, and what to do if symptoms worsen.
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:** Patient and family engagement De.5 **IOM Quality Domain**: Patient-centered 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**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	**A** **Y**☐ **N**☐

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement Accountability, Payment incentive	**C** Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☐ N☐
(for NQF staff use**)** **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	**Eval Rating**
(for NQF staff use) **Specific NPP goal**:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Leading cause of morbidity/mortality, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Heart failure (HF) is a major and growing public health problem in the United States that currently affects approximately 5.7 million Americans. More than 670,000 persons in the US are diagnosed with HF annually, and a person aged 40 years or older has a 1 in 5 chance of developing HF in their lifetime. HF is primarily a disease of the elderly, affecting more than 1 in 100 persons older than 65 years. HF is noted as the underlying cause of almost 59,000 deaths in the US annually, and the 5-year case fatality rate approaches 50%. HF was also responsible for more than 1 million hospitalizations and nearly 3.4 million ambulatory care visits in the US in 2006. Hospital discharges for HF increased by 126% between 1996 and 2006. It is the leading cause of hospitalization in persons older than 65 years. The estimated direct and indirect costs of HF in the United States for 2009, including inpatient and outpatient costs, were $37.2 billion. 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	**1a** C☐ P☐ M☐ N☐

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

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).

Subcommittee. Heart disease and stroke statistics—2010 update: a report from the American Heart Association. Circulation. 2010;121:e46–e215.

1b. Opportunity for Improvement

1b.1 Benefits (improvements in quality) envisioned by use of this measure: It is important to seize the opportunity that each hospitalization to educate patients with chronic conditions like HF. Giving the patient written discharge instructions helps to reinforce with the patient a wide range of issues, including medications, diet, activity level, and symptoms. It also gives patients the chance to ask important questions. Providing patients with discharge instructions reduces readmissions. Elderly people with heart failure have the highest rehospitalization rate of all adult patient groups, with estimated annual total direct healthcare expenditures exceeding $24.3 billion. Between 29 to 47 percent of elderly HF patients are readmitted for their condition within three to six months of an initial hospitalization. Hospital performance rates have gradually increased over the years this measure has been reported to the public but significant opportunities for improvement remain (national average 88.5%). Providers understand the importance of discharge instructions for their HF patients. Ongoing use of this measure will help ensure that high performing providers maintain high performance and the many relatively lower performing providers have an impetus to improve.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
National performance rates:
2Q09: 85.6%
3Q09: 86.9%
4Q09: 87.7%
1Q10: 88.5%

1b.3 Citations for data on performance gap:
Clinical warehouse data:
2Q09: 161,581 HF patients, 4,019 hospitals
3Q09: 145,645 HF patients, 4,000 hospitals
4Q09: 160,288 HF patients, 4,047 hospitals
1Q10: 170,505 HF patients, 4,040 hospitals

1b.4 Summary of Data on disparities by population group:
At the univariate analysis level (unadjusted odds ratios) and consistent with findings in our other HF measures, one racial/ethnic group, namely Native American, had a lower rate in this measure (76.3%) compared to the other racial/ethnic groups (Caucasian 86.3%, African-American 86.3%, Hispanic 86.6%, and Asian/Pacific Islander 87.0%).

1b.5 Citations for data on Disparities:
2009 Clinical warehouse data (Total 624,579 patients with race not missing): 414,742 Caucasian patients, 143,689 African-American patients, 51,690 Hispanic patients, 11,375 Asian/Pacific Islander patients, and 3,083 Native American patients.

1b
C☐ P☐ M☐ N☐

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*)**:** Education of heart failure patients and their families is critical. Failure of these patients to comply with physician's instructions, particularly with diet and medications, can cause exacerbation of HF. An important cause of patient's failure to comply is lack of understanding. It is, therefore, incumbent on health care professionals to be certain that patients and their families have an understanding of the causes of heart failure, prognosis, therapy, dietary restrictions, activity, importance of compliance, and the signs and symptoms of recurrent heart failure. Providing patients with discharge instructions reduces readmissions and thorough discharge planning is associated with improved patient outcomes. National guidelines strongly support the role of patient education.

1c.2-3. Type of Evidence: Cohort study, Observational study, Expert opinion, Systematic synthesis of

1c
C☐ P☐ M☐ N☐

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

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

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

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

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*)**:**
Written discharge instructions or educational material given to patient and/or caregiver at hospital discharge to home or during the hospital stay which addresses activity level, diet, discharge medications, follow-up appointment, weight monitoring, and what to do if heart failure symptoms worsen are important for care coordination and transition after discharge. Education of HF patients and their families is critical and often complex. Failure of these patients to understand how best to comply with physician's instructions is often a cause of HF exacerbation leading to subsequent hospital readmission. A retrospective study of HF patients found a correlation between documentation of compliance with the aforementioned discharge instructions and reduced readmission rates.

In terms of diet instruction, excessive dietary sodium intake is a common proximate cause of worsening symptoms and hospitalization for HF exacerbation. It is not enough to simply ask patients to follow a low salt diet. Patients need to be appropriately educated about daily sodium intake targets and how to reach targets, calorie and carbohydrate restriction, etc.

In relation to follow-up instructions, several studies have examined the effect of providing more intensive delivery of discharge instructions coupled tightly with subsequent well-coordinated follow-up care for patients hospitalized with HF, many with positive results. A meta-analysis of 18 studies representing data from 8 countries randomized 3,304 older inpatients with HF to comprehensive discharge planning plus post-discharge support or usual care. During a mean observation period of 8 months, fewer intervention patients were readmitted compared with controls. Analysis of studies reporting secondary outcomes found a trend toward lower all-cause mortality, length of stay, hospital costs, and improvement in quality-of-life scores for patients assigned to an intervention compared with usual care.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
[ACCF/AHA]: Level of Evidence C (Consensus opinion of experts, case studies, or standard of care; Very limited populations evaluated). [HFSA]: Strength of Evidence B (Cohort and Case-Control Studies; Post hoc, subgroup analysis, and meta-analysis; Prospective observational studies or registries); Strength of Evidence C (Expert Opinion, Observational studies-epidemiologic findings, Safety Reporting from large-scale use in practice)

1c.6 Method for rating evidence: [ACCF/AHA]
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.
[HFSA]
· Strength of Evidence A - Randomized, Controlled, Clinical Trials; May be assigned based on results of a single trial: Randomized controlled clinical trials provide what is considered the most valid form of guideline evidence. Some guidelines require at least 2 positive randomized clinical trials before the evidence for a recommendation can be designated level A. The HFSA guideline committee has occasionally accepted a single randomized, controlled, outcome-based clinical trial as sufficient for level A evidence when the single trial is large with a substantial number of endpoints and has consistent and robust outcomes. However, randomized clinical trial data, whether derived from one or multiple trials, have not been taken simply at face value. They have been evaluated for: (1) endpoints studied, (2) level of significance, (3) reproducibility of findings, (4) generalizability of study results, and (5) sample size and number of events on which outcome results are based.
· Strength of Evidence B - Cohort and Case-Control Studies; Post hoc, subgroup analysis, and meta-analysis; Prospective observational studies or registries: The HFSA guideline process also considers evidence

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.

arising from cohort studies or smaller clinical trials with physiologic or surrogate endpoints. This level B evidence is derived from studies that are diverse in design and may be prospective or retrospective in nature. They may involve subgroup analyses of clinical trials or have a case control or propensity design using a matched subset of trial populations. Dose-response studies, when available, may involve all or a portion of the clinical trial population. Evidence generated from these studies has well-recognized, inherent limitations. Nevertheless, their value is enhanced through attention to factors such as pre-specification of hypotheses, biologic rationale, and consistency of findings between studies and across different populations.

· Strength of Evidence C - Expert Opinion; Observational studies-epidemiologic findings; Safety Reporting from large-scale use in practice: The present HFSA guideline makes extensive use of expert opinion, or C-level evidence. The need to formulate recommendations based on level C evidence is driven primarily by a paucity of scientific evidence in many areas critical to a comprehensive guideline. For example, the diagnostic process and the steps used to evaluate and monitor patients with established HF have not been the subject of clinical studies that formally test the validity of one approach versus another. In areas such as these, recommendations must be based on expert opinion or go unaddressed.

1c.7 Summary of Controversy/Contradictory Evidence: There are no randomized trials that prove the efficacy of discharge instructions. [Patterson ME, Hernandez AF, Hammill BG, Fonarow GC, Peterson ED, Schulman KA, Curtis LH. Process of care performance measures and long-term outcomes in patients hospitalized with heart failure. Med Care. 2010 Mar;48(3):210-6.]

1c.8 Citations for Evidence (*other than guidelines***):** · VanSuch M, Naessens JM, Stroebel RJ, Huddleston JM, Williams AR. Effect of discharge instructions on readmission of hospitalised patients with heart failure: do all of the Joint Commission on Accreditation of Healthcare Organizations heart failure core measures reflect better care? Qual Saf Health Care. 2006 Dec;15(6):414-7.

· Bennet SJ, Huster GA, Baker SL, Milgrom ALB, Kirchgassner Birt J, et al. Characterization of the precipitants of hospitalization for heart failure decompensation. Am J Crit Care 1998;7:168e74.

· Michalsen A, Konig G, Thimme W. Preventable causative factors leading to hospital admission with decompensated heart failure. Heart 1998;80:437e41.

· Tsuyuki RT, McKelvie RS, Arnold JM, Avezum A Jr, Barretto AC, Carvalho AC, et al. Acute precipitants of congestive heart failure exacerbations. Arch Intern Med 2001;161:2337e42.

· McAlister FA, Stewart S, Ferrua S, et al. Multidisciplinary strategies for the management of heart failure patients at high risk for admission: a systematic review of randomized trials. J Am Coll Cardiol. 2004;44:810 -9.

· Naylor MD, Brooten DA, Campbell RL, et al. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. J Am Geriatr Soc. 2004;52:675– 84.

· Casey DE Jr., Abraham WT, Guo L, et al. Reducing heart failure hospitalizations and readmissions with heart failure advocates: A call to action for nursing. Circulation. 2007;115:e559–60.

· Windham BG, Bennett RG, Gottlieb S. Care management interventions for older patients with congestive heart failure. Am J Manag Care. 2003;9:447–59.

· Phillips CO, Wright SM, Kern DE, et al. Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: a meta-analysis. JAMA. 2004;291:1358–67.

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
[ACCF/AHA]
17. Comprehensive written discharge instructions for all patients with a hospitalization for HF and their caregivers is strongly recommended, with special emphasis on the following 6 aspects of care: diet, discharge medications, with a special focus on adherence, persistence, and uptitration to recommended doses of ACE inhibitor/ARB and beta-blocker medication, activity level, follow-up appointments, daily weight monitoring, and what to do if HF symptoms worsen. [p. 1363]
[HFSA]
6.1 Dietary instruction regarding sodium intake is recommended in all patients with HF. Patients with HF and diabetes, dyslipidemia, or severe obesity should be given specific dietary instructions. [p. 485]
12.25 It is recommended that criteria be met before a patient with HF is discharged from the hospital .. Patient and family education completed, including clear discharge instructions. [p. 500]
12.26 Discharge planning is recommended as part of the management of patients with ADHF. Discharge planning should address the following issues: Details regarding medication, dietary sodium restriction, and recommended activity level … Follow-up by phone or clinic visit early after discharge to reassess volume status .. Medication and dietary compliance … Monitoring of body weight, electrolytes and renal function.

[p. 500]

1c.10 Clinical Practice Guideline Citation: · Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, Givertz MM, Klapholz M, Moser DK, Rogers JG, Starling RC, Stevenson WG, Tang WHW, Teerlink JR, Walsh MN. Executive Summary: HFSA 2010 Comprehensive Heart Failure Practice Guideline. J Card Fail 2010;16:475e539.
· Jessup M, Abraham WT, Casey DE, Feldman AM, Francis GS, Ganiats TG, et al, writing on behalf of the 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult Writing Committee. 2009 focused update: ACCF/AHA guidelines for the diagnosis and management of heart failure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2009;53:1343–82.
1c.11 National Guideline Clearinghouse or other URL:
http://www.scpcp.org/dnn/WebDocs/HFSA%202010%20HF%20Guidelines.pdf,
http://content.onlinejacc.org/cgi/reprint/53/15/1343.pdf

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
[ACCF/AHA] - 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. [HFSA] - Strength of recommendation - "Is recommended": The recommended therapy or management process should be followed as often as possible in individual patients (part of routine care).

1c.13 Method for rating strength of recommendation (*If different from USPSTF system, also describe rating and how it relates to USPSTF)***:**
[ACCF/AHA] 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.
[HFSA}
There are several degrees of favorable recommendations and a single category for therapies felt to be not effective.
· "Is recommended": The recommended therapy or management process should be followed as often as possible in individual patients (part of routine care). Exceptions are carefully delineated and should be minimized.
· "Should be considered": A majority of patients should receive the intervention, with some discretion involving individual patients.
· "May be considered": Individualization of therapy is indicated.
· "Is not recommended": Therapeutic intervention should not be used.
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. The HFSA guidelines also characterize their recommendations according to both the weight of evidence (on an A, B, C scale) as well as the strength of the recommendation (categorized as "is recommended," "should be considered," "may be considered," and "is not recommended").

1c.14 Rationale for using this guideline over others:
The ACCF/AHA and HFSA guidelines are the only national guidelines that address the therapy of patients with HF; they use an explicit and transparent methodology; and have thus served as the foundation of national quality metrics.

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

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐**

	N☐
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Eval Rating**
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:**	
2a. Precisely Specified	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** HF patients with documentation that they or their caregivers were given written discharge instructions or other educational material addressing all of the following: 1.activity level 2.diet 3.discharge medications 4.follow-up appointment 5.weight monitoring 6.what to do if symptoms worsen	
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 1-121 through 1-122, 1-125 through 1-126, 1-129 through 1-130, 1-133 through 1-136, and 1-139 through 1-142. · Section 2 - Measurement Information | Section 2.2 – Heart Failure (HF) – pages HF-1-1 through HF-1-7.	
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):** HF patients discharged home (ICD-9-CM principal diagnosis of HF: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9); and a discharge to home, home care, or court/law enforcement	
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: 402.01: Hypertensive heart disease, malignant, with heart failure 402.11: Hypertensive heart disease, benign, with heart failure 402.91: Hypertensive heart disease, unspecified, with heart failure 404.01: Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic	**2a-specs** C☐ P☐ M☐ N☐

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

kidney disease stage I through stage IV, or unspecified
404.03: Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease
404.11: Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.13: Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease stage V or end stage renal disease
404.91: Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.93: Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease
428.0: Congestive heart failure, unspecified
428.1: Left heart failure
428.20: Unspecified systolic heart failure
428.21: Acute systolic heart failure
428.22: Chronic systolic heart failure
428.23: Acute on chronic systolic heart failure
428.30: Unspecified diastolic heart failure
428.31: Acute diastolic heart failure
428.32: Chronic diastolic heart failure
428.33: Acute on chronic diastolic heart failure
428.40: Unspecified combined systolic and diastolic heart failure
428.41: Acute combined systolic and diastolic heart failure
428.42: Chronic combined systolic and diastolic heart failure
428.43: Acute on chronic combined systolic and diastolic heart failure
428.9: Heart failure, unspecified
Discharge Disposition - Refer to http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228760129036:
· Section 1 - Data Dictionary | Alphabetical Data Dictionary – pages 1-118 through 1-120.

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
•Patients with comfort measures only documented
•Patients who had a left ventricular assistive device (LVAD) or heart transplant procedure during hospital stay (ICD-9-CM procedure code of LVAD and Heart Transplant: 33.6, 37.51, 37.52, 37.53, 37.54, 37.60, 37.62, 37.63, 37.65, 37.66, 37.68)

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, 1-201, and 1-204 through 1-205.
· Section 2 - Measurement Information | Section 2.2 – Heart Failure (HF) – pages HF-5 plus HF-1-1 through HF-1-7

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

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.

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 2.2 – Heart Failure (HF) – pages HF-5 plus HF-1-4 through HF-1-7.

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 HF as defined in section 2a.8, no ICD-9-CM Principal or Other Procedure Code of Left Ventricular Assistive Device (LVAD) or Heart Transplant as defined in section 2a.9, 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):
>= 506: 102
131-505: 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** C☐ P☐ M☐

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

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.

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)*:
Clinical Trial – 98.9%
Comfort Measures Only – 94.3%
Discharge Instructions Address Activity – 96.3%
Discharge Instructions Address Diet – 97.1%
Discharge Instructions Address Follow-up – 96.4%
Discharge Instructions Address Medications – 81.7%
Discharge Instructions Address Symptoms Worsening – 91.7%
Discharge Instructions Address Weight Monitoring – 93.6%

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 HF 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, written discharge instructions for the patient/caregiver to help ensure patient compliance post-discharge become relatively irrelevant. Although discharge instructions are arguably important in LVAD and heart transplant cases, these cases are excluded due to the population sampling methodology that this measure must share with the other HF measures in the HF measure set. Patients who leave against medical advice or who expire are appropriately excluded, and it is sensible for those who are discharged to another hospital or other health care facility (where the patient goes on to continue treatment and responsibility of care does not yet fall on him/her) to be omitted as well. Exclusions in this measure are concordant with both the 2005 ACC/AHA Clinical Performance Measures for Adults With Chronic Heart Failure and the 2010 ACC/AHA/PCPI Heart Failure Performance Measure Set.

2d.2 Citations for Evidence:

· Bonow RO, Bennett S, Casey DE, Ganiats TG, Hlatky MA, Konstam MA, et al. ACC/AHA Clinical Performance Measures for Adults With Chronic Heart Failure: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures (Writing Committee to Develop Heart Failure Clinical Performance Measures). J Am Coll Cardiol. 2005;46:1144–78.
· Bonow RO, Ganiats TG, Beam CT, Blake K, Casey DE, Goodlin SJ, et al. December 2010. American College of Cardiology Foundation/American Heart Association/Physician Consortium for Performance Improvement Heart Failure Performance Measurement Set (voting draft). In American Medical Association. Retrieved December 2010, from http://www.ama-assn.org/ama1/pub/upload/mm/370/heart-failure-measures.pdf.

2d.3 Data/sample *(description of data/sample and size)*: Clinical warehouse data: 245,783 HF patients,

2d
C☐
P☐
M☐
N☐
NA☐

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 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.

4,117 hospitals, 1Q10.

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.

2d.5 **Testing Results** *(e.g., frequency, variability, sensitivity analyses)*:
Rates of Exclusion:
- Patients with comfort measures only documented: 1.2%
- Patients enrolled in clinical trials: 0.2%
- Patients not discharged to home/home care or not discharged/transferred to court/law enforcement: 29.3%

2e. Risk Adjustment for Outcomes/ Resource Use Measures

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

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

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

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

2e
C☐ P☐ M☐ N☐ NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 **Data/sample from Testing or Current Use** *(description of data/sample and size)*: Clinical warehouse data:
2Q09: 161,581 HF patients, 4,019 hospitals
3Q09: 145,645 HF patients, 4,000 hospitals
4Q09: 160,288 HF patients, 4,047 hospitals
1Q10: 170,505 HF patients, 4,040 hospitals

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.

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: 85.6% (benchmark 99.7%)
3Q09: 86.9% (benchmark 99.8%)
4Q09: 87.7% (benchmark 99.8%)
1Q10: 88.5% (benchmark 99.9%)

2f
C☐ P☐ M☐ N☐

2g. Comparability of Multiple Data Sources/Methods

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.

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

2g
C☐ P☐ M☐ N☐ NA☐

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.

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

No tests have been performed on this measure to determine comparability of sources (paper medical record vs. EHR).

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

2h. Disparities in Care

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.

2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**
Although preliminary univariate analyses suggested a possible disparity (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.

2h
C☐ P☐ M☐ N☐ NA☐

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

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

2

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

2
C☐ P☐ M☐ N☐

3. USABILITY

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

Eval Rating

3a. Meaningful, Understandable, and Useful Information

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

3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:**
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). 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):
· 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
C☐ P☐ M☐ N☐

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

3a.5 Methods *(e.g., focus group, survey, QI project)*:
Voluntary electronic survey by visitors to website.

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:

(for NQF staff use) **Notes on similar/related endorsed or submitted measures:**

3b. Harmonization
If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population):
3b.2 Are the measure specifications harmonized? If not, why?

3b C☐ P☐ M☐ N☐ **NA** ☐

3c. Distinctive or Additive Value
3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:

5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality:
No NQF-endorsed measures with same topic and target population.

3c C☐ P☐ M☐ N☐ **NA** ☐

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

3

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

3 C☐ P☐ M☐ N☐

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)

Eval Rating

4a. Data Generated as a Byproduct of Care Processes

4a.1-2 How are the data elements that are needed to compute measure scores generated?
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 C☐ P☐ M☐ N☐

4b. Electronic Sources

4b.1 Are all the data elements available electronically? (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*)
No

4b.2 If not, specify the near-term path to achieve electronic capture by most providers.
Retooling work with HHS is expected to be completed in the near future.

4b C☐ P☐ M☐ N☐

4c. Exclusions

4c.1 Do the specified exclusions require additional data sources beyond what is required for the

4c C☐ 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).

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.

numerator and denominator specifications?
No

M☐ N☐ 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. It is important to note that this measure focuses on whether activity, diet, etc. after discharge were addressed in the written instructions sent home with the patient. It does not measure the quality of those instructions (e.g., accuracy of instructions, clarity, customized to patient needs). In some cases, quality of instruction has been sacrificed in an effort by the hospital to pass the measure. We consider measuring of the quality of discharge instructions as a different measure that should be considered in the future.
2. Abstraction of the Discharge Instructions Address Medications data element is challenging. The process of compiling a final list of all medications being prescribed at discharge and then comparing this list to the list given to the patient, to confirm completeness, requires substantial time from the abstractors, given the nature of this documentation in the medical record (e.g., conflicting documentation amongst sources, loose references such as "continue same medications", medications referenced by class and not named such as "Sent home on beta-blocker", handling of vitamins, food supplements, etc. where documentation tends to be less specific, records without documentation necessary to build a comparison list, matching up of brand or trade names vs. generic names, therapeutic substitutions made by the pharmacy). A necessary complex set of data abstraction guidelines has evolved to assist the abstractor to determine just how to classify discharge medication matches/mismatches, given the many different ways medications can be referenced. Abstraction guidelines are reviewed and revised on an ongoing basis, in an effort to reduce burden. Additionally, fact sheets which summarize important abstraction principles are published to help abstractors with data collection.
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 determine from documentation whether a copy of the instruction sheet was actually given to the patient, how to handle documentation of a plan to start a medication at discharge in terms of identifying discharge medications, what constitutes acceptable instructions for activity, diet, etc.). 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 discharge instructions elements amounted to 172, 44.1% of the total 390 Quest questions received for HF for that month (medication instructions accounted for 142 of the 172 questions – not unexpectedly, given the inherent issues with this element as briefly discussed in #2 above). 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.

4d
C☐ P☐ M☐ N☐

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

4e. Data Collection Strategy/Implementation

4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:
The decision points relating to exclusions comfort measures only, clinical trial, and discharge disposition in the algorithms were rearranged for April 2008+ discharges. The new order enabled tool developers to program tools in such a way that the abstractor could skip abstraction of Comfort Measures Only (challenging data to abstract from some medical records) if the patient was transferred to another acute care hospital, left AMA, expired, or was discharged to hospice, saving valuable abstraction time. Additionally, given the number of problems that were surfacing as abstractors attempted abstraction too soon after discharge, we now advise abstractors to hold off on data collection until the discharge summary is filed in and the chart is

4e
C☐ P☐ M☐ N☐

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

complete and closed whenever possible. Not only does this enable the abstractor to gather as much information about the hospitalization as possible (capture important information that may not have been present in the chart earlier), but if picked for validation, this will reduce the number of potential mismatches that can occur when the CDAC is abstracting from what amounts to a different chart than what the hospital abstractor used. 4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** Varies according to data collection method (use of vendor) and type of abstractor used to collect clinical data. Many hospitals have implemented standardized medical record documentation processes to reduce abstraction burden related to this measure. 4e.3 **Evidence for costs:** N/A 4e.4 **Business case documentation:** N/A	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, **met?** **Rationale:**	**4** **C☐** **P☐** **M☐** **N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐** **N☐** **A☐**

CONTACT INFORMATION
Co.1 **Measure Steward (Intellectual Property Owner)** Co.1 **Organization** Centers for Medicare & Medicaid Services, 7500 Security Boulevard , 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 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 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

Ad.2 If adapted, provide name of original measure: N/A
Ad.3-5 If adapted, provide original specifications URL or attachment

Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 **Year the measure was first released:** 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
Ad.10 **Copyright statement/disclaimers:**
Ad.11 -13 **Additional Information web page URL or attachment:**
Date of Submission (*MM/DD/YY*)**:** 12/14/2010

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

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

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 #:** 0358 **NQF Project:** Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Congestive Heart Failure (CHF) Mortality Rate (IQI 16)
De.2 **Brief description of measure**: Perecent of discharges with principal diagnosis code of CHF with in-hospital mortality
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** Mortality for Selected Conditions composite (NQF #0530)
De.4 **National Priority Partners Priority Area:** Population health, Safety De.5 **IOM Quality Domain**: Effectiveness De.6 **Consumer Care Need:** Getting better

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

every 3 years. Yes, information provided in contact section	N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	C Y☐ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	D Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	Met Y☐ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:**	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	Eval Rating
(for NQF staff use) **Specific NPP goal**:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Leading cause of morbidity/mortality, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Approximately 2 million persons in the United States have heart failure each year. [1] These numbers will likely increase as the population ages. The literature suggests that hospitals have improved care for heart failure patients. In a study of 29,500 elderly patients in Oregon, the 3-day mortality decreased by 41% from 1991 to 1995. [2] 1a.4 **Citations for Evidence of High Impact:** [1] Smith, WM. Epidemiology of congestive heart failure. Am J Cardiol 1985;55(2):3A-8A. [2] Ni H, Hershberger FE. Was the decreasing trend in hospital mortality from heart failure attributable to improved hospital care? The Oregon experience, 1991-1995. Am J Manag Care 1999;5(9):1105-15.	1a C☐ P☐ M☐ N☐
1b. Opportunity for Improvement 1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** Congestive heart failure (CHF) is a progressive, chronic disease with substantial short-term mortality, which varies from provider to provider. Better processes of care may reduce short-term mortality, which represents better quality. 1b.2 **Summary of data demonstrating performance gap (variation or overall poor performance) across providers:**	1b C☐ P☐ M☐ N☐

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

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.

Adjusted per 1,000 rates by patient and hospital characteristics, 2007

Mean	Standard error	Location	P-value: Relative to Northeast
32.076	0.372	Northeast	1.000
25.200	0.341	Midwest	0.000
27.911	0.272	South	0.000
28.870	0.429	West	0.000

1b.3 Citations for data on performance gap:
See the following report for a complete treatment of the methodology: "Methods: Applying AHRQ Quality Indicators to Healthcare Cost and Utilization Project (HCUP) Data for the National Healthcare Quality Report" [URL: http://hcupnet.ahrq.gov/QI%20Methods.pdf?JS=Y]

1b.4 Summary of Data on disparities by population group:
Adjusted per 1,000 rates by patient/hospital characteristics, 2007

Estimate	Standard error	Age: for conditions affecting any age
12.234	0.537	18-44
15.070	0.276	45-64
33.634	0.216	65 and over

Estimate	Standard error	Age: for conditions affecting elderly
17.920	0.471	65-69
22.696	0.484	70-74
26.697	0.468	75-79
36.089	0.474	80-84
47.754	0.440	85 and over

Estimate	Standard error	Gender
27.718	0.248	Male
29.119	0.235	Female

Estimate	Standard error	Median income of patient´s ZIP code
30.165	0.309	First quartile (lowest income)
27.842	0.333	Second quartile
27.121	0.353	Third quartile
27.179	0.372	Fourth quartile (highest income)

Estimate	Standard error	Location of patient residence (NCHS)
25.547	0.316	Large central metropolitan
26.118	0.339	Large fringe metropolitan
25.217	0.382	Medium metropolitan
32.740	0.562	Small metropolitan
35.863	0.526	Micropolitan
38.123	0.651	Not metropolitan or micropolitan

Estimate	Standard error	Expected payment source
35.572	0.575	Private insurance
26.881	0.184	Medicare
29.834	0.885	Medicaid
57.840	1.615	Other insurance
34.378	1.437	Uninsured / self-pay / no charge

Estimate	Standard error	Hospital Ownership/control
27.378	0.197	Private, not-for-profit
28.834	0.449	Private, for-profit
33.192	0.507	Public
Estimate	Standard error	Teaching status
26.110	0.310	Teaching
29.164	0.203	Nonteaching
Estimate	Standard error	Location of hospital
25.569	0.297	Large central metropolitan
26.294	0.358	Large fringe metropolitan
25.442	0.370	Medium metropolitan
31.519	0.521	Small metropolitan
36.442	0.544	Micropolitan
48.180	0.894	Not metropolitan or micropolitan
Estimate	Standard error	Bed size of hospital
38.751	0.494	Less than 100
27.412	0.263	100 - 299
26.437	0.312	300 - 499
26.027	0.410	500 or more

1b.5 Citations for data on Disparities:
See the following report for a complete treatment of the methodology: "Methods: Applying AHRQ Quality Indicators to Healthcare Cost and Utilization Project (HCUP) Data for the National Healthcare Quality Report" [URL: http://hcupnet.ahrq.gov/QI%20Methods.pdf?JS=Y]

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*)**:** Congestive heart failure (CHF) is a progressive, chronic disease with substantial short-term mortality, which varies from provider to provider. Better processes of care may reduce short-term mortality, which represents better quality.

1c.2-3. Type of Evidence: Expert opinion, 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*)**:**
The existence of a board quality committee was associated with higher likelihoods of adopting various oversight practices and lower mortality rates for congestive heart failure measured by the Agency for Healthcare Research and Quality´s Inpatient Quality Indicators and the State Inpatient Databases. [1]

References:
[1] Jiang, H. Joanna; Lockee, Carlin; Bass, Karma; Fraser, Irene; Kiely, Robert. (2008) Board engagement in quality: findings of a survey of hospital and system leaders. Journal of Healthcare Management, 53, 2, 121(15)

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
6 Smoothing recommended Testing, rating, and review were conducted by the project team. A full report on the literature review and empirical evaluation can be found in Refinement of the HCUP Quality Indicators by the UCSF-Stanford EPC, Detailed coding information for each QI is provided in the document Prevention Quality Indicators Technical Specifications. Rating of performance on empirical evaluations, ranged from 0 to 26. The scores were intended as a guide for summarizing the performance of each indicator on four empirical tests of precision (signal variance, area-level share, signal ratio, and R-squared) and five tests of minimum bias (rank correlation, top and bottom decile movement, absolute change, and change over two deciles), as

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.

described in the previous section.

1c.6 Method for rating evidence: The project team conducted extensive empirical testing of all potential indicators using the 1995-97 HCUP State Inpatient Databases (SID) and Nationwide Inpatient Sample (NIS) to determine precision, bias, and construct validity. The 1997 SID contains uniform data on inpatient stays in community hospitals for 22 States covering approximately 60% of all U.S. hospital discharges. The NIS is designed to approximate a 20% of U.S. community hospitals and includes all stays in the sampled hospitals. Each year of the NIS contains between 6 million and 7 million records from about 1,000 hospitals. The NIS combines a subset of the SID data, hospital-level variables, and hospital and discharge weights for producing national estimates. The project team conducted tests to examine three things: precision, bias, and construct validity.

Precision. The first step in the analysis involved precision tests to determine the reliability of the indicator for distinguishing real differences in provider performance. For indicators that may be used for quality improvement, it is important to know with what precision, or surety, a measure can be attributed to an actual construct rather than random variation.

For each indicator, the variance can be broken down into three components: variation within a provider (actual differences in performance due to differing patient characteristics), variation among providers (actual differences in performance among providers), and random variation. An ideal indicator would have a substantial amount of the variance explained by between-provider variance, possibly resulting from differences in quality of care, and a minimum amount of random variation. The project team performed four tests of precision to estimate the magnitude of between-provider variance on each indicator:

- Signal standard deviation was used to measure the extent to which performance of the QI varies systematically across hospitals or areas.
- Provider/area variation share was used to calculate the percentage of signal (or true) variance relative to the total variance of the QI.
- Signal-to-noise ratio was used to measure the percentage of the apparent variation in QIs across providers that is truly related to systematic differences across providers and not random variations (noise) from year to year.
- In-sample R-squared was used to identify the incremental benefit of applying multivariate signal extraction methods for identifying additional signal on top of the signal-to-noise ratio.

In general, random variation is most problematic when there are relatively few observations per provider, when adverse outcome rates are relatively low, and when providers have little control over patient outcomes or variation in important processes of care is minimal. If a large number of patient factors that are difficult to observe influence whether or not a patient has an adverse outcome, it may be difficult to separate the "quality signal" from the surrounding noise. Two signal extraction techniques were applied to improve the precision of an indicator:

- Univariate methods were used to estimate the "true" quality signal of an indicator based on information from the specific indicator and 1 year of data.
- Multivariate signal extraction (MSX) methods were used to estimate the "true" quality signal based on information from a set of indicators and multiple years of data. In most cases, MSX methods extracted additional signal, which provided much more precise estimates of true hospital or area quality.

Bias. To determine the sensitivity of potential QIs to bias from differences in patient severity, unadjusted performance measures for specific hospitals were compared with performance measures that had been adjusted for age and gender. All of the PQIs and some of the Inpatient Quality Indicators (IQIs) could only be risk-adjusted for age and sex. The 3M™ APR-DRG System Version 12 with Severity of Illness and Risk of Mortality subclasses was used for risk adjustment of the utilization indicators and the in-hospital mortality indicators, respectively. Five empirical tests were performed to investigate the degree of bias in an indicator:

- Rank correlation coefficient of the area or hospital with (and without) risk adjustment—gives the overall impact of risk adjustment on relative provider or area performance.
- Average absolute value of change relative to mean—highlights the amount of absolute change in performance, without reference to other providers' performance.
- Percentage of highly ranked hospitals that remain in high decile—reports the percentage of hospitals or areas that are in the highest deciles without risk adjustment that remain there after risk adjustment is performed.
- Percentage of lowly ranked hospitals that remain in low decile—reports the percentage of hospitals or areas that are in the lowest deciles without risk adjustment that remain there after risk adjustment is performed.
- Percentage that change more than two deciles—identifies the percentage of hospitals whose relative rank changes by a substantial percentage (more than 20%) with and without risk adjustment.

Construct validity. Construct validity analyses provided information regarding the relatedness or independence of the indicators. If quality indicators do indeed measure quality, then two measures of the same construct would be expected to yield similar results. The team used factor analysis to reveal underlying patterns among large numbers of variables—in this case, to measure the degree of relatedness between indicators. In addition, they analyzed correlation matrices for indicators.

1c.7 **Summary of Controversy/Contradictory Evidence:** See the following for a complete treatment of the topic:
http://www.qualityindicators.ahrq.gov/downloads/iqi/iqi_guide_v31.pdf
Note: The Literature Review Caveats column summarizes evidence specific to each potential concern on the link between the PQIs and quality of care, as described in step 3 above. A question mark (?) indicates that the concern is theoretical or suggested, but no specific evidence was found in the literature. A check mark indicates that the concern has been demonstrated in the literature.

1c.8 **Citations for Evidence (***other than guidelines***):**
http://www.qualityindicators.ahrq.gov/downloads/iqi/iqi_guide_v31.pdf

1c.9 **Quote the Specific guideline recommendation (***including guideline number and/or page number***):**
Not Applicable.

1c.10 **Clinical Practice Guideline Citation:** Not Applicable.
1c.11 **National Guideline Clearinghouse or other URL:** Not Applicable.

1c.12 **Rating of strength of recommendation (***also provide narrative description of the rating and by whom***):**
Not Applicable.

1c.13 **Method for rating strength of recommendation** (*If different from USPSTF system, also describe rating and how it relates to USPSTF)***:**
Not Applicable.

1c.14 **Rationale for using this guideline over others:**
Not Applicable.

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report*?	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Evaluation Rating**
2a. MEASURE SPECIFICATIONS	
S.1 **Do you have a web page where current detailed measure specifications can be obtained?** S.2 **If yes, provide web page URL:**	
2a. Precisely Specified	
2a.1 **Numerator Statement (***Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** Number of deaths (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator. 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-specs** **C☐** **P☐** **M☐** **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.

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.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**
Number of deaths (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
All discharges, age 18 years and older, with a principal diagnosis code of CHF.

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

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 population being measured - including all codes, logic, and definitions***):**
All discharges, age 18 years and older, with a principal diagnosis code of CHF.
ICD-9-CM CHF diagnosis codes:
39891
RHEUMATIC HEART FAILURE
40201
MAL HYPERT HRT DIS W CHF
40211
BENIGN HYP HRT DIS W CHF
40291
HYPERTEN HEART DIS W CHF
40401
MAL HYPER HRT/REN W CHF
40403
MAL HYP HRT/REN W CHF&RF
40411
BEN HYPER HRT/REN W CHF
40413
BEN HYP HRT/REN W CHF&RF
40491
HYPER HRT/REN NOS W CHF
40493
HYP HT/REN NOS W CHF&RF
4280
CONGESTIVE HEART FAILURE
4281
LEFT HEART FAILURE
42820
SYSTOLIC HEART FAILURE NOS OCT02-
42821
AC SYSTOLIC HRT FAILURE OCT02-
42822
CHR SYSTOLIC HRT FAILURE OCT02-
42823
AC ON CHR SYST HRT FAIL OCT02-
4289
HEART FAILURE NOS
42830
DIASTOLIC HRT FAILURE NOS OCT02-
42831
AC DIASTOLIC HRT FAILURE OCT02-
42832
CHR DIASTOLIC HRT FAIL OCT02-

42833
AC ON CHR DIAST HRT FAIL OCT02-
42840
SYST/DIAST HRT FAIL NOS OCT02-
42841
AC SYST/DIASTOL HRT FAIL OCT02-
42842
CHR SYST/DIASTL HRT FAIL OCT02-
42843
AC/CHR SYST/DIA HRT FAIL OCT02-

Exclude cases:
• missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1 =missing)
• transferring to another short-term hospital (DISP=2)
• MDC 14 (pregnancy, childbirth, and puerperium)

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** •
missing discharge disposition (DISP=missing)
• transferring to another short-term hospital (DISP=2)
• MDC 14 (pregnancy, childbirth, and puerperium)

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
Exclude cases:
• missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1 =missing)
• transferring to another short-term hospital (DISP=2)
• MDC 14 (pregnancy, childbirth, and puerperium)

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
Gender, age (5-year age groups), race / ethnicity, primary payer, custom

2a.12-13 Risk Adjustment Type: Risk adjustment method widely or commercially available

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 a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age in years (in 5-year age groups), All Patient Refined-Diagnosis Related Group (APR-DRG) and APR-DRG risk-of-mortality subclass. The reference population used in the model is the universe of discharges for states that participate in the HCUP State Inpatient Databases (SID) for the year 2007 (updated annually), a database consisting of 43 states and approximately 30 million adult discharges. 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., hospital, state, and region). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.
Required data elements: Patient gender; age in years at admission; International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) principal and secondary diagnosis codes. A limited license 3M APR-DRG grouper is included with the AHRQ QI Software.

2a.15-17 Detailed risk model available Web page URL or attachment: URL
http://qualityindicators.ahrq.gov/downloads/iqi/IQI_Risk_Adjustment_Tables_(Version_4_2).pdf

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***):**
RATE: Each Inpatient Quality Indicator (IQI) 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 IQI rates. 1) Discharge-level data is used to mark inpatient records containing outcomes of interest. 2) Identify

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.

populations at risk. 3) Calculate observed rates. 4) For rates that are not risk-adjusted, the risk-adjusted rate equals the observed rate. 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 IQI, the shrinkage estimate reflects a reliability adjustment unique to each indicator. Full information on IQI algorithms and specification can be found at http://qualityindicators.ahrq.gov/Iqi_download.htm.

2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:
Significance testing is not prescribed by the software. Users may calculate a confidence interval for the risk-adjusted rates and a posterior probability interval for the smoothed rates at a 95% or 99% level. Users may define the relevant benchmark and the methods of discriminating performance according to their application.

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

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.***):**
The data source is hospital discharge data such as the HCUP State Inpatient Databases (SID) or equivalent using UB-04 coding standards. The data collection instrument is public-use AHRQ QI software available in SAS or Windows versions.

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL None
http://www.qualityindicators.ahrq.gov/software.htm

2a.29-31 Data dictionary/code table web page URL or attachment: URL None
http://www.qualityindicators.ahrq.gov/downloads/winqi/AHRQ_QI_Windows_Software_Documentation_V41a.pdf

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*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 Data/sample *(description of data/sample and size)*: Veterans Integrated Service Networks´ (VISNs); and VA versus non-VA (Nationwide Inpatient Sample) using VA inpatient data (2004-2007). [1]

A survey of hospital and system leaders (presidents/chief executive officers (CEOs)) that was conducted in the first six months of 2006 with a total of 562 respondents. Hospital-level data for these composite measures were produced by applying the IQI to the State Inpatient Databases (SID) of the Healthcare Cost and Utilization Project (HCUP) sponsored by AHRQ. The SID includes all-payer data on inpatient stays from virtually all community hospitals in each participating state. [2]

Using 1995 to 2000 data from New York state (n = 7,021,065), analysts compared mortality risk (odds ratio) for individuals with and without Alzheimer´s disease. [3]

We restricted our analysis to 20 states (4) for which HCUP State Inpatient Databases (SID) were available. There were 1,601 nonfederal, urban, general hospitals in those 20 states. Over 300 hospitals were eliminated from the sample because of key missing variables in the American Hospital Association (AHA) Annual Survey of Hospital data, which was also used for this study, or because they had missing observations for some of the

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.

measures that we used. Thus, our sample consisted of 1,290 urban, acute-care hospitals for which complete data were available for 2001. [4]

2b.2 Analytic Method *(type of reliability & rationale, method for testing)*:
VA-and VISN-level IQI observed rates, risk-adjusted rates, and observed to expected ratios (O/Es). We examined the trends in VA-and VISN-level rates using weighted linear regression, variation in VISN-level O/Es, and compared VA to non-VA trends. [1]

A t-test was used to determine the significance of differences in quality measures. [2]

Odds Ratio. [3]

A likelihood ratio test of the hypothesis that the coefficients on all of these variables were equal to 0 (lambda) = 35.3, p< .01). [4]

2b.3 Testing Results *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
VA in-hospital mortality rates for CHF Mortality were unchanged over time. The IQIs are easily applied to VA administrative data. They can be useful to tracks rate trends over time, reveal variation between sites, and for trend comparisons with other healthcare systems. [1]

The existence of a board quality committee was associated with higher likelihoods of adopting various oversight practices and lower mortality rates for congestive heart failure measured by the Agency for Healthcare Research and Quality´s Inpatient Quality Indicators and the State Inpatient Databases. [2]

Among men, adjusted odds of death were greater for those with Alzheimer´s disease (AD) for gastrointestinal congestive heart failure (CHF) (+42 percent, p < .0001). Among women, AD did not affect risks for most conditions although their risk for death from CHF was less than that for men with AD. [3]

The risk-adjusted mortality rate for congestive heart failure (CHF) is not significantly associated with costs. The AHRQ QIs have the advantage of taking the multidimensional nature of hospital quality into account. As the coefficients on the AHRQ QIs show, measures of hospital quality can have conflicting effects on hospital costs. A single measure that combines these effects into one variable offers less insight into hospital performance than the outcomes for each measure. [4]

References
[1] Borzecki AM, Christiansen CL, Loveland S, Chew P, Rosen AK. Trends in the inpatient quality indicators: the Veterans Health Administration experience. Med Care. 2010 Aug;48(8):694-702.
[2] Jiang, H. Joanna; Lockee, Carlin; Bass, Karma; Fraser, Irene; Kiely, Robert. (2008) Board engagement in quality: findings of a survey of hospital and system leaders. Journal of Healthcare Management, 53, 2, 121(15)
[3] Laditka JN, Laditka SB, Cornman CB. Evaluating hospital care for individuals with Alzheimer´s disease using inpatient quality indicators. Am J Alzheimers Dis Other Demen. 2005 Jan-Feb;20(1):27-36. PMID: 15751451.
[4] Laditka JN, Laditka SB, Cornman CB. Evaluating hospital care for individuals with Alzheimer´s disease using inpatient quality indicators. Am J Alzheimers Dis Other Demen. 2005 Jan-Feb;20(1):27-36. PMID: 15751451.

2c. Validity testing

2c.1 Data/sample *(description of data/sample and size)*: Retrospective cohort study based on 2.07 million inpatient admissions between 1998 and 2000 in the California State Inpatient Database. [1]

We used 2004-2007 Veterans Health Administration (VA) discharge and Vital Status files. [2]

2c.2 Analytic Method *(type of validity & rationale, method for testing)*:
The AHRQ IQI software was used to calculate risk-adjusted mortality rates using either (1) routine administrative data that included all the International Classification of Diseases (ICD)-9-CM codes or (2) enhanced administrative data that included only the ICD-9-CM codes representing preexisting conditions. [1]

2c
C☐
P☐
M☐
N☐

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

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

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

We derived 4-year facility-level in-hospital and 30-day observed mortality rates and observed/expected ratios (O/Es) for admissions with a principal diagnosis of acute myocardial infarction, congestive heart failure, stroke, gastrointestinal hemorrhage, hip fracture, and pneumonia. We standardized software-calculated O/Es to the VA population and compared O/Es and outlier status across sites using correlation, observed agreement, and kappas. [2]

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
Without using POA data, for congestive heart failure 25% of hospitals classified as low-quality hospitals using enhanced administrative data were misclassified as intermediate-quality hospitals using routine administrative data. Despite the fact that the AHRQ IQIs were primarily intended to serve as a screening tool, they are being increasingly used to publicly report hospital quality. These findings emphasized the need (which the AHRQ QI have now adopted by incorporating POA data in the risk-adjustment) to improve the "quality" of administrative data by including a POA indicator if these data are to serve as the information infrastructure for quality reporting. [1]

Of 119 facilities, in-hospital versus 30-day mortality O/E correlations were generally high (median: r = 0.78; range: 0.31-0.86). Examining outlier status, observed agreement was high (median: 84.7%, 80.7%-89.1%). Kappas showed at least moderate agreement (k > 0.40) for all indicators except stroke and hip fracture (k = 0.22). Across indicators, few sites changed from a high to nonoutlier or low outlier, or vice versa (median: 10, range: 7-13). The AHRQ IQI software can be easily adapted to generate 30-day mortality rates. Although 30-day mortality has better face validity as a hospital performance measure than in-hospital mortality, site assessments were similar despite the definition used. [3]

References
[1] Glance L.G.; Osler T.M.; Mukamel D.B.; Dick A.W. Impact of the present-on-admission indicator on hospital quality measurement: Experience with the Agency for Healthcare Research and Quality (AHRQ)Inpatient Quality Indicators. Medical Care, v. 46, no. 2, Feb. 2008, p. 112-119. DOI: 10.1097/MLR.0b013e318158aed6.
[2] Borzecki AM, Christiansen CL, Chew P, Loveland S, Rosen AK. Comparison of in-hospital versus 30-day mortality assessments for selected medical conditions. Med Care. 2010 Dec;48(12):1117-21. PMID: 20978451

2d. Exclusions Justified

2d.1 **Summary of Evidence supporting exclusion(s):**
Exclusions remove cases where the outcome of interest is less likely to be preventable or more likely to be preventable or with no or very low risk

2d.2 **Citations for Evidence:**
Refinement of the HCUP Quality Indicators (Technical Review), May 2001
http://qualityindicators.ahrq.gov/downloads/technical/qi_technical_review.zip

2d.3 **Data/sample** *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2d.4 **Analytic Method** *(type analysis & rationale)*:
Expert panel and descriptive analyses stratified by exclusion categories

2d.5 **Testing Results** *(e.g., frequency, variability, sensitivity analyses)*:
Refinement of the HCUP Quality Indicators (Technical Review), May 2001
http://qualityindicators.ahrq.gov/downloads/technical/qi_technical_review.zip

2d
C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 **Data/sample** *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2e
C☐ P☐ M☐ N☐

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;[Error! Bookmark not defined.] OR
rationale/data support no risk adjustment.

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

Risk-adjustment models use a standard set of categories based on readily available classification systems for demographics, severity of illness and comorbidities. Within each category, covariates are initially selected based on a minimum of 30 cases in the outcome of interest. Then a stepwise regression process on a development sample is used to select a parsimonious set of covariates where p<.05. Model is then tested on a validation sample.

2e.3 Testing Results *(risk model performance metrics)*:
c 0.787

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: Not applicable

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.

2f. Identification of Meaningful Differences in Performance

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
Posterior probability distribution parameterized using the Gamma distribution

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)*:

5th	25th	Median	75th	95th
0.017245	0.025607	0.032831	0.041305	0.055832

2f C☐ P☐ M☐ N☐

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.

2g. Comparability of Multiple Data Sources/Methods

2g.1 Data/sample *(description of data/sample and size)*: Not applicable

2g.2 Analytic Method *(type of analysis & rationale)*:
Not applicable

2g.3 Testing Results *(e.g., correlation statistics, comparison of rankings)*:
Not applicable

2g C☐ P☐ M☐ N☐ NA ☐

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

2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*: Median income of patient´s ZIP code:
1) Estimate 2) Standard error 3) P-value: Relative to marked group-c 4) P-value:
2007 relative to 2006
First quartile (lowest income) 30.165 0.309 0.000 0.000
Second quartile 27.842 0.333 0.184 0.000
Third quartile 27.121 0.353 0.909 0.000
Fourth quartile (highest income)c 27.179 0.372 0.000

[1] Although we did find overall disparities in care, we found that indicators for blacks, Hispanics, and Asians were not statistically worse than corresponding quality indicators for whites in the same hospital. Only a few hospitals provide lower quality of care to minorities than to whites.

[1] Darrell J. Gaskin, Christine S. Spencer, Patrick Richard, Gerard F. Anderson, Neil R. Powe and Thomas A. LaVeist. Do Hospitals Provide Lower-Quality Care To Minorities Than To Whites? Health Affairs, 27, no. 2 (2008): 518-527 doi: 10.1377/hlthaff.27.2.518

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:
Users may stratify based on gender and race/ethnicity

2h C☐ P☐ M☐ N☐ NA ☐

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

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific* **2**

Acceptability of Measure Properties?	
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* **C** ☐ **P** ☐ **M** ☐ **N** ☐

3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Evaluation**
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** In use 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** Arizona (NY QIO) Why Not the Best? http://www.http://whynotthebest.org/ California (state) Hospital Inpatient Mortality Indicators for California http://www.oshpd.ca.gov/HID/Products/PatDischargeData/AHRQ/iqi-imi_overview.html Colorado (state hospital association) Colorado Hospital Report Card http://www.cohospitalquality.org/index.php?option=com_frontpage&Itemid=1 Florida (state) Florida Health Finder http://www.floridahealthfinder.gov/ Illinois (state) Illinois Hospital Report Card and Consumer Guide to Health Care http://www.healthcarereportcard.illinois.gov/ Iowa (Iowa Healthcare Collaborative) Iowa Healthcare Collaborative http://www.ihconline.org/aspx/publicreporting/iowareport.aspx Kentucky (Norton Healthcare, a hospital system) Norton Healthcare Quality Report http://www.nortonhealthcare.com/body.cfm?id=157 Kentucky (state) Health Care Information Center http://chfs.ky.gov/ohp/healthdata Kentucky (state hospital association) Kentucky Hospital Association Quality Data http://info.kyha.com/QualityData/IQISite/ Maine (state)	**3a** **C** ☐ **P** ☐ **M** ☐ **N** ☐

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

Maine Health Data Organization
http://gateway.maine.gov/mhdo2008Monahrq/home.html

Massachusetts (state)
My HealthCare Options
http://www.mass.gov/healthcareqc

New Hampshire (NY QIO)
New York State Health Accountability Foundation
http://nyshaf.org/juice/IPROSpikeChart.html

New Jersey (state)
Find and Compare Quality Care in NJ Hospitals
http://www.nj.gov/health/healthcarequality/

New York (health care coalition)
New York State Hospital Report Card
http://www.myhealthfinder.com/

Oregon (state)
Oregon Hospital Quality Indicators
http://www.oregon.gov/OHPPR/HQ/

Rhode Island (NY QIO)
Why Not the Best?
http://www.http://whynotthebest.org/

Texas (state)
Reports on Hospital Performance
http://www.dshs.state.tx.us/thcic/

Utah (state)
Utah Hospital Comparison Reports
http://health.utah.gov/myhealthcare/

Washington (health care coalition)
Washington State Hospital Report Card
http://www.myhealthfinder.com/wa09/index.php

Wisconsin (state hospital association)
CheckPoint
http://www.wicheckpoint.org/index.aspx

The measures is also reported on HCUPnet:
http://hcupnet.ahrq.gov/HCUPnet.jsp?Id=EB57801381F71C41&Form=MAINSEL&JS=Y&Action=%3E%3ENext%3E%3E&_MAINSEL=AHRQ%20Quality%20Indicators

This measure is used in the MONAHRQ system that is provided for public reporting and quality improvement throughout the United States: http://monahrq.ahrq.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***):**
University Healthcare Consortium - An alliance of 103 academic medical centers and 219 of their affiliated hospitals. Reporting the AHRQ QIs to their member hospitals. (see www.uhc.edu. Note: measure results reported to hospitals; not reported on site).

Dallas Fort Worth Hospital Council – Reporting on measure results to over 70 hospitals in Texas (see

www.dfwhc.ord. Note: measure results reported to hospitals; not reported on site).

Norton Healthcare - a multi-hospital system in Kentucky (see http://www.nortonhealthcare.com/about/Our_Performance/index.aspx)
Ministry Health Care - a multi-hospital system in Wisconsin (see http://ministryhealth.org/display/router.aspx. Note: measure results reported to hospitals; not reported on site).

Minnesota Hospital Association
http://www.mnhospitals.org/ Note: measure used in quality improvement. Not reported publicly by the association)

Premier - Premier´s "Quality Advisor" tool provides performance reports to approximately 650 hospitals for their use in monitoring and improving quality. Hospitals receive facility specific reports on this measure in Quality Advisor.

This measure is used in the MONAHRQ system that is provided for public reporting and quality improvement throughout the United States: http://monahrq.ahrq.gov/

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)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

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). 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.
The Model Reports (discussed immediately above) are based on:
• 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.

3a.6 **Results** *(qualitative and/or quantitative results and conclusions)*:
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:**
CMS CHF Mortality Measure

(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):
3b.2 **Are the measure specifications harmonized? If not, why?**
The specifications are harmonized, but CMS uses 30-day mortality

C☐ P☐ M☐ N☐ NA

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

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

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

	☐
3c. Distinctive or Additive Value 3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** The AHRQ is all-payer (not Medicare FFS only) and uses in-hospital mortality, which is available in real-time 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:** The AHRQ measure provides a real-time indication of hospital performances, reflects the patient´s experience in the hospital, and is available for all-payers	**3c** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, met? **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	Eval Rating
4a. Data Generated as a Byproduct of Care Processes 4a.1-2 **How are the data elements that are needed to compute measure scores generated?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources 4b.1 **Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) Yes 4b.2 **If not, specify the near-term path to achieve electronic capture by most providers.**	**4b** C☐ P☐ M☐ N☐
4c. Exclusions 4c.1 **Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No 4c.2 **If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA ☐
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.** Coding professionals follow detail guidelines, are subject to training and credentialing requirements, peer review and audit. Risk-adjusted measures of mortality may lead to an increase in coding of comorbidities. All in-hospital mortality measures may encourage earlier post-operative discharge, and thereby shift deaths to skilled nursing facilities or outpatient settings. However, Rosenthal et al. found no evidence that hospitals with lower in-hospital standardized mortality had higher (or lower) early post-discharge mortality. [1] Coding professionals follow detailed guidelines, are subject to training and credentialing requirements, peer review and audit.	**4d** C☐ P☐ M☐ N☐

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.

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

References:
[1] Rosenthal GE, Baker DW, Norris DG, et al. Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. Health Serv Res 2000;34(7):1449-68.

4e. Data Collection Strategy/Implementation

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).

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:**
Relative to other indicators, a lower percentage of the variation occurs at the provider level rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is moderate, at 53.5%, indicating that some of the observed differences in provider performance likely do not represent true differences.

4e.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:**
All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at:
http://www.qualityindicators.ahrq.gov/software.htm

4e.3 **Evidence for costs:**
All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at:
http://www.qualityindicators.ahrq.gov/software.htm

4e.4 **Business case documentation:** All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at:
http://www.qualityindicators.ahrq.gov/software.htm

4e
C☐ P☐ M☐ N☐

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*? **4**

Steering Committee: Overall, to what extent was the criterion, *Feasibility*, met? **4**
C☐ P☐ M☐ N☐
Rationale:

RECOMMENDATION

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

Steering Committee: Do you recommend for endorsement?
Y☐ N☐ A☐
Comments:

CONTACT INFORMATION

Co.1 **Measure Steward (Intellectual Property Owner)**
Co.1 **Organization**
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.2 **Point of Contact**
John, Bott, MSSW, MBA, david.atkins@ahrq.hhs.gov, 301-427-1608-

Measure Developer If different from Measure Steward
Co.3 **Organization**
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.4 Point of Contact
John, Bott, MSSW, MBA, david.atkins@ahrq.hhs.gov, 301-427-1608-

Co.5 Submitter If different from Measure Steward POC
John, Bott, MSSW, MBA, david.atkins@ahrq.hhs.gov, 301-427-1608-, Agency for Healthcare Research and Quality

Co.6 Additional organizations that sponsored/participated in measure development
UC Davis,
Stanford University,
Battelle Memorial Institute

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.
None

Ad.2 If adapted, provide name of original measure: None
Ad.3-5 If adapted, provide original specifications URL or attachment

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 Year the measure was first released: 2001
Ad.7 Month and Year of most recent revision: 10, 2010
Ad.8 What is your frequency for review/update of this measure? Annual
Ad.9 When is the next scheduled review/update for this measure? 05, 2011

Ad.10 Copyright statement/disclaimers: The AHRQ QI software is publicly available; no copyright disclaimers

Ad.11 -13 Additional Information web page URL or attachment:

Date of Submission (*MM/DD/YY*)**:** 02/01/2011

Page 4: [1] Comment [k4]	Karen Pace	10/5/2009 8:59:00 AM

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.

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 #:** 0277 **NQF Project:** Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Congestive Heart Failure Admission Rate (PQI 8)
De.2 **Brief description of measure***:* Percent of county population with an admissions for CHF.
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** Prevention Quality Indicator (PQI) composite
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**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	**A** **Y**☐ **N**☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** **Y**☐ **N**☐

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

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	**Eval Rating**
(for NQF staff use) **Specific NPP goal**:	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** High resource use, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** Billings et al. found that low-income ZIP codes in New York City had 4.6 times more CHF hospitalizations per capita than high-income ZIP codes. Millman et al. reported that low-income ZIP codes had 6.1 times more CHF hospitalizations per capita than high-income ZIP codes.65 Based on empirical results, areas with high rates of CHF also tend to have high rates of admission for other ACSCs. 1a.4 **Citations for Evidence of High Impact:** 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. Millman M, editor. Committee on Monitoring Access to Personal Health Care Services. Washington DC: National Academy Press.	**1a** C☐ P☐ M☐ N☐
1b. Opportunity for Improvement 1b.1 **Benefits (improvements in quality) envisioned by use of this measure:** Congestive heart failure is a PQI that would be of most interest to comprehensive health care delivery systems. This indicator is measured with high precision, and most of the observed variance reflects true differences across areas. Risk adjustment for age and sex appears to affect the areas with the highest and lowest raw rates. Areas with high rates may wish to examine the clinical characteristics of their patients to check for a more complex case mix. Patient age, clinical measures such as heart function, and other management issues may affect admission rates.	**1b** C☐ P☐ M☐ N☐

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

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).

As the causes for admissions may include poor quality care, lack of patient compliance, or problems accessing care, areas may wish to review CHF patient records to identify precipitating causes and potential targets for intervention.

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
Adjusted per 100,000 rates by patient and hospital characteristics, 2007

Mean	Standard error	Location	P-value: Relative to Northeast
402.605	22.318	Northeast	1.000
446.773	21.686	Midwest	0.156
474.166	17.900	South	0.012
293.022	11.579	West	0.000

1b.3 Citations for data on performance gap:
See the following report for a complete treatment of the methodology: "Methods: Applying AHRQ Quality Indicators to Healthcare Cost and Utilization Project (HCUP) Data for the National Healthcare Quality Report" [URL: http://hcupnet.ahrq.gov/QI%20Methods.pdf?JS=Y]

1b.4 Summary of Data on disparities by population group:
Adjusted per 100,000 rates by patient characteristics, 2007

Estimate	Standard error	Age: for conditions affecting any age
38.527	1.828	18-44
298.394	10.627	45-64
1912.391	43.139	65 and over

Estimate	Standard error	Age: for conditions affecting elderly
835.456	22.964	65-69
1243.6	30.172	70-74
1845.486	43.594	75-79
2841.152	69.354	80-84
4453.902	114.115	85 and over

Estimate	Standard error	Gender
474.842	11.383	Male
370.707	8.504	Female

Estimate	Standard error	Median income of patient´s ZIP code
561.781	25.3	First quartile (lowest income)
420.838	16.952	Second quartile
361.98	14.697	Third quartile
319.623	20.016	Fourth quartile (highest income)

Estimate	Standard error	Location of patient residence (NCHS)
442.037	34.923	Large central metropolitan
413.407	31.738	Large fringe metropolitan
380.89	36.494	Medium metropolitan
398.905	45.931	Small metropolitan
417.946	23.022	Micropolitan
430.314	20.094	Not metropolitan or micropolitan

1b.5 Citations for data on Disparities:
See the following report for a complete treatment of the methodology: "Methods: Applying AHRQ Quality Indicators to Healthcare Cost and Utilization Project (HCUP) Data for the National Healthcare Quality Report" [URL: http://hcupnet.ahrq.gov/QI%20Methods.pdf?JS=Y]

1c. Outcome or Evidence to Support Measure Focus — **1c** C☐

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.

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*)**:** Congestive heart failure (CHF) can be controlled in an outpatient setting for the most part. If area rates for CHF are high even after risk adjustment and stratification, the quality of preventive services in that region are held to be insufficient in preparing CHF patients to manage their condition.

P☐ M☐ N☐

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

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*)**:**
Congestive heart failure is a PQI that would be of most interest to comprehensive health care delivery systems. This indicator is measured with high precision, and most of the observed variance reflects true differences across areas.
Risk adjustment for age and sex appears to affect the areas with the highest and lowest raw rates. Areas with high rates may wish to examine the clinical characteristics of their patients to check for a more complex case mix. Patient age, clinical measures such as heart function, and other management issues may affect admission rates.
As the causes for admissions may include poor quality care, lack of patient compliance, or problems accessing care, areas may wish to review CHF patient records to identify precipitating causes and potential targets for intervention.

1c.5 Rating of strength/quality of evidence (*also provide narrative description of the rating and by whom***):**
RATING: 14 Testing, rating, and review were conducted by the project team. A full report on the literature review and empirical evaluation can be found in Refinement of the HCUP Quality Indicators by the UCSF-Stanford EPC, Detailed coding information for each QI is provided in the document Prevention Quality Indicators Technical Specifications. Rating of performance on empirical evaluations, ranged from 0 to 26. The scores were intended as a guide for summarizing the performance of each indicator on four empirical tests of precision (signal variance, area-level share, signal ratio, and R-squared) and five tests of minimum bias (rank correlation, top and bottom decile movement, absolute change, and change over two deciles), as described in the previous section.

1c.6 Method for rating evidence: The project team conducted extensive empirical testing of all potential indicators using the 1995-97 HCUP State Inpatient Databases (SID) and Nationwide Inpatient Sample (NIS) to determine precision, bias, and construct validity. The 1997 SID contains uniform data on inpatient stays in community hospitals for 22 States covering approximately 60% of all U.S. hospital discharges. The NIS is designed to approximate a 20% of U.S. community hospitals and includes all stays in the sampled hospitals. Each year of the NIS contains between 6 million and 7 million records from about 1,000 hospitals. The NIS combines a subset of the SID data, hospital-level variables, and hospital and discharge weights for producing national estimates. The project team conducted tests to examine three things: precision, bias, and construct validity.
Precision. The first step in the analysis involved precision tests to determine the reliability of the indicator for distinguishing real differences in provider performance. For indicators that may be used for quality improvement, it is important to know with what precision, or surety, a measure can be attributed to an actual construct rather than random variation.
For each indicator, the variance can be broken down into three components: variation within a provider (actual differences in performance due to differing patient characteristics), variation among providers (actual differences in performance among providers), and random variation. An ideal indicator would have a substantial amount of the variance explained by between-provider variance, possibly resulting from differences in quality of care, and a minimum amount of random variation. The project team performed four tests of precision to estimate the magnitude of between-provider variance on each indicator:
• Signal standard deviation was used to measure the extent to which performance of the QI varies systematically across hospitals or areas.
• Provider/area variation share was used to calculate the percentage of signal (or true) variance relative to the total variance of the QI.
• Signal-to-noise ratio was used to measure the percentage of the apparent variation in QIs across providers that is truly related to systematic differences across providers and not random variations (noise) from year to year.
• In-sample R-squared was used to identify the incremental benefit of applying multivariate signal extraction

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.

methods for identifying additional signal on top of the signal-to-noise ratio.
In general, random variation is most problematic when there are relatively few observations per provider, when adverse outcome rates are relatively low, and when providers have little control over patient outcomes or variation in important processes of care is minimal. If a large number of patient factors that are difficult to observe influence whether or not a patient has an adverse outcome, it may be difficult to separate the "quality signal" from the surrounding noise. Two signal extraction techniques were applied to improve the precision of an indicator:
• Univariate methods were used to estimate the "true" quality signal of an indicator based on information from the specific indicator and 1 year of data.
• Multivariate signal extraction (MSX) methods were used to estimate the "true" quality signal based on information from a set of indicators and multiple years of data. In most cases, MSX methods extracted additional signal, which provided much more precise estimates of true hospital or area quality.
Bias. To determine the sensitivity of potential QIs to bias from differences in patient severity, unadjusted performance measures for specific hospitals were compared with performance measures that had been adjusted for age and gender. All of the PQIs and some of the Inpatient Quality Indicators (IQIs) could only be risk-adjusted for age and sex. The 3M™ APR-DRG System Version 12 with Severity of Illness and Risk of Mortality subclasses was used for risk adjustment of the utilization indicators and the in-hospital mortality indicators, respectively. Five empirical tests were performed to investigate the degree of bias in an indicator:
• Rank correlation coefficient of the area or hospital with (and without) risk adjustment—gives the overall impact of risk adjustment on relative provider or area performance.
• Average absolute value of change relative to mean—highlights the amount of absolute change in performance, without reference to other providers' performance.
• Percentage of highly ranked hospitals that remain in high decile—reports the percentage of hospitals or areas that are in the highest deciles without risk adjustment that remain there after risk adjustment is performed.
• Percentage of lowly ranked hospitals that remain in low decile—reports the percentage of hospitals or areas that are in the lowest deciles without risk adjustment that remain there after risk adjustment is performed.
• Percentage that change more than two deciles—identifies the percentage of hospitals whose relative rank changes by a substantial percentage (more than 20%) with and without risk adjustment.
Construct validity. Construct validity analyses provided information regarding the relatedness or independence of the indicators. If quality indicators do indeed measure quality, then two measures of the same construct would be expected to yield similar results. The team used factor analysis to reveal underlying patterns among large numbers of variables—in this case, to measure the degree of relatedness between indicators. In addition, they analyzed correlation matrices for indicators.

1c.7 Summary of Controversy/Contradictory Evidence: See the following for a complete treatment of the topic: http://www.qualityindicators.ahrq.gov/downloads/pqi/pqi_guide_v31.pdf
Note: The Literature Review Findings column summarizes evidence specific to each potential concern on the link between the PQIs and quality of care, as described in step 3 above. A question mark (?) indicates that the concern is theoretical or suggested, but no specific evidence was found in the literature. A check mark indicates that the concern has been demonstrated in the literature.

1c.8 Citations for Evidence (*other than guidelines***):**
http://www.qualityindicators.ahrq.gov/downloads/pqi/pqi_guide_v31.pdf

1c.9 Quote the Specific guideline recommendation (*including guideline number and/or page number***):**
Not Applicable.

1c.10 Clinical Practice Guideline Citation: Not Applicable.
1c.11 National Guideline Clearinghouse or other URL: Not Applicable.

1c.12 Rating of strength of recommendation (*also provide narrative description of the rating and by whom***):**
Not Applicable.

1c.13 Method for rating strength of recommendation (*If different from USPSTF system, also describe rating and how it relates to USPSTF)***:**
Not Applicable.

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: Not Applicable.	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report*, **met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Eval Rating**
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:**	
2a. Precisely Specified	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** All discharges of age 18 years and older with ICD-9-CM principal diagnosis code for CHF. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** Time period is user defined. Users of the measure typically use a 12 month time period. **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** All discharges of age 18 years and older with ICD-9-CM principal diagnosis code for CHF. Include ICD-9-CM diagnosis codes: 39891 RHEUMATIC HEART FAILURE 4280 CONGESTIVE HEART FAILURE 4281 LEFT HEART FAILURE 42820 SYSTOLIC HRT FAILURE NOS OCT02- 42821 AC SYSTOLIC HRT FAILURE OCT02- 42822 CHR SYSTOLIC HRT FAILURE OCT02- 42823 AC ON CHR SYST HRT FAIL OCT02- 42830 DIASTOLC HRT FAILURE NOS OCT02- 42831 AC DIASTOLIC HRT FAILURE OCT02- 42832 CHR DIASTOLIC HRT FAIL OCT02- 42833 AC ON CHR DIAST HRT FAIL OCT02- 42840 SYST/DIAST HRT FAIL NOS OCT02- 42841 AC SYST/DIASTOL HRT FAIL OCT02-	**2a-specs** **C☐** **P☐** **M☐** **N☐**

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

42842
CHR SYST/DIASTL HRT FAIL OCT02-
42843
AC/CHR SYST/DIA HRT FAIL OCT02-
4289
HEART FAILURE NOS
Include ICD-9-CM diagnosis codes ONLY for discharges before 2002Q3 (ending September 30, 2002):
40201
MAL HYPERT HRT DIS W CHF
40211
BENIGN HYP HRT DIS W CHF
40291
HYPERTEN HEART DIS W CHF
40401
MAL HYPER HRT/REN W CHF
40403
MAL HYP HRT/REN W CHF/RF
40411
BEN HYPER HRT/REN W CHF
40413
BEN HYP HRT/REN W CHF/RF
40491
HYPER HRT/REN NOS W CHF
40493
HYP HT/REN NOS W CHF/RF

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 cardiac procedure code
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
AORTCOR 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 REVASC
363
OTH HEART REVASCULAR
3631
OPEN CHEST TRANS REVASC
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)
3751
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
AHRQ Quality Indicators Web Site: http://www.qualityindicators.ahrq.gov
Prevention Quality Indicators Technical Specifications Version 4.2- 2010
PQI #8 Congestive Heart Failure (CHF) Admission Rate Page 3
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

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: 18 and older

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator***):**
Time period is user defined. Users of the measure typically use a 12 month time period.

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions***):**
Population in Metro Area or county, age 18 years and older.

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** none

2a.10 Denominator Exclusion Details (*All information required to collect exclusions to the denominator, including all codes, logic, and definitions***):**
Not applicable

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 gender, age (5-year age groups), race / ethnicity

2a.12-13 Risk Adjustment Type: Risk adjustment method widely or commercially available

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 a logistic regression model and covariates for gender and age in years (in 5-year age groups). The reference population used in the model is the universe of discharges for states that participate in the HCUP State Inpatient Databases (SID) for the year 2007 (updated annually), a database consisting of 43 states and approximately 30 million adult discharges. 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, state, and region). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate

2a.15-17 Detailed risk model available Web page URL or attachment: URL None
http://qualityindicators.ahrq.gov/downloads/pqi/PQI_Risk_Adjustment_Tables_(Version_4_2).pdf

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***):**
Each indicator is expressed as a rate, is defined as outcome of interest / population at risk or numerator / denominator. The AHRQ Quality Indicators (AHRQ QI) software performs five steps to produce the rates. 1) Discharge-level data is used to mark inpatient records containing the outcome of interest and 2) the population at risk. For provider indicators, the population at risk is also derived from hospital discharge records; for area indicators, the population at risk is derived from U.S. Census data. 3) Calculate observed

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.

rates. Using output from steps 1 and 2, rates are calculated for user-specified combinations of stratifiers. 4) Calculate expected rates. Regression coefficients from a reference population database are applied to the discharge records and aggregated to the provider or area level. 5) Calculate risk-adjusted rate. Use the indirect standardization to account for case-mix. 6) Calculate smoothed rate. A Univariate shrinkage factor is applied to the risk-adjusted rates. The shrinkage estimate reflects a reliability adjustment unique to each indicator. Full information on calculation algorithms and specifications can be found at http://qualityindicators.ahrq.gov/PQI_download.htm

2a.22 **Describe the method for discriminating performance** *(e.g., significance testing)***:**
Significance testing is not prescribed by the software. Users may calculate a confidence interval for the risk-adjusted rates and a posterior probability interval for the smoothed rates at a 95% or 99% level. Users may define the relevant benchmark and the methods of discriminating performance according to their application.

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

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.***):**
The data source is hospital discharge data such as the HCUP State Inpatient Databases (SID) or equivalent using UB-04 coding standards. The data collection instrument is public-use AHRQ QI software available in SAS or Windows versions.

2a.26-28 **Data source/data collection instrument reference web page URL or attachment:** URL None
http://www.qualityindicators.ahrq.gov/software.htm

2a.29-31 **Data dictionary/code table web page URL or attachment:** URL None
http://www.qualityindicators.ahrq.gov/downloads/winqi/AHRQ_QI_Windows_Software_Documentation_V41a.pdf

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

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

2a.38-41 **Clinical Services** (*Healthcare services being measured, check all that apply*)
Clinicians: Physicians (MD/DO)

TESTING/ANALYSIS

2b. Reliability testing

2b.1 **Data/sample** *(description of data/sample and size)***:** AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2b.2 **Analytic Method** *(type of reliability & rationale, method for testing)***:**
Expert panels and empirical analysis

2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
Relatively precise estimates of admission rates for CHF can be obtained, although random variation may be important for small hospitals and rural areas. Based on empirical evidence, this indicator is very precise, with a raw area level rate of 521.0 per 100,000 population and a standard deviation of 286.5. 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 93.0%, indicating that the observed differences in age-sex adjusted rates very likely represent true differences across areas.

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.

2c. Validity testing

2c.1 **Data/sample** *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2c.2 **Analytic Method** *(type of validity & rationale, method for testing)*:
Expert panels and empirical analysis

2c.3 **Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
Billings et al. found that low-income ZIP codes in New York City had 4.6 times more CHF hospitalizations per capita than high-income ZIP codes.64 Millman et al. reported that low-income ZIP codes had 6.1 times more CHF hospitalizations per capita than high-income ZIP codes.65
Based on empirical results, areas with high rates of CHF also tend to have high rates of admission for other ACSCs.

2c
C☐ P☐ M☐ N☐

2d. Exclusions Justified

2d.1 **Summary of Evidence supporting exclusion(s)**:
Exclusions remove cases where the outcome of interest is less likely to be preventable or with no or very low risk

2d.2 **Citations for Evidence:**
Refinement of the HCUP Quality Indicators (Technical Review), May 2001
http://qualityindicators.ahrq.gov/downloads/technical/qi_technical_review.zip

2d.3 **Data/sample** *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2d.4 **Analytic Method** *(type analysis & rationale)*:
Expert panel and descriptive analyses stratified by exclusion categories

2d.5 **Testing Results** *(e.g., frequency, variability, sensitivity analyses)*:
Refinement of the HCUP Quality Indicators (Technical Review), May 2001
http://qualityindicators.ahrq.gov/downloads/technical/qi_technical_review.zip

2d
C☐ P☐ M☐ N☐ NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 **Data/sample** *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

2e.2 **Analytic Method** *(type of risk adjustment, analysis, & rationale)*:
Risk-adjustment models use a standard set of categories based on readily available classification systems for demographics, severity of illness and comorbidities. Within each category, covariates are initially selected based on a minimum of 30 cases in the outcome of interest. Then a stepwise regression process on a development sample is used to select a parsimonious set of covariates where p<.05. Model is then tested on a validation sample

2e.3 **Testing Results** *(risk model performance metrics)*:
c-statistic not reported

2e.4 **If outcome or resource use measure is not risk adjusted, provide rationale:** Not applicable

2e
C☐ P☐ M☐ N☐ NA☐

2f. Identification of Meaningful Differences in Performance

2f.1 **Data/sample from Testing or Current Use** *(description of data/sample and size)*: AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

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

2f
C☐ P☐ M☐ N☐

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
... [1]

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

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

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 w
... [3]

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 cessatio
... [4]

Posterior probability distribution parameterized using the Gamma distribution

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)*:

5th	25th	Median	75th	95th
0.001361	0.002526	0.003658	0.005090	0.007724

2g. Comparability of Multiple Data Sources/Methods

2g.1 **Data/sample** *(description of data/sample and size)*: Not applicable

2g.2 **Analytic Method** *(type of analysis & rationale)*:
Not applicable

2g.3 **Testing Results** *(e.g., correlation statistics, comparison of rankings)*:
Not applicable

2g C☐ P☐ M☐ N☐ NA☐

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

2h. Disparities in Care

2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: Median income of patient´s ZIP code:
1) Estimate 2) Standard error 3) P-value: Relative to marked group-c 4) P-value:
2007 relative to 2006
First quartile (lowest income) 100.330 5.768 0.000 0.069
Second quartile 60.771 2.840 0.000 0.021
Third quartile 47.923 2.472 0.007 0.011
Fourth quartile (highest income)c 38.217 2.572 0.176

2h.2 **If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:**
Users may stratify based on gender and race/ethnicity

2h C☐ P☐ M☐ N☐ NA☐

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

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

2

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

2 C☐ P☐ M☐ N☐

3. USABILITY

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

Eval Rating

3a. Meaningful, Understandable, and Useful Information

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

3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*):
1) State of California:
http://www.oshpd.ca.gov/hid/products/preventable_hospitalizations/pdfs/PH_REPORT_WEB.pdf
2) State of New Jersey: Find and Compare Quality Care in New Jersey Hospitals,
http://www.nj.gov/health/healthcarequality/

3a C☐ P☐ M☐ N☐

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

3) Niagara Health Quality Coalition and Alliance for Quality Health Care: New York State Hospital Report Card, http://www.myhealthfinder.com/
4) State of Texas: Reports on Hospital Performance, http://www.dshs.state.tx.us/thcic/
5) Maine: Maine Health Data Organization: http://gateway.maine.gov/mhdo2008Monahrq/home.html
6) Hawaii: awaii Health Information Corporation: http://hhic.org/publicreports.asp
7) Nevada: Nevada Compare Care: http://www.nevadacomparecare.net/monahrq/home.html

In use as a part of the AHRQ Quality Indicators. They are reported in numerous forums including: http://hcupnet.ahrq.gov/HCUPnet.jsp?Id=EB57801381F71C41&Form=MAINSEL&JS=Y&Action=%3E%3ENext%3E%3E&_MAINSEL=AHRQ%20Quality%20Indicators

This measure is used in the Monahrq system that is provide for public reporting and quality improvement throughout the United States: http://monahrq.ahrq.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). If not used for QI, state the plans to achieve use for QI within 3 years***):**
The software is publicly available free of charge (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 Indicator software.

This measure is used in the Monahrq system that is provided for public reporting and quality improvement throughout the United States: http://monahrq.ahrq.gov/

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)***:** AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges

3a.5 **Methods** *(e.g., focus group, survey, QI project)***:**
AHRQ has developed the Quality Indicators Mapping Tool to facilitate use of the Prevention Quality Indicators and incorporated the tool into the MONAHRQ software, which has undergone user beta testing and is now available for download.

3a.6 **Results** *(qualitative and/or quantitative results and conclusions)***:**
Several states including Maine, Hawaii and Nevada have begun public reporting using the MONAHRQ tool. See http://monahrq.ahrq.gov/

3b/3c. Relation to other NQF-endorsed measures

3b.1 **NQF # and Title of similar or related measures:**
None found.

(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):
3b.2 **Are the measure specifications harmonized? If not, why?**
No competing measures found.

C☐ P☐ M☐ N☐ NA☐

3c. Distinctive or Additive Value — **3c**
3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:**
No competing measures found.

C☐ P☐ M☐ N☐ NA

5.1 **If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same**

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).

target population), Describe why it is a more valid or efficient way to measure quality: No competing measures found.	☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, met? **Rationale:**	**3** **C☐ P☐ M☐ N☐**
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	**Eval Rating**
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** **C☐ P☐ M☐ N☐**
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**4b** **C☐ P☐ M☐ N☐**
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** **C☐ P☐ M☐ N☐ NA☐**
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.** Coding professionals follow detail guidelines, are subject to training and credentialing requirements, peer review and audit. As a PQI, CHF is not a measure of hospital quality, but rather one measure of outpatient and other health care. Providers may reduce admission rates without actually improving quality by shifting care to an outpatient setting. Some CHF care takes place in emergency rooms. As such, combining inpatient and emergency room data may give a more accurate picture of this indicator. Physician management of patients with congestive heart failure differs significantly by physician specialty. [1, 2] Such differences in community practices may be reflected in differences in CHF admission rates. [1] Edep ME, Shah NB, Tateo IM, et al. Differences between primary care physicians and cardiologists in management of congestive heart failure: relation to practice guidelines. J Am Coll Cardiol 1997;30(2):518-26. [2] Reis, SE, Holubkov R, Edmundowicz D, et al. Treatment of patients admitted to the hospital with congestive heart failure: specialty-related disparities in practice patterns	**4d** **C☐ P☐ M☐ N☐**
4e. Data Collection Strategy/Implementation	**4e**

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.

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).

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: This indicator is measured with high precision, and most of the observed variance reflects true differences across areas. Risk adjustment for age and sex appears to affect the areas with the highest and lowest raw rates. Areas with high rates may wish to examine the clinical characteristics of their patients to check for a more complex case mix. Patient age, clinical measures such as heart function, and other management issues may affect admission rates. As the causes for admissions may include poor quality care, lack of patient compliance, or problems accessing care, areas may wish to review CHF patient records to identify precipitating causes and potential targets for intervention. **4e.2 Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:** All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm **4e.3 Evidence for costs:** All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm **4e.4 Business case documentation:** All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm	**C☐ P☐ M☐ N☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, met? **Rationale:**	**4** **C☐ P☐ M☐ N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐ N☐ A☐**

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.2 Point of Contact
John, Bott, MSSW, MBA, john.bott@ahrq.hhs.gov, 301-427-1317-

Measure Developer If different from Measure Steward
Co.3 Organization
Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850

Co.4 Point of Contact
John, Bott, MSSW, MBA, john.bott@ahrq.hhs.gov, 301-427-1317-

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
Co.6 Additional organizations that sponsored/participated in measure development UC Davis, Stanford University, Battelle Memorial Institute
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.** None
Ad.2 If adapted, provide name of original measure: None **Ad.3-5 If adapted, provide original specifications URL or attachment**
Measure Developer/Steward Updates and Ongoing Maintenance **Ad.6 Year the measure was first released:** 2001 **Ad.7 Month and Year of most recent revision:** 10, 2010 **Ad.8 What is your frequency for review/update of this measure?** Annual **Ad.9 When is the next scheduled review/update for this measure?** 05, 2011
Ad.10 Copyright statement/disclaimers: The AHRQ QI software is publicly available; no copyright disclaimers
Ad.11 -13 Additional Information web page URL or attachment:
Date of Submission (*MM/DD/YY*)**:** 02/01/2011

Page 14: [1] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

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

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

AND

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

AND

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

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

Page 14: [2] Comment [KP16] **Karen Pace** **10/5/2009 8:59:00 AM**

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

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

rationale/data support no risk adjustment.

Page 14: [3] Comment [k17] **Karen Pace** **10/5/2009 8:59:00 AM**

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

Page 14: [4] Comment [k19] **Karen Pace** **10/5/2009 8:59:00 AM**

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

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 #:** 0229 **NQF Project:** Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) 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 HF.
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 HF 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. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	**A** **Y**☒ **N**☐

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** Y☒ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	**C** Y☒ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☒ N☐
(for NQF staff use**)** **Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☒ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:** How is missing data handled? Gender is a risk factor in the model rather than a stratification.	
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. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	**Eval Rating**
(for NQF staff use) **Specific NPP goal**: Safety: 1)All hospitals will reduce preventable and premature hospital-level mortality rates to best-in-class. 2) All hospitals and their community partners will improve 30-day mortality rates following hospitalization for select conditions (acute myocardial infarction, heart failure, pneumonia) to best-in-class.	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Leading cause of morbidity/mortality, High resource use, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** HF incidence approaches 10 per 1000 population after 65 years of age (NHLBI 2007,), and is the most common discharge diagnosis among the elderly (Jessup and Brozena 2003); prevalence of HF in the U.S. is estimated at nearly 6 million. (Lloyd-Jones 2009), and is suspected as the leading cause of death in people over age 65. Many current hospital interventions are known to decrease the risk of death within 30 days of hospital admission. (Jha 2007) 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:** Jessup M, Brozena S. Medical progress: heart failure. N Engl J Med	**1a** C☐ P☐ M☐ N☐

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

2003;348:2007–18.

National Heart, Lung, and Blood Institute. Unpublished tabulation of NHANES, 1971-1975, 1976-1980, 1988-1994, 1999-2002, 2003-2006, and extrapolation to the U.S. population, 2007.

Lloyd-Jones D et al,American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010 Feb 23;121(7):e46-e215. Epub 2009 Dec 17

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.

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 HF. 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 HF RSMRs among hospitals. For the most recently reported three years of data (7/2006-6/2009) the mean hospital RSMR was 10.8% with a range of 6.6% to 18.2%. The 5th percentile was 8.4% and the 95th percentile was 13.4%. The interquartile range was 9.9% to 11.7%.

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 HF hospitalizations from July 1, 2006-June 30, 2009 and includes 1,096,751 hospitalizations from 4,743 hospitals. The index hospitalizations are those included in the measure and reported in the 2010 update to Hospital Compare.

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 slightly better performance for hospitals with higher proportions of African-American patients, but that the range of performance is similar to other hospitals. We divided hospitals into deciles based on the proportions 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 HF RSMR of 11.3% (range 6.4%- 19.4%) vs hospitals in the highest decile with >25% African American patients and a median HF RSMR of 10.5% (range 6.7%-15.1%).

Similar analyses were completed to evaluate hospital differences in performance based on socioeconomic status (SES) of their patients. These analyses suggest a slightly higher median HF RSMR at the hospitals in the lowest quartile based on the SES of their patients (as measured by median income of the patient's zip code). The lowest quartile hospitals' median RSMR is 11.3% compared to median RSMR of 10.8% for hospitals in highest quartile. However the range for the two groups is largely overlapping (6.7%-19.4% vs 6.9%-16.1%), respectively, demonstrating that substantial numbers of hospitals serving low SES patients perform well on

1b
C☐
P☐
M☐
N☐

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

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

the measure.

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 HF cases over the 3 year period, a total of 4,175 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 HF. 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*)**:**
Numerous studies have demonstrated that appropriate and timely treatment for HF patients can reduce the risk of mortality within 30 days of hospital admission. (Hunt 2009, Jha 2007) Additionally, trials of interventions which improve patient education upon discharge have been shown to improve survival for HF patients. (Mcalister 2001) 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.

References:

Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, Jessup M,Konstam MA, Mancini DM, Michl K, Oates JA, Rahko PS, Silver MA, Stevenson LW,Yancy CW; American College of Cardiology Foundation; American Heart Association.2009 Focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the International Society for Heart and Lung Transplantation. J Am Coll Cardiol. 2009 Apr 14;53(15):e1-e90.

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.

McAllister FA, Lawson FME, Teo KK, Armstrong PW: A systematic review of
randomized trials of disease management programs in heart failure. Am J Med 2001, 110:378-384

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.7 Summary of Controversy/Contradictory Evidence: Use of Hierarchical Generalized Linear Modeling
Hierarchical modeling 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 hierarchiacal 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 readmission. 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

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:
  - 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. ... [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.

purchasing programs.

Effect of Patient Preferences Regarding End of Life Care
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) We have added an exclusion for patients who are enrolled in hospice prior to, or on the day of, admission.

(2) We 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 we include markers of frailty within our risk-adjustment variables, including: protein-calorie malnutrition, dementia or senility, and hemiplegia, paraplegia, paralysis and functional disability.

(4) CMS will further consider clinical and measurement issues for patients for whom survival is not an objective.

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

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 *Importance to Measure and Report*?	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report***, met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Eval Rating**
2a. MEASURE SPECIFICATIONS	
S.1 **Do you have a web page where current detailed measure specifications can be obtained?** S.2 **If yes, provide web page URL:** **2a. Precisely Specified**	**2a-specs** **C☐**

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.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):**
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 HF.

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator***):**
Patients who die within 30 days of the index admission date.

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions***):**
Measure includes deaths from any cause within 30 days from admission date of index hospitalization.

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):**
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 and to define exclusions to 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 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 the 12 months prior to admission. Patients who are transferred from one acute care facility to another must have a principal discharge diagnosis of HF 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 HF 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 (*The time period in which cases are eligible for inclusion in the denominator***):**
This measure was developed with 12 months of data. Currently the measure is publicly-reported with three 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 HF defined by a principal discharge diagnosis of (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 the 12 months prior to admission.

ICD-9-CM codes that define the patient cohort:
402.01 Hypertensive heart disease, malignant, with heart failure
402.11 Hypertensive heart disease, benign, with heart failure
402.91 Hypertensive heart disease, unspecified, with heart failure
404.01 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.03 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease
404.11 Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.13 Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease

P☐
M☐
N☐

stage V or end stage renal disease
404.91 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.93 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease
428.0 Congestive heart failure, unspecified
428.1 Left heart failure
428.20 Unspecified systolic heart failure
428.21 Acute systolic heart failure
428.22 Chronic systolic heart failure
428.23 Acute on chronic systolic heart failure
428.30 Unspecified diastolic heart failure
428.31 Acute diastolic heart failure
428.32 Chronic diastolic heart failure
428.33 Acute on chronic diastolic heart failure
428.40 Unspecified combined systolic and diastolic heart failure
428.41 Acute combined systolic and diastolic heart failure
428.42 Chronic combined systolic and diastolic heart failure
428.43 Acute on chronic combined systolic and diastolic heart failure
428.9 Heart Failure, unspecified

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 HF diagnosis);
• 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 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-

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.

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
- Acute myocardial infarction
- Unstable angina
- Chronic atherosclerosis
- Cardio-respiratory failure and shock
- Valvular and rheumatic heart disease

Comorbidity
- Hypertension
- Stroke
- Renal failure
- 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 last year
- Major psych 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 a 1998 sample of 222,424 cases from 5,087 hospitals. 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:** URL N/A www.qualitynet.org

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 reliability of the model was tested by randomly selecting 50% of patients in the initial one-year cohort and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients. Furthermore, in each subsequent year of measure maintenance we have re-fit the model and compared the frequencies of comorbidities and model fit across 3 years.

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 discriminant ability, overall fit, and generated hospital RSMRs and corresponding interval estimates for the development sample.

2b.3 **Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
See results under "Risk-Adjustment Strategy" below.

2b
C☐
P☐
M☐
N☐

2c. Validity testing

2c.1 **Data/sample** *(description of data/sample and size)***:** Medical-record validation:
For the derivation of the chart-based model, we used cases identified through a Health Care Financing Administration (now CMS) quality initiative, which sampled admissions from fee-for-service Medicare beneficiaries for several clinical conditions, including HF. Cases were identified over a 6-month period within each state, plus the District of Columbia and Puerto Rico, during the period April 1, 1998 through October 31, 1999. Based on the principal discharge diagnosis, approximately 800 HF discharges per state were identified, and the corresponding medical records were abstracted by 2 clinical data abstraction centers. In states with fewer than 900 HF discharges, all cases were used. The abstractors first sorted the universe of eligible claims by age, race, sex, and hospital, then systematically sampled cases from a random starting point. Patients must have been enrolled in fee-for-service Medicare; Medicare managed care (Medicare + Choice) beneficiaries were excluded. CMS subsequently conducted a re-measurement using the same data collection methodology for 2000 and 2001 discharges, and the combined 1998-2001 data, including 73,832 patients, served as the national heart failure (NHF) dataset for development of the chart-based model.

2c.2 **Analytic Method** *(type of validity & rationale, method for testing)***:**
Medical-record validation: We developed a medical record measure to compare with the administrative measure. We developed a measure cohort with the medical record data using the inclusion/exclusion criteria

2c
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.

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.

and risk-adjustment strategy that was 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 matched sample included 46,700 patients. We compared the output of the two measures, that is the state performance results, in the same group of patients.

2c.3 Testing Results *(statistical results, assessment of adequacy in the context of norms for the test conducted)*:
The results of the medical-record validation were produced at the state level. The mortality medical record model had a c-statistic of 0.78. The correlation coefficient for the results of the administrative model compared to the medical-record model was very high, at 0.95.

2d. Exclusions Justified

2d.1 Summary of Evidence supporting exclusion(s):
Rationale for exclusions described in "Denominator Exclusions"

2d.2 Citations for Evidence:
See "Denominator Exclusions"

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

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

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

2d
C☐
P☐
M☐
N☐
NA☐

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: Prior years of data from Medicare Part A inpatient 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.

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 accounting for differences in hospital case-mix. (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 initial measure development, we tested the performance of the model developed in a random selected half of the 1998 hospitalizations for HF (representing 222,424 cases discharged from the 5,087

2e
C☐
P☐
M☐
N☐
NA☐

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;[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.

hospitals) against hospitalizations from the other half(representing 222,157 cases discharged from 5,088 hospitals). The performance was not substantively different in the validation sample (ROC area = 0.70) compared with the development sample (1998). The models appear well calibrated, with over-fitting indices of (-0.0035, 0.9928).

For the development cohort the results are summarized below:
Residuals lack of fit (<2, [-2,0),[0,2),[2+): 0.00, 87.85, 3.76, 8.39
Model Chi-square [# of covariates]: 11,521 [24]
Predictive ability (lowest decile %, highest decile %): 3.0%-28.5%
Area under ROC curve: .71

For the validation cohort the results are summarized below:
Residuals lack of fit (<2, [-2,0),[0,2),[2+): 0.00, 87.76, 3.83, 8.41
Model Chi-square [# of covariates]: 11444 [24]
Predictive ability (lowest decile %, highest decile %): 2.8%- 29.0%
Area under ROC curve: .70

In subsequent years, during annual measure maintenance 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 each subsequent year since 2005 and the and the parameters have remained consistent. For the 2005-2007 and 2006-2008 calendar year datasets, we reported each individual year results as well as the 3-year combined results. Model performance was stable over all time periods.

References:

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/

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

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

2f. Identification of Meaningful Differences in Performance

2f.1 **Data/sample from Testing or Current Use** *(description of data/sample and size)***:** This data below is based on RSMRs calculated for HF hospitalizations from July 1, 2006- June 30, 2009 and includes 1,096,751 hospitalizations from 4,743 hospitals. The index hospitalizations are those included in the measure and reported in the 2010 update to Hospital Compare.

2f.2 **Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)***:**
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 HF cases in the three-year period.

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

2f
C☐
P☐
M☐
N☐

Comment [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.

Recent analyses show substantial variation in HF RSMRs among hospitals. For the most recently reported three years of data (7/2006-6/2009) the mean hospital RSMR was 10.8% with a range of 6.6% to 18.2%. The 5th percentile was 8.4% and the 95th percentile was 13.4%. The interquartile range was 9.9% to 11.7%.

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.

2g. Comparability of Multiple Data Sources/Methods

2g.1 **Data/sample** *(description of data/sample and size)*: No current comparable data source was available that has complete data for a nationally representative sample.

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

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

2g
C☐ P☐ M☐ N☐ NA☐

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

2h. Disparities in Care

2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: N/A – Measure is not stratified

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 indicate little hospital-level disparities. The analyses performed by CMS (described in section 1b) demonstrate that hospitals have similar and overlapping performance on the measure regardless of the proportion of patients of low socioeconomic status or of African-American race. Importantly, the analyses show that hospitals with high proportions of low socioeconomic status patients or high proportions of African-American patients are able to perform well on the measure. For this reason CMS does not plan to stratify the measure.

2h
C☐ P☐ M☐ N☐ NA☐

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

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

2

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

2
C☐ P☐ M☐ N☐

3. USABILITY

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

Eval Rating

3a. Meaningful, Understandable, and Useful Information

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

3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*):
The measure has been publicly reported on Hospital Compare since June 2007. Used in CMS' Hospital Inpatient Quality Reporting Program (Formerly RHQDAPU). The measure is reported on Hospital Compare, www.hospitalcompare.hhs.gov.

3a
C☐ P☐ M☐ N☐

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

3a.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***):** **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)***:** 3a.5 **Methods** *(e.g., focus group, survey, QI project)***:** This measure was NQF endorsed in 2007. 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.6 **Results** *(qualitative and/or quantitative results and conclusions)***:**	
3b/3c. Relation to other NQF-endorsed measures 3b.1 **NQF # and Title of similar or related measures:** NQF # 0230 – Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization; NQF # 0468 - Pneumonia (PN) 30-Day Mortality Rate	
(for NQF staff use) **Notes on similar/related endorsed or submitted measures:** Related to #0358 Heart failure inpatient mortality (AHRQ)	
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?** Yes, the risk-adjustment strategy is similar.	**3b** C☐ P☐ M☐ N☐ NA ☐
3c. Distinctive or Additive Value 3c.1 **Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** The measure looks at a different condition, HF, than the AMI and pneumonia measures listed in 3b.1. 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:** NQF #0358 Congestive Heart Failure Mortality (IQI 16). 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 a diverse group of hospitals.	**3c** C☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability***, met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	**Eval Rating**
4a. Data Generated as a Byproduct of Care Processes	**4a** C☐

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.)

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☐ M☐ N☐

4b. Electronic Sources

4b.1 Are all the data elements available electronically? (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*)
Yes

4b.2 If not, specify the near-term path to achieve electronic capture by most providers.

4b C☐ P☐ M☐ N☐

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.

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☐ P☐ M☐ N☐ NA☐

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.

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.
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 HF 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.

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 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.

4d C☐ P☐ M☐ N☐

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

4e. Data Collection Strategy/Implementation

4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:
N/A

4e.2 Costs to implement the measure (*costs of data collection, fees associated with proprietary measures*)**:**
The measure is developed using administrative claims data and does not necessitate any additional cost/burden on hospitals.

4e.3 Evidence for costs:
N/A

4e C☐ P☐ M☐ N☐

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

4e.4 Business case documentation: N/A	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Feasibility*?	**4**
Steering Committee: Overall, to what extent was the criterion, *Feasibility*, met? **Rationale:**	**4** **C☐** **P☐** **M☐** **N☐**
RECOMMENDATION	
(for NQF staff use) **Check if measure is untested and only eligible for time-limited endorsement.**	**Time-limited ☐**
Steering Committee: Do you recommend for endorsement? **Comments:**	**Y☐** **N☐** **A☐**

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner)
Co.1 Organization
Centers for Medicare & Medicaid Services, 7500 Security Boulevard , Mail Stop S3-02-01, Baltimore, Maryland, 21244-9045

Co.2 Point of Contact
Lein, Han, PhD, Government Task Leader, lein.han@cms.hhs.gov, 410-786-0205-

Measure Developer If different from Measure Steward
Co.3 Organization
Yale New Haven Health Services Corporation (YNHHSC), 1 Church Street, Suite 200, New Haven, Connecticut, 06510

Co.4 Point of Contact
Susannah, Bernheim, MD, MHS, susannah.bernheim@yale.edu, 203-764-3271-

Co.5 Submitter If different from Measure Steward POC
Susannah, Bernheim, MD, MHS, susannah.bernheim@yale.edu, 203-764-7231-, Yale New Haven Health Services Corporation (YNHHSC)

Co.6 Additional organizations that sponsored/participated in measure development
MPR: Mathematica Policy Research; RTI-Research Triangle Institute

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.
The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org

Ad.2 If adapted, provide name of original measure: Heart Failure 30-day Mortality
Ad.3-5 If adapted, provide original specifications URL or attachment URL N/A www.qualitynet.org

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? 07, 2011

Ad.10 Copyright statement/disclaimers: N/A

Ad.11 -13 Additional Information web page URL or attachment: URL N/A www.qualitynet.org for Measure

Methodology report and Maintenance reports
Date of Submission (*MM/DD/YY*)**:** 12/14/2010

Page 4: [1] Comment [k4]	Karen Pace	10/5/2009 8:59:00 AM

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.

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 #:** 0330 **NQF Project:** Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION
De.1 **Measure Title:** Hospital 30-day, all-cause, risk-standardized readmission rate following heart failure hospitalization
De.2 **Brief description of measure**: The measure estimates a hospital 30-day risk-standardized readmission rate (RSRR), defined as readmission for any cause within 30 days after the date of discharge of the index admission for patients discharged from the hospital with a principal diagnosis of heart failure (HF).
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 mortality rate (RSMR) following an HF hospitalization.
De.4 **National Priority Partners Priority Area:** Patient and family engagement, Care coordination, Safety De.5 **IOM Quality Domain**: Effectiveness, Patient-centered, Efficiency, Safety De.6 **Consumer Care Need:** Getting better, 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**
A. The measure is in the public domain or an intellectual property (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* A.1 **Do you attest that the measure steward holds intellectual property rights to the measure and the right to use aspects of the measure owned by another entity (e.g., risk model, code set)?** Yes A.2 **Indicate if Proprietary Measure** (*as defined in measure steward agreement*)**:** A.3 **Measure Steward Agreement:** Government entity and in the public domain - no agreement necessary A.4 **Measure Steward Agreement attached:**	**A** **Y**☒ **N**☐

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** Y☒ N☐
C. The intended use of the measure includes both public reporting and quality improvement. ►**Purpose:** Public reporting, Internal quality improvement	**C** Y☒ N☐
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1**Testing:** Yes, fully developed and tested D.2 **Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** Yes	**D** Y☒ N☐
(for NQF staff use**) Have all conditions for consideration been met?** **Staff Notes to Steward** (*if submission returned*)**:**	**Met** Y☒ N☐
Staff Notes to Reviewers (*issues or questions regarding any criteria*)**:** Gender is a risk factor rather than stratified. How is missing data handled?	
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. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria) **1a. High Impact**	**Eval Rating**
(for NQF staff use) **Specific NPP goal**: Care Coordination: All healthcare organizations and their staff will work collaboratively with patients to reduce 30-day readmission rates.	
1a.1 **Demonstrated High Impact Aspect of Healthcare:** Affects large numbers, Leading cause of morbidity/mortality, High resource use, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 **Summary of Evidence of High Impact:** The Medicare Payment Advisory Commission (MedPAC) has called for hospital-specific public reporting of readmission rates, identifying HF as a priority condition (MedPAC, 2007). MedPAC finds that readmissions are common, costly, and often preventable. Based on 2005 Medicare data, MedPAC estimates that about 12.5% of Medicare HF admissions were followed by a readmission within 15 days, accounting for more than 90,000 admissions at a cost of $590 million. HF is the most common principal discharge diagnosis among Medicare beneficiaries and the third highest for hospital reimbursements in 2005 (CMS/OIS, 2006), and the leading cause of readmission among Medicare beneficiaries, with nearly half of HF patients expected to return to the hospital within 6 months of discharge. (Jencks 2009, Krumholz 1997) All-cause 30-day readmission rates per thousand patients discharged with HF increased by 11 percent between 1992 and 2001 (CMS/MPR/MQMS, 2003). HF readmission is a costly event and represents an undesirable outcome of care from the patient's perspective, and highly disparate HF readmission rates among hospitals suggest there is room for improvement. (MedPAC 2007, Bernheim 2010)	**1a** C☐ P☐ M☐ N☐

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

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).

1a.4 Citations for Evidence of High Impact: Report to the Congress: Promoting Greater Efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission, 2007.

Centers for Medicare & Medicaid Services, Office of Information Services (OIS). Available at http://www.cms.hhs.gov/MedicareFeeforSvcPartsAB/Downloads/SSDischarges0405.pdf , accessed October 21, 2006.

Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009 Apr 2;360(14):1418-28.

Krumholz HM, Parent EM, Tu N, Vaccarino V, Wang Y, Radford MJ, Hennen J. Readmission after hospitalization for congestive heart failure among Medicare beneficiaries. Arch Intern Med. 1997;157:99 –104.

Centers for Medicare & Medicaid Services, Mathematica Policy Research, Medicare Quality Monitoring System (MQMS) Report: Heart Failure, 1992-2001. Available at: http://www.mathematica-mpr.com/publications/PDFs/mqmsheart.pdf. accessed December 06, 2010

Report to the Congress: Promoting Greater Efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission, 2007.

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.

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 readmission rates following hospitalization for HF. 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 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:
Our recent analyses show substantial variation in RSRRs among hospitals. For the most recently reported three years of data (7/2006-6/2009) the mean hospital RSRR was 24.6% with a range of 17.3% to 32.4%. The 5th percentile was 21.4 and the 95th was 28.1. The interquartile range was 23.4% to 25.8%.

We have also demonstrated ongoing geographic variation in hospital RSRRs for HF.

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. 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 RSRRs calculated for HF hospitalizations from July 1, 2006-June 30, 2009 and includes 1,319,065 hospitalizations from 4759 hospitals. The index hospitalizations are

1b
C ☐
P ☐
M ☐
N ☐

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

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

those included in the measure and reported in the 2010 update to hospital compare.

1b.4 Summary of Data on disparities by population group:
CMS has performed analyses to evaluate disparities in performance by hospitals based on the proportion of patients that they serve that are African-American. These analyses show that though the median RSRR is slightly higher for hospitals with higher proportions of African-American patients compared with lower proportions,the range of performance is similar. 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 5% African-American patients and a median HF RSRR 24.3 (range 18.2-33.2) compared to hospitals in the highest decile with greater than 25% African American patients and a median HF RSRR 26.0 (range 20.6- 32.8).

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 HF RSRR at the hospitals in the lowest quartile based on the socioeconomic status of their patients (as measured by median income of the patient's zip code). The lowest quartile hospitals median RSRR is 25.0 compared to median RSRR of 24.4 for hospitals in highest quartile. However the range for the two groups is largely overlapping (19.0-33.2 vs 18.8-31.0) demonstrating that substantial numbers of hospitals serving low SES patients perform well on the measure.

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 HF cases over the 3 year period, a total of 4,260 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 readmission rates after hospitalization for HF. The goal is to directly affect patient outcomes by measuring risk-standardized rates of readmission.

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*)**:**
Studies have shown that interventions during and after a hospitalization can be effective in reducing readmission rates in geriatric populations (Benbassat and Taragin, 2000; Naylor et al., 1999; Coleman et al., 2006) and for elderly HF patients particularly (Phillips et al., 2004; Naylor et al., 2004; Koelling et al., 2005; Krumholz et al., 2002). Such interventions can be cost saving (Coleman et al., 2006; Krumholz et al., 2002; Naylor et al., 2004; Koelling et al., 2005; Phillips et al., 2004).

References:

Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Arch Intern Med. 2000 Apr 24;160(8):1074-81.

Naylor MD, Brooten D, Campbell R, et al. Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA. 1999 Feb 17;281(7):613-20.

Coleman EA, Parry C, Chalmers S, Min SJ. The care transitions intervention:results of a randomized controlled trial. Arch Intern Med. 2006 Sep 25;166(17):1822-8.

Phillips CO, Wright SM, Kern DE, et al. Comprehensive discharge planning with postdischarge support for older patients with congestive heart failure: a meta-analysis. JAMA. 2004 Mar 17;291(11):1358-67.

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:
  - 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.

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.

Naylor MD, Brooten DA, et al. Transitional care of older adults hospitalized with heart failure: a randomized, controlled trial. J Am Geriatr Soc. 2004 May;52(5):675-84.

Koelling TM, Johnson ML, Cody RJ, Aaronson KD. Discharge education improves clinical outcomes in patients with chronic heart failure. Circulation. 2005 Jan 18;111(2):179-85.

Krumholz HM, Amatruda J, Smith GL, et al. Randomized trial of an education and support intervention to prevent readmission of patients with heart failure. J Am Coll Cardiol. 2002 Jan 2;39(1):83-9.

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.7 Summary of Controversy/Contradictory Evidence: All-cause Readmission
This measure calculates a 30-day all-cause readmission rate. CMS measures all-cause readmission for rather than readmission due to certain conditions (e.g. heart failure readmissions) for a number of reasons. First, a narrow focus on specific causes of readmission may simply provide an incentive to shift patients away from those codes. Second, within the chain of events that lead to a patient being readmitted to the hospital there is often some aspect of care that could be improved, thereby reducing the risk of readmission. This is not to suggest that all readmissions are preventable, but the goal of the measure is to encourage broad approaches to quality improvement which will thereby lower all patients' risk of readmission. More narrowly defining readmission measures to those that are disease specific may incentivize a limited focus on improvements in care as opposed to thinking comprehensively about the patient's full medical and social needs at discharge. Factors which may influence readmission rates include medication reconciliation, patient education, follow-up care and communication between inpatient and outpatient providers. The goal is not to reduce the readmission rate to zero but to reduce overall readmission rates to what is achievable by the best hospitals.

Use of Hierarchical Generalized Linear Modeling
Hierarchical modeling 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 readmission rates. However, CMS frequently receives comments and questions about this approach, so we are concisely reiterating the rationale for and merits of using hierarchiacal 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 readmission. 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.

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

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

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

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 *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report*, **met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	**Eval Rating**
2a. MEASURE SPECIFICATIONS	
S.1 Do you have a web page where current detailed measure specifications can be obtained? **S.2 If yes, provide web page URL:** **2a. Precisely Specified**	
2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome***):** 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 readmission. We define this as readmission for any cause within 30 days from the date of discharge of the index HF admission. In addition, if a patient has one or more admissions within 30 days of discharge from the index admission, only one was counted as a readmission. **2a.2 Numerator Time Window (***The time period in which cases are eligible for inclusion in the numerator***):** Defined as readmission for any cause within 30 days from the date of discharge of the index admission. **2a.3 Numerator Details (***All information required to collect/calculate the numerator, including all codes, logic, and definitions***):** Measure includes readmissions to any acute care hospital for any cause within 30 days of the index HF admission discharge date.	
2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured***):** 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 and to define exclusions to the patient cohort. The cohort includes admissions for Medicare fee-for service (FFS) beneficiaries age 65 years or older discharged from the hospital with a principal 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 the 12 months prior to admission. **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 (***The time period in which cases are eligible for inclusion in the denominator***):** This measure was developed with 12 months of data. Currently the measure is publicly-reported with three	**2a-specs** **C☐** **P☐** **M☐** **N☐**

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

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 HF defined by a principal discharge diagnosis of the following (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 the 12 months prior to admission.

ICD-9-CM codes that define the patient cohort:
402.01 Hypertensive heart disease, malignant, with heart failure
402.11 Hypertensive heart disease, benign, with heart failure
402.91 Hypertensive heart disease, unspecified, with heart failure
404.01 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.03 Hypertensive heart and chronic kidney disease, malignant, with heart failure and with chronic kidney disease stage V or end stage renal disease
404.11 Hypertensive heart and chronic kidney disease, benign, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.13 Hypertensive heart and chronic kidney disease, benign, with heart failure and chronic kidney disease stage V or end stage renal disease
404.91 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and with chronic kidney disease stage I through stage IV, or unspecified
404.93 Hypertensive heart and chronic kidney disease, unspecified, with heart failure and chronic kidney disease stage V or end stage renal disease
428.0 Congestive heart failure, unspecified
428.1 Left heart failure
428.20 Unspecified systolic heart failure
428.21 Acute systolic heart failure
428.22 Chronic systolic heart failure
428.23 Acute on chronic systolic heart failure
428.30 Unspecified diastolic heart failure
428.31 Acute diastolic heart failure
428.32 Chronic diastolic heart failure
428.33 Acute on chronic diastolic heart failure
428.40 Unspecified combined systolic and diastolic heart failure
428.41 Acute combined systolic and diastolic heart failure
428.42 Chronic combined systolic and diastolic heart failure
428.43 Acute on chronic combined systolic and diastolic heart failure
428.9 Heart Failure, unspecified

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population***):** We excluded admissions for patients:

- with an in-hospital death (because they are not eligible for readmission);

- without at least 30 days post-discharge enrollment in Medicare FFS (because the 30-day readmission outcome cannot be assessed in this group);

- transferred to another acute care facility (When a patient is transferred from one acute care hospital to another, these multiple contiguous hospitalizations are considered one episode of care. Readmissions for transferred patients are attributed to the hospital that ultimately discharges the patient to a non-acute care setting.);

- discharged against medical advice (AMA) (because providers did not have the opportunity to deliver full care and prepare the patient for discharge);

- admitted with HF within 30 days of discharge from an index admission (Admissions within 30 days of discharge of an index admission will be considered readmissions. No admission is counted as a readmission

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.

and an index admission. The next eligible admission after the 30-day time period following an index admission will be considered another index admission.)

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 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 RSRR. 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 readmission 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 readmission 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 readmission, 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 readmission, 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. 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. We did not risk-adjust for CCs that were possible adverse events of care and that were only recorded in the index admission,

The final set of risk-adjustment variables are:

Demographic

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

Cardiovascular

- History of CABG
- Cardio-respiratory failure or shock
- Congestive heart failure
- Acute coronary syndrome
- Coronary atherosclerosis or angina
- Valvular or rheumatic heart disease
- Specified arrhythmias
- Other or unspecified heart disease

- Vascular or circulatory disease

Comorbidity

- Metastatic cancer or acute leukemia
- Cancer
- Diabetes or DM complications
- Protein-calorie malnutrition
- Disorders of fluid, electrolyte, acid-base
- Liver or biliary disease
- Peptic ulcer, hemorrhage, other specified gastrointestinal disorders
- Other gastrointestinal disorders
- Severe hematological disorders
- Iron deficiency or other anemias and blood disease
- Dementia or other specified brain disorders
- Drug/alcohol abuse/dependence/psychosis
- Major psychiatric disorders
- Depression
- Other psychiatric disorders
- Hemiplegia, paraplegia, paralysis, functional disability
- Stroke
- Chronic obstructive pulmonary disease
- Fibrosis of lung or other chronic lung disorders
- Asthma
- Pneumonia
- End stage renal disease or dialysis
- Renal failure
- Nephritis
- Other urinary tract disorders
- Decubitus ulcer or chronic skin ulcer

--

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://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1219069855841

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 RSRR is calculated as the ratio of the number of "predicted" to the number of "expected" readmissions, multiplied by the national unadjusted readmission rate. For each hospital, the "numerator" of the ratio is the number of readmissions within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the "denominator" is the number of readmissions 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 readmission or better quality and a higher ratio indicates higher-than-expected readmission 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 readmission, 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 readmissions (the denominator) is obtained by regressing the risk factors and a common intercept on the readmission 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 a 2004 sample of 283,919 cases from 4,669 hospitals. The models have been maintained and re-evaluated each year since public reporting of the measures began in 2009.

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: URL N/A
http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1219069855841

2a.29-31 Data dictionary/code table web page URL or attachment: URL N/A
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 reliability of the model was tested by randomly selecting 50% of patients in the initial one-year cohort and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients. Furthermore, in each subsequent year of measure maintenance we have re-fit the model and compared the frequencies of comorbidities and model fit across 3 years. **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 discriminant ability, overall fit, and generated hospital-level RSRRs and corresponding interval estimates for the development sample. **2b.3 Testing Results** *(reliability statistics, assessment of adequacy in the context of norms for the test conducted)*: See results under "Risk-Adjustment Strategy" below.	**2b** C☐ P☐ M☐ N☐
2c. Validity testing **2c.1 Data/sample** *(description of data/sample and size)*: Medical-record validation: For the derivation of the chart-based model, we used cases identified through a Health Care Financing Administration (now CMS) quality initiative, which sampled admissions from fee-for-service Medicare beneficiaries for several clinical conditions, including HF. Cases were identified over between April 1998 and March 1999 or between July 2000 and June 2001. Based on the principal discharge diagnosis, approximately 800 HF discharges per state were identified, and the corresponding medical records were abstracted by data central data abstraction center. In states with fewer than 800 HF discharges, all cases were used. The abstractors first grouped the claims by state, then sorted the universe of eligible claims by age, race, sex, and treating hospital, and then systematically sampled cases from a random starting point. Patients must have been enrolled in fee-for-service Medicare, resulting in a dataset of 78,882 records. **2c.2 Analytic Method** *(type of validity & rationale, method for testing)*: Medical-record validation: We developed a medical record measure to compare with the administrative measure. We defined a measure cohort with the medical record data using the inclusion/exclusion criteria that was 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 matched sample included 64,329 patients. We compared the output of the two measures, that is, the state performance results, in the same group of patients. **2c.3 Testing Results** *(statistical results, assessment of adequacy in the context of norms for the test conducted)*: The results of the medical-record validation were produced at the state level. The mortality medical record model had a c-statistic of 0.58 as compared to 0.60 for the claims based measure. The correlation coefficient for the results of the administrative model compared to the medical-record model was very high, at 0.97 showing excellent consistency of the two models. Reference: Keenan PS, Normand SL, Lin Z, Drye EE, Bhat KR, Ross JS, Schuur JD, Stauffer BD, Bernheim SM, Epstein AJ, Wang Y, Herrin J, Chen J, Federer JJ, Mattera JA,Wang Y, Krumholz HM. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. Circ Cardiovasc Qual Outcomes. 2008 Sep;1(1):29-37.	**2c** C☐ P☐ M☐ N☐
2d. Exclusions Justified **2d.1 Summary of Evidence supporting exclusion(s):** Rationale for exclusions described in "Denominator Exclusions" **2d.2 Citations for Evidence:**	**2d** C☐ P☐ M☐ N☐ NA

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 ... [1]

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.

See "Denominator Exclusions" ☐

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

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

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

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: Prior years of data from Medicare Part A inpatient and outpatient data and Part B outpatient data are used to identify variables for risk-adjustment.

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 RSRRs accounting for differences in hospital case-mix. (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 initial measure development, we tested the performance of the model developed in a random selected half of the 2004 hospitalizations for HF (representing 283,919 cases discharged from the 4,669 hospitals) against hospitalizations from the other half(representing 283,528 cases discharged from 4,680 hospitals). The performance was not substantively different in the validation sample (ROC area = 0.60) compared with the development sample (2004). The models appear well calibrated, with the over-fitting indices of (0,089, 1.05).

For the development cohort the results are summarized below:
Residuals lack of fit {<-2, [-2,0),[0,2),[2+}: {0,76.40,17.62,5.98}
Model Chi-Sq [# of covariates]: 6,462 [37]
Predictive ability (lowest decile %, highest decile %): (15%,37%)
Area under ROC curve: .60

For the validation cohort the results are summarized below:
Residuals lack of fit {<-2, [-2,0),[0,2),[2+}: {0,76.29,17.83,5.88}
Model Chi-Sq [# of covariates]: 6,632 [37]
Predictive ability (lowest decile%, highest decile %): (15%,37%)
Area under ROC curve: .60

In subsequent years, during annual measure maintenance we looked at the distributions of comorbid conditions, hospital volume, crude rates, hospital RSRR, risk-adjusted odds ratios and 95% confidence intervals, and between-hospital variance over each subsequent year since 2006 and the and the parameters have remained consistent. For example, for the 2006-2008 calendar year dataset, we reported each

2e
C☐
P☐
M☐
N☐
NA☐

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.

individual year results as well as the 3-year combined results. Model performance was stable over all time periods.

References:

Krumholz HM, Normand S-LT, Keenan PS, et al. 2008. Hospital 30-Day Heart Failure Readmission Measure: Methodology. Report prepared for the Centers for Medicare & Medicaid Services.

Bernheim SM, Lin Z, Bhat KR, et al. 2010. 2010 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-Day Risk-Standardized Readmission Measures. Report prepared for the Centers for Medicare & Medicaid Services.

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

2f. Identification of Meaningful Differences in Performance

2f.1 **Data/sample from Testing or Current Use** *(description of data/sample and size)*: This data below is based on RSRRs calculated for HF hospitalizations from July 1, 2006- June 30, 2009 and includes 1,319,065 hospitalizations from 4,759 hospitals. The index hospitalizations are those included in the measure and reported in the 2010 update to Hospital Compare.

2f.2 **Methods to identify statistically significant and practically/meaningfully differences in performance** *(type of analysis & rationale)*:
For each RSRR, CMS characterizes the uncertainty associated with the RSRR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSRR's interval estimate does not include the national crude readmission rate (is lower or higher than the rate), then CMS is confident that the hospital's RSRR 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 HF cases in the three-year period.

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)*:
Recent analyses show substantial variation in hospital RSRR's for HF:
Mean: 24.6%
Minimum: 17.3%
5th percentile: 21.4%
25th percentile: 23.4%
Median: 24.5%
75th percentile: 25.8%
95th percentile: 28.1%
Maximum: 32.4%

2f
C☐
P☐
M☐
N☐

2g. Comparability of Multiple Data Sources/Methods

2g.1 **Data/sample** *(description of data/sample and size)*: No current comparable data source was available that has complete data for a nationally representative sample.

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

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

2g
C☐
P☐
M☐
N☐
NA☐

2h. Disparities in Care

2h
C☐

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.

2h.**1 If measure is stratified, provide stratified results** *(scores by stratified categories/cohorts)*: N/A – Measure is not stratified 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 indicate that performance on RSRR's is similar, with wide overlap across hospitals with different proportions of African American or low SES patients. Hospitals with higher proportions of African-American or low-SES patients can perform at least as well on our measures. The analyses performed by CMS (described in section 1b) demonstrate that hospitals have largely overlapping performance on the measure regardless of the proportion of patients of low socioeconomic status or of African-American race. Importantly, the analyses show that hospitals with high proportions of low socioeconomic status patients or high proportions of African-American patients are able to perform well on the measure. For this reason CMS does not plan to stratify the measure.	P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties?*	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* C☐ P☐ M☐ N☐
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)	**Eval Rating**
3a. Meaningful, Understandable, and Useful Information 3a.1 **Current Use:** In use 3a.2 **Use in a public reporting initiative (disclosure of performance results to the public at large)** (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:** Used in CMS' Hospital Inpatient Quality Reporting Program (Formerly RHQDAPU) 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***):** **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)***:** 3a.5 **Methods** *(e.g., focus group, survey, QI project)***:** This measure was originally NQF endorsed in 2008. Prior to public reporting in 2009, CMS conducted a dry run in 2008 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.6 **Results** *(qualitative and/or quantitative results and conclusions)***:**	**3a** C☐ P☐ M☐ N☐
3b/3c. Relation to other NQF-endorsed measures 3b.1 **NQF # and Title of similar or related measures:** NQF # 0505- Thirty-day all-cause risk standardized readmission rate following acute myocardial infarction	

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.

(AMI) hospitalization. NQF # 0506- Thirty-day all-cause risk standardized readmission rate following pneumonia hospitalization.	
(for NQF staff use) **Notes on similar/related endorsed or submitted measures:**	
3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): **3b.2 Are the measure specifications harmonized? If not, why?** Yes, they used a similar risk adjustment strategy.	**3b** C☐ P☐ M☐ N☐ NA ☐
3c. Distinctive or Additive Value **3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:** This measure looks at a different condition for the readmission outcome, HF, from the two other related readmission measures for AMI and pneumonia. **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☐ P☐ M☐ N☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	**Eval Rating**
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	**4b** C☐ P☐ M☐ N☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	**4c** C☐ P☐ M☐ N☐ NA ☐
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.**	**4d** C☐ P☐ M☐

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.

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

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 HF and that the model produced estimates of RSRRs 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.

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 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.

N ☐

4e. Data Collection Strategy/Implementation

4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues:
N/A

4e.2 Costs to implement the measure (*costs of data collection, fees associated with proprietary measures*)**:**
The measure is developed using administrative claims data and does not necessitate any additional cost/burden on hospitals.

4e.3 Evidence for costs:
N/A

4e.4 Business case documentation: N/A

4e
C ☐
P ☐
M ☐
N ☐

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

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

4

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

4
C ☐
P ☐
M ☐
N ☐

RECOMMENDATION

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

Time-limited ☐

Steering Committee: Do you recommend for endorsement?
Comments:

Y ☐
N ☐
A ☐

CONTACT INFORMATION

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

21244-9045

Co.2 **Point of Contact**
Lein, Han, PhD, Government Task Leader, Lein.han@cms.hhs.gov, 410-786-0205-

Measure Developer If different from Measure Steward
Co.3 **Organization**
Yale New Haven Health Services Corporation YNHHSC, 1 Church St., Suite 200, New Haven, Connecticut, 06510

Co.4 **Point of Contact**
Susannah, Bernheim, MD, MHS, susannah.bernheim@yale.edu, 203-764-7231-

Co.5 **Submitter If different from Measure Steward POC**
Susannah, Bernheim, MD, MHS, susannah.bernheim@yale.edu, 203-764-7231-, Centers for Medicare & Medicaid Services

Co.6 **Additional organizations that sponsored/participated in measure development**
MPR- Mathematica Policy Research, RTI- Research Triangle Institute

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.**
The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org

Ad.2 **If adapted, provide name of original measure:** N/A
Ad.3-5 **If adapted, provide original specifications URL or attachment** URL N/A www.qualitynet.org

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 **Year the measure was first released:** 2008
Ad.7 **Month and Year of most recent revision:** 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?** 07, 2011

Ad.10 **Copyright statement/disclaimers:** N/A

Ad.11 -13 **Additional Information web page URL or attachment:** URL N/A www.qualitynet.org for Measure Methodology report and Maintenance reports

Date of Submission (*MM/DD/YY*)**:** 12/14/2010

Page 11: [1] Comment [KP14] **Karen Pace** **10/5/2009 8:59:00 AM**

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

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

AND

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

AND

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

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

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 #:** 962 **NQF Project:**
De.1 Title of Measure: Composite Measure of Hospital Quality for Heart Failure (HF)
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*): A composite measure of in-hospital process- and outcome-of-care for Heart Failure (HF) patients.
De.3 Type of Measure: ☐ Composite with component measures combined at patient-level (e.g., all-or-none) ☒ 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** ☐ patient and family engagement ☐ population health ☐ safety ☒ care coordination ☐ palliative and end of life care ☐ overuse

De.5 IOM Quality Domain ☒ effectiveness ☐ efficiency ☐ equity ☐ patient-centered ☐ safety ☒ timeliness

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**
A. The measure is in the public domain or an intellectual property agreement (measure steward agreement) is signed. *Public domain only applies to governmental organizations. All non-government organizations must sign a measure steward agreement even if measures are made publicly and freely available.* **A.1 Do you attest that the measure steward holds intellectual property rights to the measure and the right to use any aspects of the measure owned by another entity (e.g., component measures, risk model, code set)?** ☒ Yes **A.2 Measure Steward Agreement** ☐ Signed and Submitted **OR** ☒ Government entity–public domain (If measure steward agreement not signed for non-government entities, do not submit) **A.3 Please check if either of the following apply:** ☐ Proprietary Measure ☐ Proprietary Complex Measure w/fees	**A** Y☐ N☐
B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. **B.1** ☒ Yes (If no, do not submit)	**B** Y☐ N☐
C. The intended use of the measure includes both public reporting and quality improvement. **C.1 Purpose:** ☒ Public reporting ☒ Internal quality improvement **C.2** ☒ Accountability ☐ Accreditation ☐ Payment incentive ☐ Other, describe: (If not intended for both public reporting and quality improvement, do not submit)	**C** Y☐ N☐
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. **D.1 Testing:** ☒ Fully developed and tested (If composite measure not tested, do not submit) **D.2 Have NQF-endorsed measures been reviewed to identify if there are similar or related measures?** ☒ 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?** ☒ Yes (If no, do not submit)	**D** Y☐ N☐
De.7 If component measures of the composite **are aggregate-level measures, all** must be either NQF-endorsed or submitted for consideration for NQF endorsement (*check one*) ☒ All component measures are NQF-endorsed measures ☐ Some or all component measures are not NQF-endorsed and have been submitted using the online measure submission tool (If not, do not submit)	Y☐ N☐
(for NQF staff use) **Have all conditions for consideration been met?** Staff Notes to Steward (if submission returned):	**Met** Y☐ N☐
Staff Notes to Reviewers (issues or questions regarding any criteria):	
Staff Reviewer Name(s):	

TAP/Workgroup Reviewer Name:	

Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (composite measure evaluation criteria)	**Eval**
(for NQF staff use) **Specific NPP goal**:	
1d. Purpose/objective of the Composite **1d.1 Describe the purpose/objective of the composite measure:** 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 various measures adopted for the program 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 Heart Failure (HF) 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. Specifically, this measure summarizes clinical process- and outcome-of-care indicators associated with the treatment of HF 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 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. 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 may 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 assess: "Overall, how well is our system serving patients with this condition?" 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. **1d.2 Describe the quality construct used in developing the composite:** The composite measure of quality of hospital care for HF 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 four individual process-of-care indicators and two outcome-of-care indicators meant to capture the quality of hospital care provided to patients with HF. NQF has endorsed all six indicators. The proposed composite combines these in the form of process- and outcome-of-care domain scores. CMS realizes that some HF indicators that appear *on Hospital Compare* and are included in the composite measure may later lose their endorsed status. Should that occur, we will reconfigure the composite and resubmit to NQF for endorsement at the next available opportunity.	**1d** **C**☐ **P**☐ **M**☐ **N**☐

However, CMS wishes the composite to include all HF indicators that are endorsed at the time of its submission.

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 focuses on diseases important to the Medicare population: Acute Myocardial Infarction (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 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, 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	• Include the same set of process and outcome measures as Hospital Compare
Produce differences in composite values that are clinically and statistically meaningful and reflect true differences in underlying quality	• Enter component indicators 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 indicators, adjust component indicators for reliability so that random variation does not drive small hospitals to extremes
Results available for a large number of hospitals	• Process indicators are available when the number of eligible discharges is five or more; outcome indicators are available when the number of eligible discharges is 25 or more
Focus more on measures relevant to more patients	• Construct process and outcome domains using weights based on national denominators
Method is scientifically acceptable and acceptable to stakeholders	• 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 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
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	
Method is relatively simple Hospitals can duplicate results	• Use equal weighting to combine process and outcome domains • Reliability weights are a function of a hospital's number of cases and national parameters

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, the HF composite is a formative summary of all HF indicators reported on Hospital Compare. Measures were adopted for the Hospital Inpatient Quality Reporting Program through a consensus process that deemed them to be indicators of well-coordinated high-quality care for HF. The measures that make up the composite include both process- and outcome-of-care indicators.

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 the HF composite measure, the process-of-care indicators

1e
C☐
P☐
M☐
N☐

evaluate whether a patient received:

- Discharge instructions [HF1]
- Evaluation of Left Ventricular Systolic (LVS) Function [HF2]
- ACE Inhibitor or ARB for Left Ventricular Systolic Dysfunction (LVSD) [HF3]
- Smoking Cessation advice/counseling [HF4]

These NQF-endorsed process-of-care indicators represent established best practices for HF care[1,2] and were adopted by CMS 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.

In addition to reflecting current clinical guidelines, studies have shown a clear relationship between execution of these practices and decreased mortality for HF patients[3-5], one of the two outcome-of-care indicators also included in the proposed HF composite measure. The two HF outcome-of-care component measures are: 1) 30-day risk-standardized mortality and 2) 30-day risk-standardized 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 HF patients at a given hospital experience, and therefore may be critical to understanding the quality of care received.[i]

The combination of these component indicators ultimately serves to deliver a single, useful, condition-specific summary of HF care for consumer use.

Citations

1. Hunt SA. ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure). J American College of Cardiology 2005; 46(6):e1-82.
2. Heart Failure Society of America. HFSA 2006 Comprehensive Heart Failure Practice Guideline. J Card Fail. 2006 Feb; 12(1):e1-2.
3. 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.
4. Garg R, Yusuf S. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. Collaborative Group on ACE Inhibitor Trials. JAMA. 1995 May 10; 273(18):1450-6. Erratum in: JAMA 1995 Aug 9; 274(6):462.
5. Heart Failure Society of America. HFSA 2010 Comprehensive Heart Failure Practice Guideline. J Card Fail. 2010 Jun; 16(6):e1-e2.

If the component measures are combined at the patient level, complete 1a, 1b, and 1c.

If the component measures are combined at the aggregate level, skip to criterion 2, *Scientific Acceptability of Measure Properties* (individual measures are either NQF-endorsed or submitted individually).

1a. High Impact

1a.1 Demonstrated high impact aspect of healthcare (*Select the most relevant*)

☐ affects large numbers ☐ frequently performed procedure ☐ leading cause of morbidity/mortality ☐ high resource use ☐ severity of illness ☐ patient/societal consequences of poor quality ☐ other, describe: **1a.2**

1a.3 Summary of Evidence of High Impact:

1a.4 Citations for Evidence of High Impact:

1a
H☐
M☐
L☐
N☐

[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.

1b. Opportunity for Improvement **1b.1 Briefly explain benefits (improvements in quality) envisioned by use of this measure:** **1b.2 Summary of data demonstrating performance gap** (*variation or overall poor performance across providers*)**:** **1b.3 Citations for data on performance gap:** **1b.4 Summary of Data on disparities by population group:** **1b.5 Citations for data on Disparities:**	**1b** **H☐** **M☐** **L☐** **N☐**
1c. Evidence-based **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.)* **1c.2 Type of Evidence** *(Check all that apply)* ☐ Cohort study ☐ Evidence-based guideline ☐ Expert opinion ☐ Meta-analysis ☐ Observational study ☐ Randomized controlled trial ☐ Systematic synthesis of research ☐ Other (*Please describe*): **1c.3** **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*)**:** **1c.5 Rating of strength/quality of evidence** (*also provide narrative description of the rating and by whom*) **1c.6 Method for rating evidence:** **1c.7 Summary of Controversy/Contradictory Evidence:** **1c.8 Citations for Evidence** (*other than guidelines*) **1c.9 Quote the Specific guideline recommendation** (*including guideline number and/or page number*) **1c.10 Clinical Practice Guideline Citation:** **1c.11 National Guideline Clearinghouse or other URL:** **1c.12 Rating of strength of recommendation** (*also provide narrative description of the rating and by whom*) **1c.13 Method for rating strength of recommendation** (*If different from USPSTF system, also describe rating and how it relates to USPSTF)*: **1c.14 Rationale for using this guideline over others:**	**1c** **H☐** **M☐** **L☐** **N☐**
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	**1**
Steering Committee: Was the threshold criterion, *Importance to Measure and Report*, **met?** **Rationale:**	**1** **Y☐** **N☐**
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (composite measure evaluation criteria)	**Eval**
2a. COMPOSITE MEASURE SPECIFICATIONS	
In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained? **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:	**2a-specs** **C☐** **P☐**

http://www.hospitalcompare.hhs.gov/

M☐
N☐

S.2 If yes, provide web page URL: http://www.hospitalcompare.hhs.gov/

2a. Precisely Specified

2a.0.1 Components of the Composite (*List the components, i.e., domains/sub-composites, individual measures. If component measures are NQF-endorsed, include NQF measure number*; if not NQF-endorsed, provide date of submission to NQF)

HOSPITAL PROCESS-OF-CARE INDICATORS		
NQF #0136	Percent of HF Patients that Received Discharge Instructions	Endorsed May 9, 2007
NQF #0135	Percent of HF Patients with Evaluation of LVS Function	Endorsed May 9, 2007
NQF #0162	Percent of HF Patients Given ACE Inhibitor or ARB for LVSD	Endorsed May 9, 2007
NQF #0027	Percent of HF Patients Given Smoking Cessation Advice/Counseling	Endorsed May 1, 2006
HOSPITAL OUTCOME-OF-CARE INDICATORS		
NQF #0229	HF 30-day Risk-Standardized Mortality	Endorsed May 9, 2007
NQF #0330	HF 30-day Risk-Standardized Readmission	Endorsed May 15, 2008

If the composite measure cannot be specified with a numerator and denominator, please consult with NQF staff.

If the component measures are combined at the aggregate level, do not include the individual measure specifications below.

2a.1 Composite Numerator Statement:

For the process-of-care domain, the numerator is equal to the weighted sum of four terms. Each term is equal to the ratio of the hospital's raw performance rate to the national performance rate for the indicator. The weight is equal to the total number of observations, that is, the number of patients 'at risk' for the indicator.

For the outcome-of-care domain, the numerator is equal to the weighted sum of two terms. Each term is equal to the ratio of the hospital's risk-standardized performance rate to the national performance rate for the indicator. The weight is equal to the total number of eligible discharges for the indicator.

2a.2 Numerator Time Window: July 2006 - June 2009

2a.3 Numerator Details: Successes in the following heart failure process-of-care and outcome-of-care indicators:

HOSPITAL PROCESS-OF-CARE INDICATORS
1. Percent of HF Patients that Received Discharge Instructions (NQF #0136)
2. Percent of HF Patients with Evaluation of LVS Function (NQF #0135)
3. Percent of HF Patients Given ACE Inhibitor or ARB for LVSD (NQF #0162)
4. Percent of HF Patients Given Smoking Cessation Advice/Counseling (NQF #0027)

HOSPITAL OUTCOME-OF-CARE INDICATORS
1. HF 30-day Risk-Standardized Mortality (NQF #0229)
2. HF 30-day Risk-Standardized Readmission (NQF #0330)

2a.4 Composite Denominator Statement:

For the process-of-care domain, the denominator is equal to the total number of observations for all HF process indicators. It is thus equal to the number of patients 'at risk for the four process indicators.

For the outcome-of-care domain, the denominator is equal to the total number of observations for all HF outcome indicators. It is thus equal to the number of eligible discharges for the two outcome indicators.

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 heart failure quality reports. Counts of outcome-of-care opportunities are based on claims data.

2a.9 Composite Denominator Exclusions:

The following two criteria were applied as exclusion restrictions:

1. Hospitals with less than five eligible patient cases for the process-of-care indicators and less than 25 eligible discharges for the outcome-of-care indicators.
2. Hospitals that were missing rates for one or more process-of-care and/or outcome-of-care indicators.

2a.10 Denominator Exclusion Details: See above (2a.9)

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions***):**
The composite measure was not stratified.

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 as the simple average of process and outcome domain scores. The outcome domain score was computed as the denominator-weighted sum of the ratio of actual to expected values of the two outcome indicators. The process domain score was computed as the ratio of actual to expected values of the four process indicators. All indicators are publically reported by the CMS on *Hospital Compare* and are NQF endorsed. The method of scoring is described in detail below. Additional documentation is available in Section 2 of the attached appendix (Appendix A).

CMS began publically reporting 30-day risk-standardized mortality and readmission rates, used in construction of the outcome domains score, in June 2007 and in July 2009, respectively. In computing the indicators, Yale researchers employed a method known as 'shrinkage' or 'Bayesian smoothing' to increase the overall accuracy of the indicators. The method is well-known and widely accepted in the statistical literature (Morris 1983; Carlin and Louis 2000). In order to bring the process-of-care indicators into conformity with outcome indicators in constructing the composite, reliability weights to each individual process-of-care indicator. Each indicator is thus computed as a weighted average of the hospital's own value for the indicator and the national mean for that indicator. Each indicator was then standardized by dividing by the national mean of the indicator. Outcome-indicators were also was standardized by dividing by the national mean of the indicator.

In order to remain consistent with the approach used for AHRQ measures, CMS used denominator weighting in constructing the process- and outcome-of-care domains. Denominator weighting places greater weight on indicators that apply to higher numbers of patients nationally, 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-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.

Lastly, the overall composite score was calculated as a simple average of the two domain scores. In Table 2a.42.1, we provide a summary of the composite measure. Since the process- and outcome-of-care indicators are standardized by the national rate of each of the indicators, hospitals with a composite score of >1 have a performance score that is greater than the national rate and hospitals with a composite score of <1 have a performance score that is less than the national rate. However, it should be noted that the differences in performance from the national rate should be interpreted with caution since it may not be statistically significant.

Therefore, our method of discrimination of performance is described in greater detail in Section 2a.22.

Table 2a.42.1: Summary of Composite and Composite Domains

Domain	Description	Interpretation
Process-of-Care	Denominator weighted average of standardized (by the national mean) probabilities that patients with HF will receive the appropriate care.	Hospitals with a process-of-care domain score >1 have a score that is better than average. Hospitals with a process-of-care domain score <1 have a score that is worse than average. Hospitals with a process-of-care domain score =1 have a score that is equal to the average.
Outcome-of-Care	Denominator weighted average of standardized (by the national mean) probabilities of survival and of avoidance of readmission after 30 days of admission to a hospital with HF.	Hospitals with an outcome-of-care domain score >1 have a score that is better than average. Hospitals with an outcome-of-care domain score <1 have a score that is worse than average. Hospitals with an outcome-of-care domain score =1 have a score that is equal to the average.
Overall Composite	Simple average of the process- and outcome-of-care domain scores.	Hospitals with a composite score >1 have a score that is better than average. Hospitals with a composite score <1 have a score that is worse than average. Hospitals with a composite score =1 have a score that is equal to the average.

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

Composite scores for a hospital were calculated if:

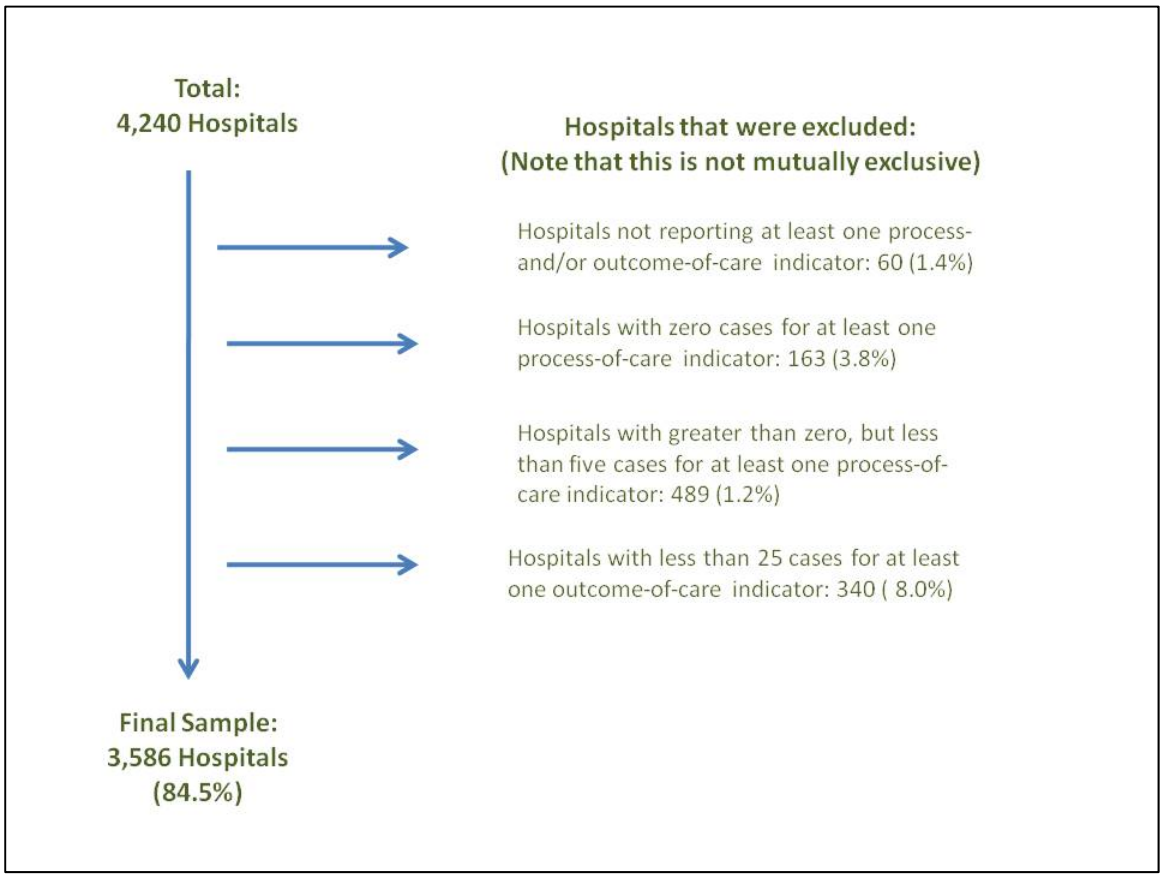
1. The hospitals reported rates for all four process and all two outcome-of-care indicators
2. Each process-of-care indicator had at least five cases and each outcome-of-care indicator had at least 25 cases.

Composite scores were not estimated for hospitals that did not satisfy the above two criteria. Table 2a.44.1 summarizes the time at which the data was released on *Hospital Compare* and the collection period of the quality indicators. In addition, Figure 2a.44.1 shows how the final sample of hospitals was derived.

Table 2a.44.1: Data Release and Collection Period

Data Release	Indicators Used	Time Period
March 2008	Process-of-care	July 2006-June 2007
March 2009	Process-of-care	July 2007-June 2008
March 2010	Process-of-care	July 2008-June 2009
June 2010	Outcome-of-care	July 2006-June 2009

Figure 2a.44.1: Sample of Hospitals



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- and outcome-of-care domains. Denominator weighting places relatively more weight on measures that apply to relatively more patients nationally, 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. Technical documentation on the scoring approach is provided in Section 2.1 of Appendix A, attached)

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

Table 2a.21.1: Steps to Construct the Composite Score

Key Steps	Process-of-Care Domain	Outcome-of-Care Domain	Overall Composite
Step 1a Exclude hospitals that do not meet the minimum case size requirement Step 1b Exclude hospitals missing one or more indicators	Exclude hospitals if there are less than five cases for any of the four process-of-care indicators. Exclude hospitals missing one or more process-of-care indicators.	Exclude hospitals if there are less than 25 cases for any of the two outcome-of-care indicators. Exclude hospitals missing one or more outcome-of-care indicators.	N/A
Step 2 Weight the indicators by a reliability weight	The value of each process-of-care indicators is set to a weighted average of the hospital's own rate and the national rate. *Example* *Suppose the performance rate for the "percentage of HF patients with evaluation of LVS function" at Heartcare Regional Hospital is 80% and the national rate for this indicator is 77%. Also, suppose that the hospital's weight is 0.8. Then the hospital's reliability-weight adjusted rates is:* $0.8(80\%) + (1 - 0.8)(77\%) = 79.4\%$	N/A	N/A
Step 3 Standardize the indicators by dividing by the national mean of each indicator	The value of each (reliability weight adjusted) process-of-care indicator is divided by the national rate. *Example* *Given the previous example in Step 2, if Heartcare Regional Hospital's reliability-weight adjusted rates is79.4% and the national reliability-rate adjusted rate is 81%, then the standardized indicator is:* $\frac{79.4}{81.0} = 0.98$	The value of each outcome-of-care indicator is divided by the national rate. *Example* *If the 30-day risk-adjusted survival rate at Heartcare Regional Hospital is 91% and the national survival rate is 88.8%, then the standardized indicator is:* $\frac{91.0}{88.8} = 1.02$	N/A

Table 2a.21.1: Steps to Construct the Composite Score (cont.)

Key Steps	Process-of-Care Domain	Outcome-of-Care Domain	Overall Composite
Step 4 Combine the indicators using a denominator weighted average	Take a denominator-weighted average of the standardized process-of-care indicators. *Example* *Suppose the standardized rates and the national number of cases for the four process-of-care for Heartcare Hospital respectively are*:* *HF1: 1.10 (N=4000)* *HF2: 0.98 (N=5000)* *HF3: 1.32 (N=3500)* *HF4: 0.95 (N=4000)* *Then the process-of-care domain score is:* $\frac{4000}{16500}(1.10)+\frac{5000}{16500}(0.98)+\frac{3500}{16500}(1.32)+\frac{4000}{16500}(0.95)=1.06$	Take a denominator-weighted average of the standardized outcome-of-care indicators. *Example* *Suppose the standardized rates and the national number of cases for the two outcome-of-care for Heartcare Hospital respectively are**:* *Survival: 1.02 (N=4500)* *Readmission: 0.95 (N=4500)* *Then the outcome-of-care domain score is:* $\frac{4500}{9000}(1.02)+\frac{4500}{9000}(0.95)=0.99$	N/A
Step 5 Combine the process- and outcome-of-care domains to create a composite score	N/A	N/A	Take a simple average of the process- and outcome-of-care domain scores *Example* *Given the standardized rates for the process- and outcome-of-care domains, the composite score is:* $\frac{1}{2}(1.06)+\frac{1}{2}(0.99)=1.03$

Notes:
* HF1: Percent of HF Patients that Received Discharge Instructions; HF2: Percent of HF Patients with Evaluation of LVS Function; HF3: Percent of HF Patients Given ACE Inhibitor or ARB for LVSD; HF4: Percent of HF Patients Given Smoking Cessation Advice/Counseling.
** Survival: 30-day risk-adjusted survival rate; Readmission: 30-day risk-adjusted lack of readmission.

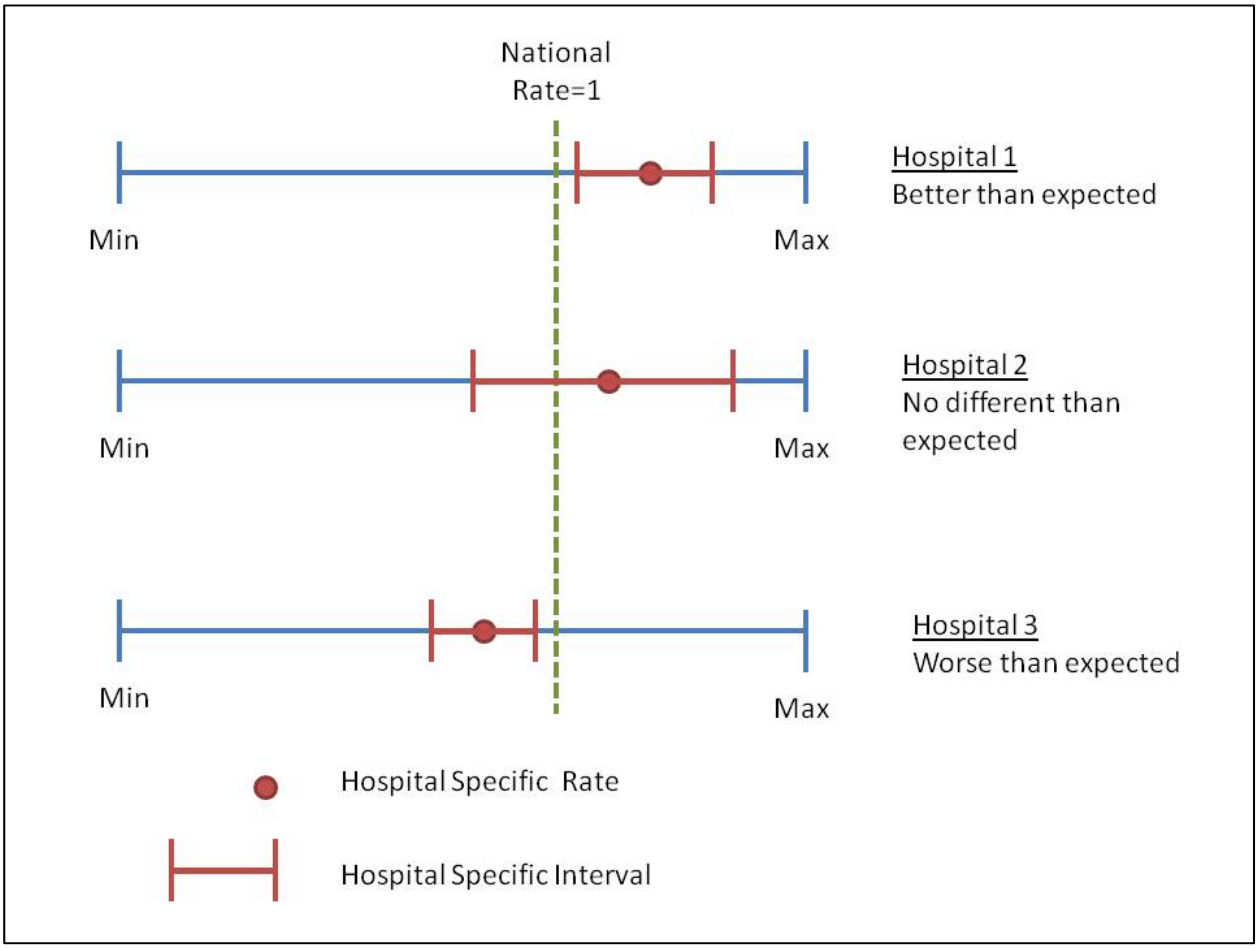
2a.22 Describe the method for discriminating performance *(e.g., significance testing)*:

To examine meaningful differences in composite measures among hospitals, we compared hospitals' confidence interval estimates with the overall mean and assigned hospitals into one of three performance categories: "better-than-expected' hospitals, if the interval estimate is entirely above the mean; 'no-different-than-expected' hospitals, if the interval estimate includes the mean; and 'worse-than-expected' 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.

We derived the standard error for each hospital and estimated 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. Figure 2a.22.1 shows how the hospitals are categorized into one of three performance categories. Complete information on the technical methodology for discriminating performance is contained in Appendix A, Section 2.3.

Figure 2a.22.1: Hospital Categorization



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.*

- ☐ Documentation of original self-assessment *(e.g., SF-36)*
- ☒ Electronic administrative data/ claims
- ☐ Electronic Clinical Data *(e.g., MDS)*
- ☐ Electronic Health/Medical Record
- ☐ External audit
- ☐ Lab data
- ☐ **M**anagement data
- ☐ Organizational policies and procedures
- ☐ Paper Medical Record/flowsheet
- ☐ Pharmacy data
- ☐ Public health data/vital statistics
- ☐ Registry data
- ☐ Survey-patient *(e.g., CAHPS)*
- ☐ Survey-provider
- ☐ Special or unique data, specify:

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: ☐ Individual ☐ Group ☐ Other
☒ Facility/Agency (e.g., hospital, nursing home)
☐ Health plan
☐ Integrated delivery system
☐ Multi-site/corporate chain
Population: ☐ National ☐ Regional/network
☐ State ☐ Counties/Cities

☐ Prescription drug plan

Program: ☐ Disease management ☐ QIO
☐ Other

☐ Measured at all levels
☐ Other (*Please describe*):

2a.26 Care Settings (*Check the settings for which the measure is specified and tested; check all that apply*)
Ambulatory Care: ☐ Amb Surgery Center ☐ Office ☐ Clinic ☐ Emergency Dept ☐ Hospital Outpatient

☐ Assisted Living
☐ Behavioral health/psychiatric unit
☐ Dialysis Facility
☐ Emergency medical services/ambulance
☐ Group Home
☐ Home
☐ Hospice

☒ Hospital
☐ Long term acute care hospital
☐ Nursing home/ Skilled Nursing Facility (SNF)
☐ Rehabilitation Facility
☐ All settings
☐ Unspecified or "not applicable"
☐ Other (*Please describe*):

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

Behavioral Health:
- ☐ Mental health
- ☐ Substance use treatment
- ☐ Other

Clinicians:
- ☐ Audiologist
- ☐ Chiropractor
- ☐ Dentist/Oral surgeon
- ☐ Dietician/Nutritional professional
- ☒ Nurses
- ☐ Optometrist
- ☐ PA/NP/Advanced Practice Nurse
- ☐ Pharmacist

- ☒ Physicians (MD/DO)
- ☐ Podiatrist
- ☐ Psychologist/LCSW
- ☐ PT/OT/Speech
- ☐ Respiratory Therapy
- ☐ Other

☐ Dialysis
☐ Home health
☐ Hospice/Palliative care
☐ Imaging services
☐ Laboratory
☐ 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 HF that are now reported under the Hospital Inpatient Quality Reporting Program. Our analysis aims to document the strength of associations among them.

The analysis reported here relies on data that are publicly reported on Hospital Compare. We merged process-of-care indicators and outcome-of-care indicators for HF collected between July 2006 and June 2009. We estimated composite measures for 3,586 hospitals (out of a potential 4,240 hospitals) for which:

1. The hospitals reported rates for all four process and all two outcome-of-care indicators
2. Each process-of-care indicator had at least five cases and each outcome-of-care indicator had at least 25 cases.

Background on Indicators Reported on *Hospital Compare*:
The indicators used in the construction of composites were drawn from *Hospital Compare*. The process-of-care indicators were drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2006 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for HF were based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.

2i.2 Analytic Method:

We carried out two analyses to explore the structure of the HF 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:

Although the HF composite was not intended as a reflective measure, psychometric properties do indicate a single underlying quality construct.

Table 2i.3.1 shows correlations across the process and outcome indicators. The correlations across the process-of-care indicators are significant and positive, and all are greater than 0.4, which indicates moderate correlation. Correlations between the process and outcome indicators are positive, albeit are weak, with values below 0.10. There is a weak negative correlation between mortality and readmission, which may reflect competing risks. That is, higher rates of mortality reduce the opportunity for readmission. Cronbach's alpha was estimated as 0.73, surpassing the commonly desired value of 0.70, suggesting that indicators are internally consistent.

The factor analysis of component measures produced a single factor with an eigenvalue greater than one. The eigenvalue for the first factor was almost 10 times that of the second factor, strongly suggesting that the component indicators represent one underlying construct.

2i
C☐
P☐
M☐
N☐

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

	HF 1	HF 2	HF 3	HF 4	Mort	Read
HF 1	1.00					
HF 2	0.47	1.00				
HF 3	0.40	0.51	1.00			
HF 4	0.59	0.51	0.53	1.00		
Mort*	0.07	0.05	0.10	0.18	1.00	
Read*	0.09	0.03	0.03	0.07	-0.13	1.00
Cronbach Alpha	0.73					

Notes:
* 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

	Factor Loadings			
	Factor 1	Factor 2	Factor 3	Uniqueness
HF 1	0.63	0.05	0.11	0.59
HF 2	0.73	0.01	-0.08	0.47
HF 3	0.75	-0.03	-0.09	0.43
HF 4	0.78	-0.02	0.06	0.38
Mort*	0.10	-0.30	0.06	0.90
Read*	0.08	0.30	0.04	0.90
Eigenvalues	2.12	0.19	0.04	
Proportion	1.12	0.10	0.02	
N	3,586			

Notes:
* 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:

The analysis of the component indicators' contribution to variability of the composite relies on data that are publicly reported on Hospital Compare. We merged process-of-care indicators and outcome-of-care indicators for HF collected between July 2006 and June 2009. We estimated composite measures for 3,586 hospitals (out of potential 4,240 hospitals) for which:

1. The hospitals reported rates for all four process and all two outcome-of-care indicators
2. Each process-of-care indicator had at least five cases and each outcome-of-care indicator had at least 25 cases.

Background on Indicators Reported on *Hospital Compare*:
The indicators used in the construction of composites were drawn from *Hospital Compare*. The process-of-care

2j
C☐
P☐
M☐
N☐

indicators were drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2006 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for HF were based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.

2j.2 Analytic Method:

In order to assess the contribution of each indicator to variability in the HF composite, we compare the percent change in (1) the variance and (2) the inter-quartile range (IQR) of the composite and of the process and outcome domain scores when a process or outcome indicator is removed. Results appear in Table 2j.3.1.

2j.3 Results:

In Table 2j.3.1, positive values indicate that addition of the component indicator tends to *reduce* the variance or IQR. Only one indicator, HF2 (Percent of HF Patients with Evaluation of LVS Function), exhibits a positive effect on the composite variance. Because the outcome domain contains only two component indicators, readmission and mortality both have strong negative effects on the variance of the domain score. The strong variance-reducing effect of mortality appears to be the result of its tight distribution.

Table 2j.3.1. Change in Inter-quartile Range and Variance of the Composite, Process and Outcome Domains with the Removal of Indicators

	Overall Composite		Process Domain		Outcome Domain	
Remove:	Change in Variance (%)	Change in Inter-quartile Range (%)	Change in Variance (%)	Change in Inter-quartile Range (%)	Change in Variance (%)	Change in Inter-quartile Range (%)
HF 1	21.10	7.93	21.78	8.29	-	-
HF 2	-32.83	-33.84	-34.22	-36.41	-	-
HF 3	4.65	4.86	4.77	5.16	-	-
HF 4	42.62	36.55	44.66	37.92	-	-
Mortality	2.54	2.04	-	-	194.17	72.82
Readmission	0.09	1.28	-	-	25.83	17.24

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 national 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 uses data that are publicly reported on Hospital Compare. We merged process indicators and outcome indicators for HF collected between July 2006 and June 2009. We estimated composite measures for 3,586 hospitals (out of potential 4,240 hospitals) for which:

1. The hospitals reported rates for all four process and all two outcome-of-care indicators
2. Each process-of-care indicator had at least five cases and each outcome-of-care indicator had at least 25 cases.

Background on Indicators Reported on Hospital Compare:

The indicators used in the construction of composites were drawn from Hospital Compare. The process-of-care indicators were drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2006 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for HF were based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.

2k
C☐
P☐
M☐
N☐

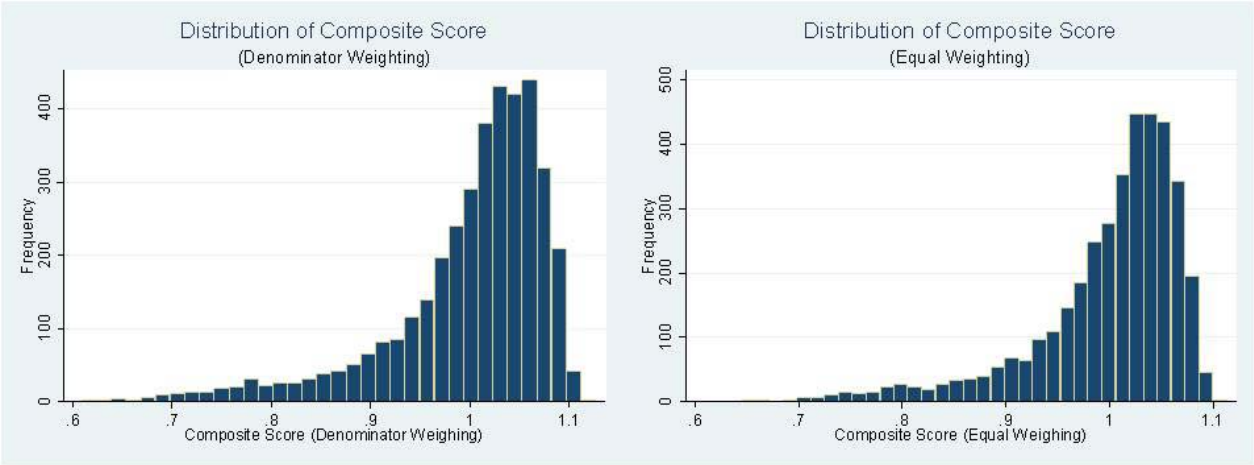
2k.2 Analytic Method:

We compared the distribution of the HF composite measure with equal and differential weighting.

2k.3 Results:

Figure 2k.3.1 displays the distribution of the HF composite measure with equal and differential weighting. As the figure shows, denominator weighting slightly increases the percentage of hospitals with higher composite scores. A table of the distribution of composite scores is also provided in the appendix (Table 2k.3.1)

Figure 2k.3.1: Distribution of Composite Score with Denominator and Equal Weighting



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 large numbers 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 HF 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 HF process and outcome-of-care indicators and represents the quality construct expressed earlier.

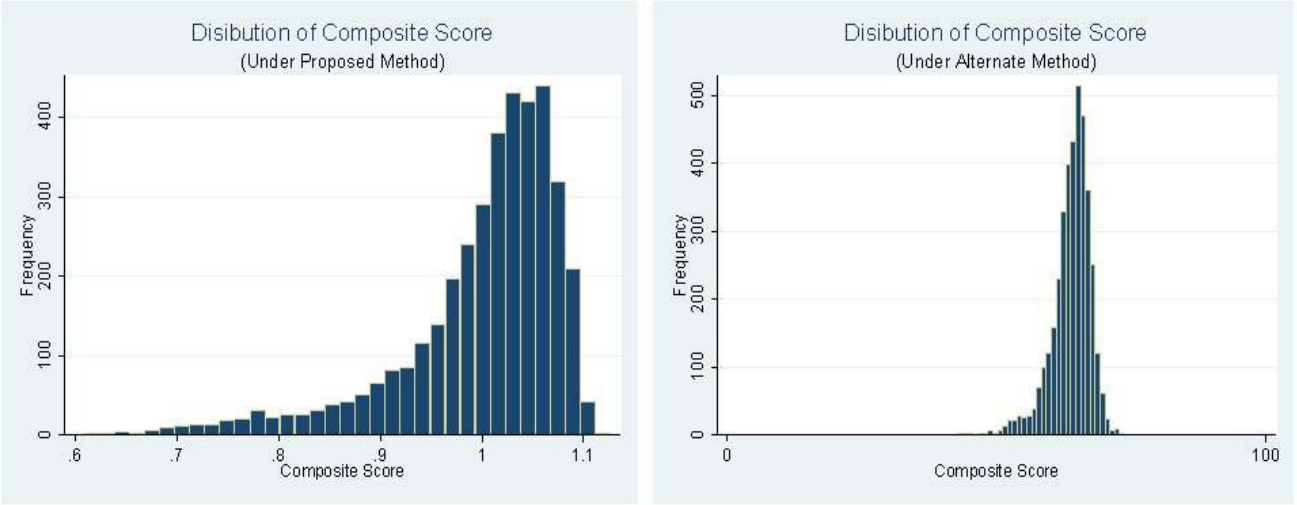
2k.5 Indicate if any alternative scoring/aggregation methods were tested and why not chosen:

In addition to the preferred compositing approach, we tested an alternative scoring approach that differed on two levels (Alternate Method). First, we estimated composite scores for hospitals that were missing less than half of

the process- and outcome-of-care indicators. That is, if a hospital had two or more process and one or more outcome indicator, a composite score was estimated. We imputed missing values with the national mean. Second, we used an alternative standardization approach by subtracting the national mean and dividing by the standard deviation, before taking the simple average of the two domain scores. Because this could result in negative composite values for some hospitals, the score was then rescaled to a range between zero and one hundred. It should be noted that this approach was a method we used when we initially presented our composite measure to the NQF in February, 2011.

In Figure 2k.5.1, we present distributions of the two alternative scoring methods. The figures show that the second approach (Alternate Method) leads to composite scores with a tight distribution as a result of the standardization approach; therefore, our proposed approach should provide users with a distribution that is easier for consumers to view. Furthermore, our reevaluated compositing approach reduces potential misinterpretations by consumers that the composite score is an actual rate between zero and 100 percent. A table of the distribution of composite scores is also provided in the appendix (Table 2k.5.1)

Figure 2k.5.1: Comparision of Compositing Approaches



Furthermore, we considered, but rejected, alternative weighting schemes that would reduce the weight assigned to indicators that were strongly left-skewed (often referred to as "topped off"). This can be done, for example, by constructing weights that depend on the difference between the national mean for an indicator and the highest possible score. First, we are disinclined to make judgments about the relative importance of endorsed indicators. It does not appear reasonable to argue that an element of care becomes "less important" in a composite because many hospitals report providing it. Second, at a purely practical level, the distributions of the four HF process indicators do not sharply differ from one another, so weighting in this fashion would produce a result resembling equal weighting. Finally, and perhaps most importantly, such an approach to weighting would make a hospital's score dependent on the behavior of other hospitals. For example, a hospital that performed well on indicator A and poorly on indicator B would receive a higher score if other hospitals performed poorly on A and well on B than it would if other hospitals performed well on A and poorly on B. This is not, in our view, a desirable property for a composite to have.

2l. Analysis of missing component scores

2l.1 Data/sample:

Construction of the composite scores relies on data that are publicly reported on Hospital Compare. We merged process-of-care indicators and outcome-of-care indicators for HF collected between July 2006 and June 2009. We estimated composite measures for 3,586 hospitals (out of potential 4,240 hospitals) for which:

1. The hospitals reported rates for all four process and all two outcome-of-care indicators
2. Each process-of-care indicator had at least five cases and each outcome-of-care indicator had at least 25 cases.

2l
C☐
P☐
M☐
N☐

Of the 4,240 hospitals 654 did not receive a composite score for one or more of the following reasons:

1. The hospital was missing a rate for one or more of the process- and/or outcome-of-care indicators (1.2%)
2. The hospital reported a case size of zero for one or more of the process-of-care indicators; therefore a hospital specific rate was not reported (3.8%)
3. The hospital reported a case size of greater than zero, but less than five cases for one or more process-of-care indicator (1.2%)
4. The hospital reported a case size of less than 25 cases for one or more outcome-of-care indicator (0.8%)

Background on Indicators Reported on *Hospital Compare*:

The indicators used in the construction of composites were drawn from *Hospital Compare*. The process-of-care indicators were drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2006 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for HF were based on Medicare claims for hospital stays with discharge dates between July 2006 and June 2009.

2I.2 Analytic Method:

We examined whether there were differences in the distribution of the process- and outcome-of care rates for all hospitals compared to those hospitals for which there were no missing process- and outcome-of-care indicators so that composites were estimated for these hospitals.

2I.3 Results:

Figures 2I.3.1 and 2I.3.2 show that there is very little difference in the distribution of each of the components indicators between those hospitals that had a composite score calculated (i.e., those with no missing process- or outcome-of-care indicators and for the full sample of hospitals. Specific distributions for each of the indicators are available in Table 2I.3.1 in the appendix.

Figure 2I.3.1: Comparison of Distribution for Process-of-Care Indicators

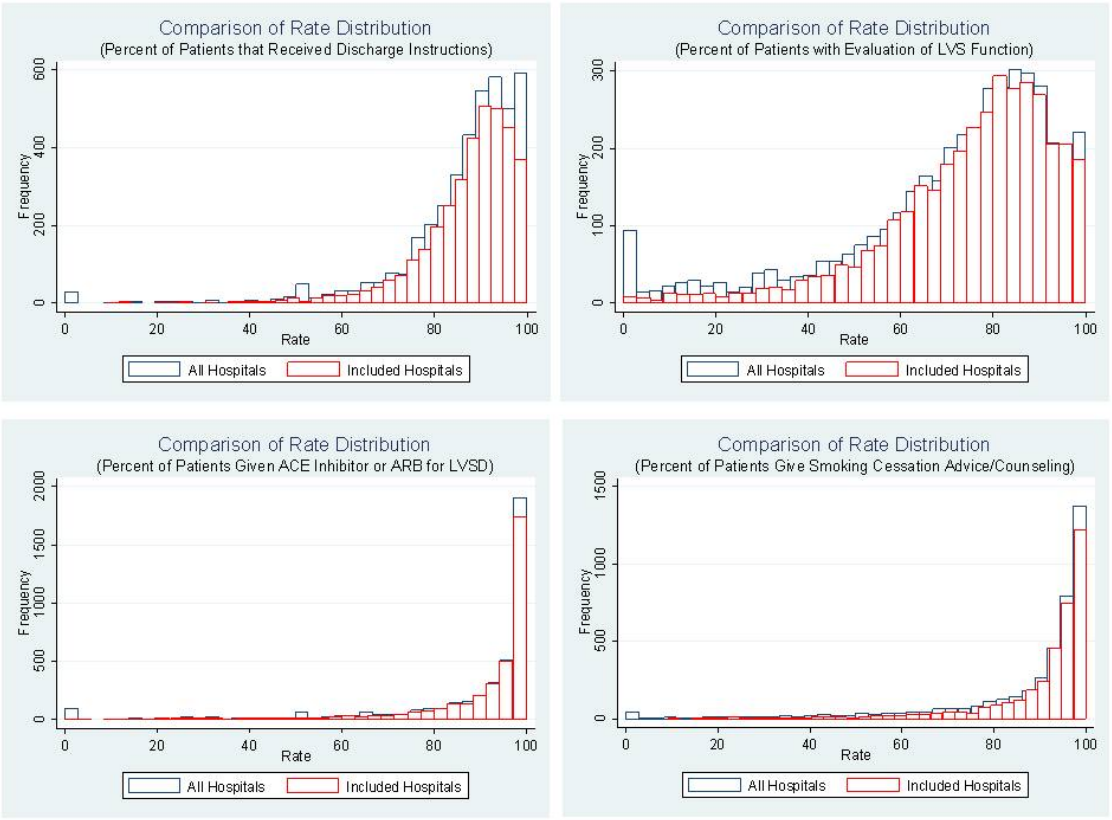
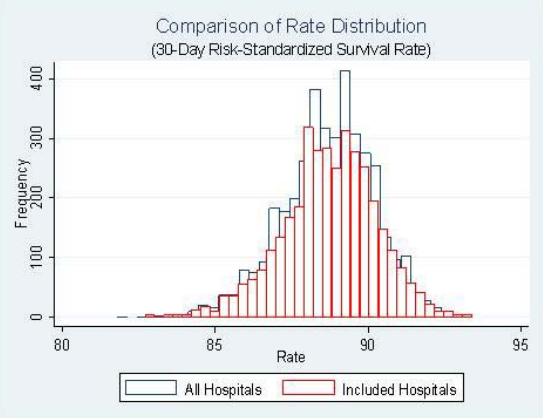
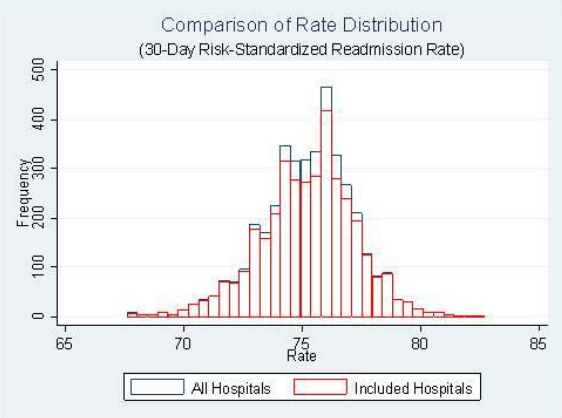


Figure 2I.3.2: Comparison of Distribution for Outcome-of-Care Indicators





2b. Reliability testing of composite score

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

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

1. **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 examined the reliability of all four (4) HF process-of-care indicators that make up the HF composite. Their sample included 30 hospitals, representing a diverse range of geographic locations, sizes, settings (urban/rural), and ownership categories (profit/not-for-profit); 17 of these collected HF data. A randomly selected set of de-identified, previously abstracted medical records was transmitted from the hospitals' performance measurement vendors and HF 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 HF-4 (Smoking cessation counseling) the sample size was less than 50.

2. **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. The reliability of the component measures is assessed on a quarterly basis by CMS' contractor, CDAC (Clinical Data Abstraction Center). This assessment uses a sample of five (5) randomly patient records from each hospital participating in the RHQDAPU program, which includes hospitals from all states but Maryland and Puerto Rico.[ii]

2b
C☐
P☐
M☐
N☐

[ii] 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)*:

1. **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 percent agreement for continuous variable elements and chance-corrected agreement using Cohen's kappa for binary data elements.

2. **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% 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)*:

1. **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 below summarizes the reliability statistics for the HF 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 HF-3 (ACEI for LVSD), substantial agreement (kappa ranging from 0.61 – 0.80) for HF-2 (LVSD evaluation) and HF-4 (smoking cessation), and moderate agreement (kappa ranging from 0.41 – 0.60) for HF-1 (discharge instructions).

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

HF Component Measure	N	Agreement (%)	Kappa
HF-1*			
Discharge instructions to address activity	180	86.1	0.65
Discharge instructions to address diet	180	90.0	0.73
Discharge instructions address follow-up	180	87.8	0.47
Discharge instructions address medications	180	90.6	0.53
Discharge instructions address symptoms	180	86.1	0.71
Discharge instructions address weight	180	90.6	0.81
HF-2	201	88.6	0.78
HF-3	116	94.0	0.88
HF-4	35	88.6	0.68

   Notes:
   *HF-1 includes written instructions or educational material given to patient or caregiver at discharge or during the hospital stay addressing all of the following: activity level, diet, discharge medications, follow-up appointment, weight monitoring, and what to do if symptoms worsen.

2. **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% of hospitals exceeded the 80% 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-of-care measures from July 2008-June 2009 with outcome-of-care measures from July 2006-June 2009. The second data set merges process-of-care measures from July 2007-June 2008 with outcome-of-care measures from July 2006-June 2009. Composite measures are calculated for hospitals where:

1. The hospitals reported rates for all four process and all two outcome-of-care indicators
2. Each process-of-care indicator had at least five cases and each outcome-of-care indicator had at least 25 cases.

The composite measures from these time periods were then compared. Across these two data collection periods, 2,906 hospitals had valid composite measures for HF.

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 0.41 ($p<0.001$), indicating moderate predictive validity of the composite. (See Table 2c.3.1) A large number of hospitals (around 55 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 (around 1 percent) occupy the first quartile in 2007-2008 and the fourth quartile in 2008-2009, and vice versa. Across the two separate time periods, around 36 percent of hospitals' categorizations differ by one quartile (i.e., 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.

Table 2c.3.1. Comparison of Composite Measures, by Reporting Period

	2008-2009 Reporting**				
2007-2008 Reporting*	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Total
Quartile 1***	486	167	56	18	727
Quartile 2	183	310	186	47	726
Quartile 3	45	200	327	155	727
Quartile 4	13	49	158	506	726
Total	727	726	727	726	2,906
Spearman Correlation****	0.77 (0.00)				
Kappa Statistic	0.41 (0.00)				

Notes:
* 2007-2008 reporting: process- and outcome-of-care measures for HF with a data collection period of July 2007 to June 2008
** 2008-2009 reporting: process- and outcome-of-care measures for HF with a data collection period of July 2008 to June 2009
*** Higher quartile categories indicate that the hospital had higher (i.e., better quality) composite measures.
**** P-values in parentheses.

2c
C☐
P☐
M☐
N☐

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. We merged process-of-care indicators and outcome-of-care indicators for HF collected between July 2006 and June 2009. We estimated composite measures for 3,586 hospitals (out of potential 4,240 hospitals) for which:

1. The hospitals reported rates for all four process and all two outcome-of-care indicators
2. Each process-of-care indicator had at least five cases and each outcome-of-care indicator had at least 25 cases.

Background on Indicators Reported on *Hospital Compare*:
The indicators used in the construction of composites were drawn from *Hospital Compare*. The process-of-care indicators were drawn from Medicare hospital administrative claims data and medical record documents with discharge dates between July 2006 and June 2009. The hospital outcome-of-care indicators for 30-day risk-adjusted mortality and readmission for HF were 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/meaningfully differences in performance *(type of analysis & rationale)***:**

To examine meaningful differences in composite measures across hospitals, we compare 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 meaningfully differences in performance)* **:**

Note: CMS has not decided how hospital performance will ultimately be displayed 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.

The total number of hospitals in each performance category is displayed in Table 2f.3.1. The table shows that there are meaningful differences in the overall composite score as 1,745 or around 48 percent of hospitals are categorized as being statistically different from the national average. Of the remaining 52 percent, around half of the hospitals' performances are significantly worse than the national average. The hospital performance category for the outcome-of-care domain is consistent with the hospital performance categories displayed on Hospital Compare for each of the indicators. That is, very few number of hospital s are in the "better than" or "worse than" the national rate categories.

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

Type of Composite	Performance Category		
	Worse than National Rate	No Different than National Rate	Better than National Rate
Overall	955	886	1,745
Process Domain	1,051	614	1,921
Outcome Domain	130	3,274	182

2f
C☐
P☐
M☐
N☐

Figure 2f.3.1 displays the distribution of hospitals' composite scores. In Figure 2f.3.2, the distribution of scores for the process- and outcome-of-care domains are also displayed.

Figure 2f.3.1: Distribution of Composite Score

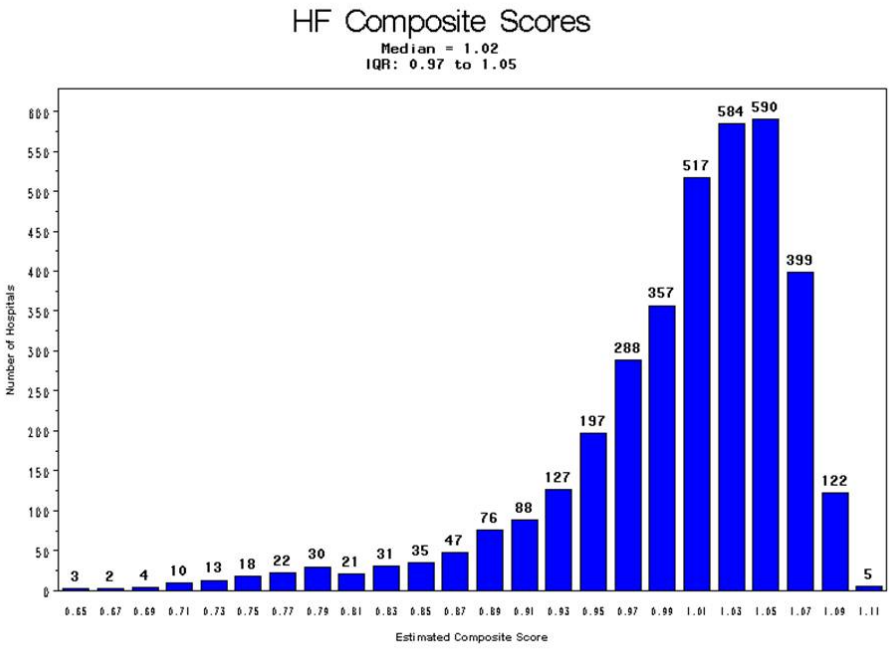
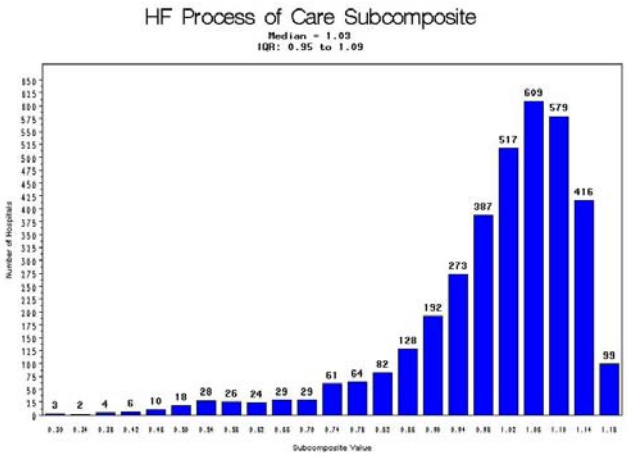
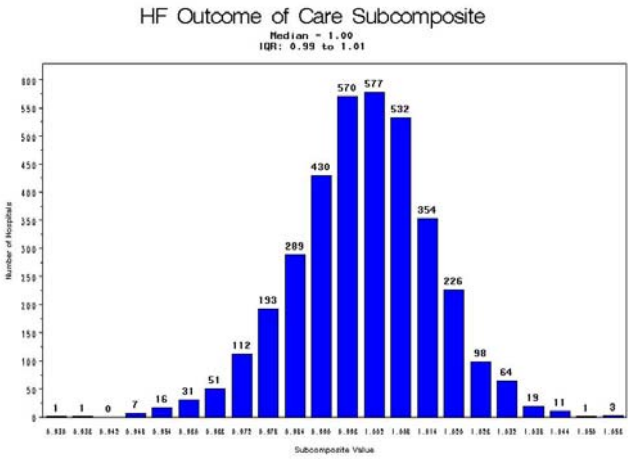


Figure 2f.3.2: Distribution of Process- and Outcome-of-Care Domains





2h. Disparities in Care

2h.1 If measure is stratified, provide stratified results *(scores by stratified categories/cohorts)*:
The measure is not stratified.

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:
The distribution of composite scores by the following hospital characteristics:

1. Hospital bed size
2. Ownership status

2h
C☐
P☐
M☐
N☐
NA☐

3. Teaching status
4. Census region
5. Percentage of patients that was black.

Slight differences in the distribution were observed for hospital bed size, teaching status, census region, and race. Figures 2h.3.1-2h.3.4 present distributions for these characteristics. This analysis demonstrates that composite scores increase at most points along the distribution when hospital bed sizes increases as well as when the hospital is a teaching hospital (although teaching hospitals may also be more likely to be larger hospitals). This analysis also finds that there is very little difference in the distribution of the composite measure by the percentage of blacks served by hospital.

Figure 2h.3.1: Distribution of Composite Score, by Hospital Bed Size

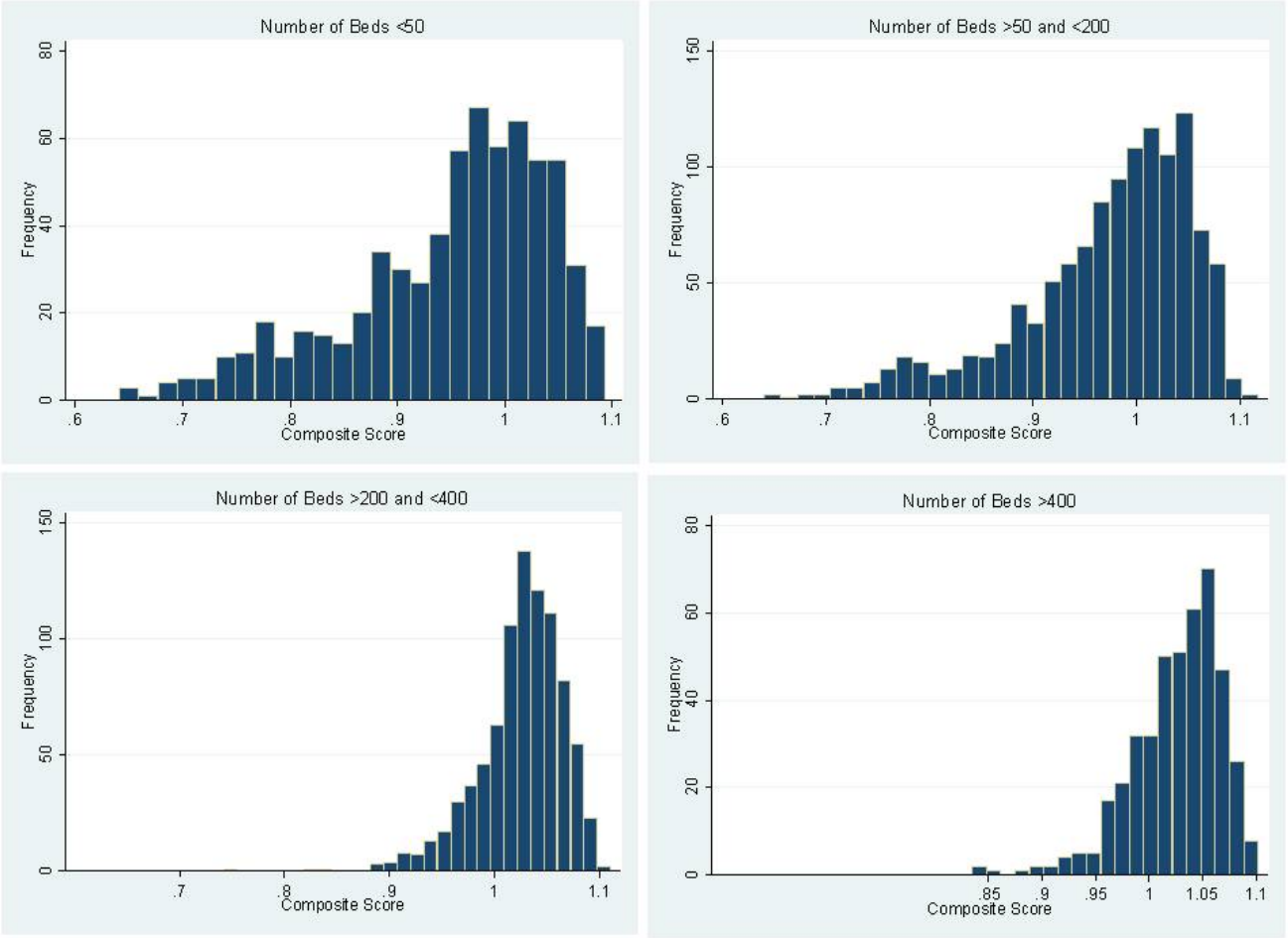
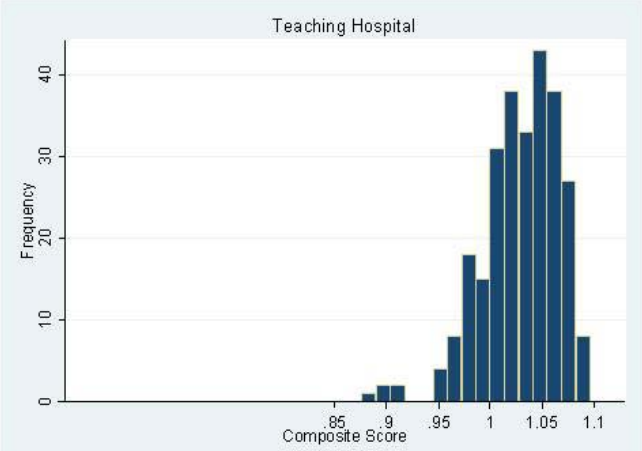


Figure 2h.3.2: Distribution of Composite Score, by Hospital Teaching Status



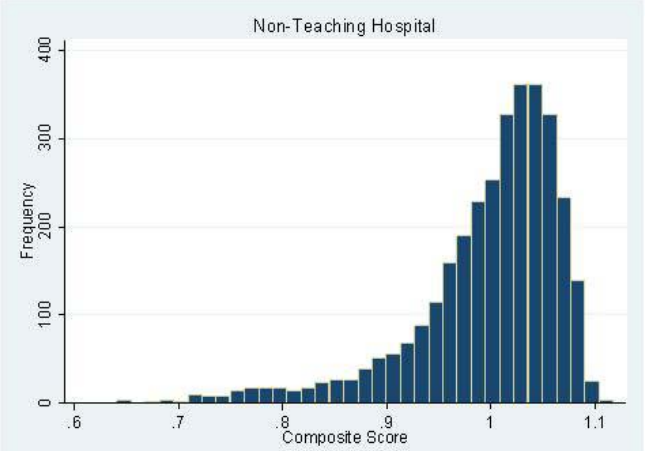
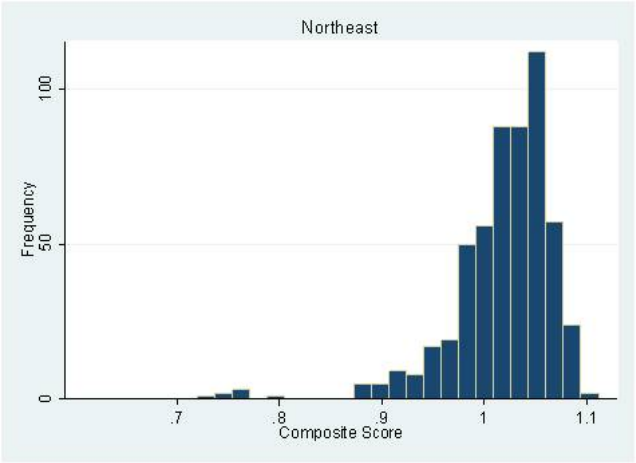
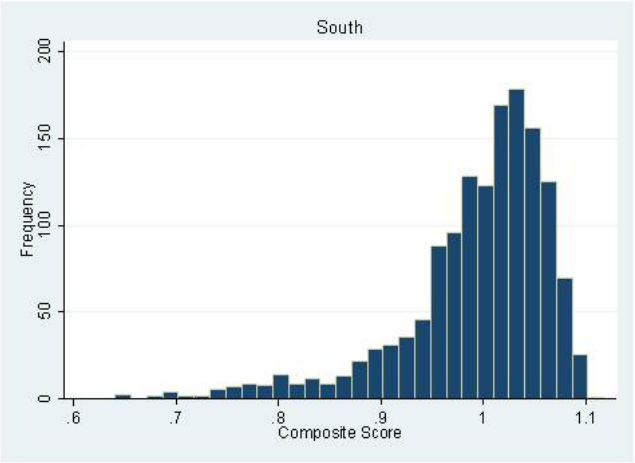
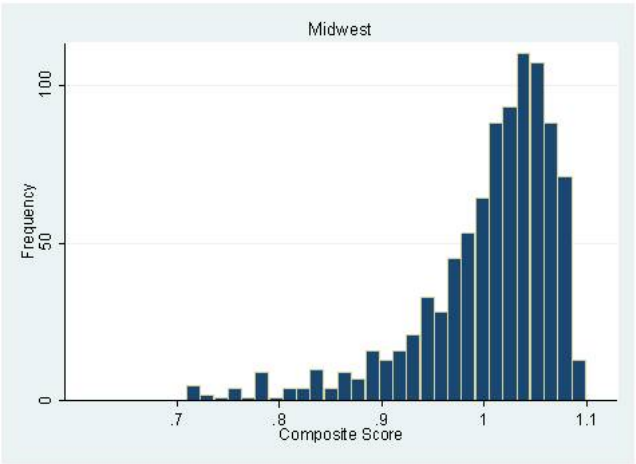


Figure 2h.3.3: Distribution of Composite Score, by Census Region







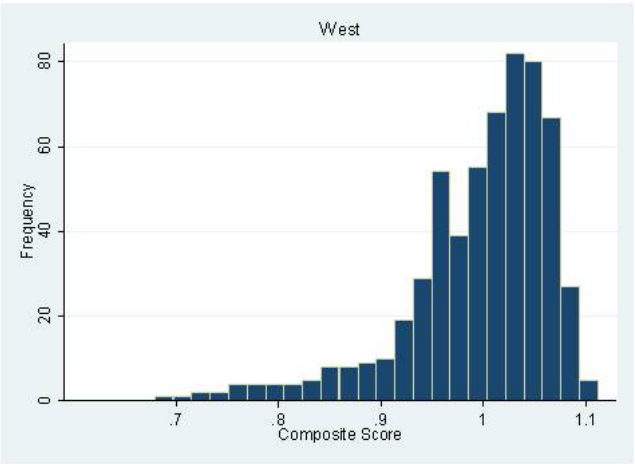
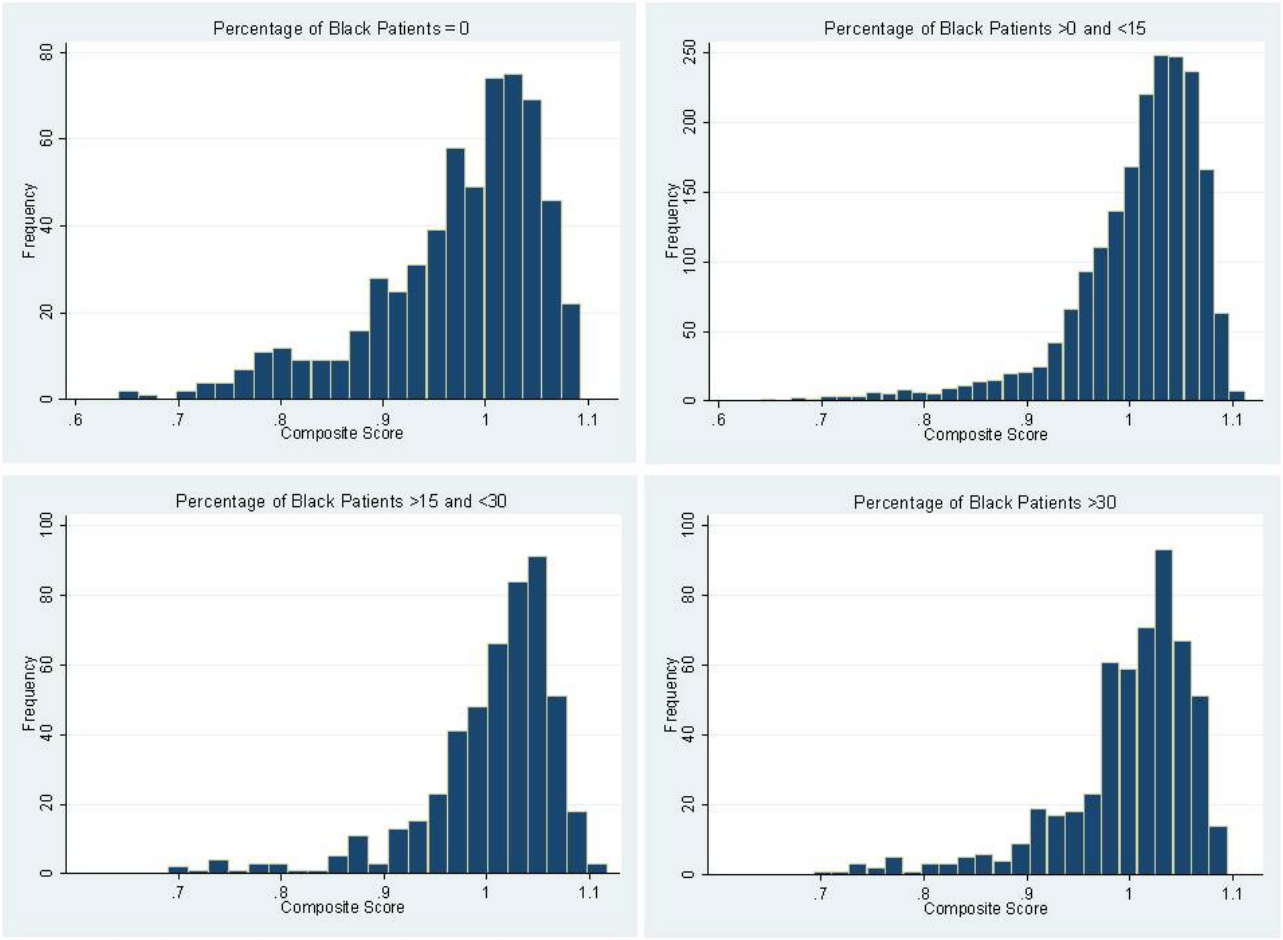


Figure 2h.3.4: Distribution of Composite Score, by Race



If the component measures are combined at the patient level, complete 2d.

2d. Exclusions Justified

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

2d.2 Citations for Evidence:

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

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

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

2d
H☐ M☐ L☐ N☐ NA☐

If the component measures are combined at the patient level and include outcomes, complete 2e.

2e. Risk Adjustment

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

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

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

2e
H☐ M☐ L☐ N☐ NA☐

2e.4 If outcome or resource use measure is not risk adjusted, provide rationale:	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties*?	**2**
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, **met?** **Rationale:**	*2* **C☐** **P☐** **M☐** **N☐**
3. USABILITY	
Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (composite measure evaluation criteria)	**Eval**

3a. Meaningful, Understandable, and Useful Information

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

3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large) (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*)**:**

Following NQF endorsement, public reporting is expected on *Hospital Compare* sometime in 2012.

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, CMS plans to publicly report this composite on *Hospital Compare*. CMS' current timetable calls for this public reporting to occur in 2012. 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 HF patients, and be an asset to clinicians. In work that is directly relevant to the proposed measure, Borck et al held a series of focus groups that evaluated consumer and clinician understanding of condition-specific composite measures for AMI, HF, Pneumonia and SCIP that are very similar to the proposed measure. As well, their work evaluated understanding of AHRQ and Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) composite measures. In addition, work by Smith et al examined the interpretability of Hospital Compare data, including several of the component measures in the proposed composite. A further study by Peters et al also provides insight into consumer understanding of publicly reported hospital quality measures, while L&M Policy Research LLC specifically reports on consumer understanding of the 'readmissions' outcome-of-care indicator, one of two possible outcome-of-care indicators included in this composite.

1. **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 used a convenience sample of 21 consumers in the Baltimore, MD area. Participants ranged from 45-70 years old, were 67% women, and 48% Medicare beneficiaries.
   Round 2: Borck et al used a convenience sample of 18 consumers and 5 physicians from the Miami, FL

3a
C☐
P☐
M☐
N☐

area. The group had an age range of 45 to 70 years old, and were made up of a majority of men and Medicare beneficiaries.

2. **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 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 out of 51 (92%) were over 65 years, and of the over 65 group, 53% were Medicare beneficiaries at risk for heart disease. Among the health care providers, 30% were nurses, 38% were primary care physicians, and the remainder were cardiologists and pulmonologists.

3. **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 employed a convenience sample of employed-age adults (18 – 64 years old, mean age of 37, 48% female, and 76% white) to determine whether providing only the most important quality information increase comprehension and information use. Half of the sample had lower levels of education (high school or less), 45% had health insurance and 74% had an annual household income of less than $20,000.

4. **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, aged 50 – 70 years, most of whom were previously diagnosed with heart disease; 8 caregivers, aged 40 – 60 years; and 6 physicians who were primary care physicians, cardiologists, and pulmonologists.

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

1. **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 were utilized to assess comprehension of composite measures, organization and presentation of the site, and composite labels and descriptions.

2. **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 were first allowed to navigate the website independently and then asked a series of open-ended questions using an approved protocol during an approximately two-hour period.

3. **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 were presented with hospital quality data in a different format. In the first group, data on cost, quality, and non-quality information was unordered. In the second, cost and quality data was highlighted and presented first, while non-quality 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.

4. **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 them and their family members.

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

1. **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, Pneumonia and SCIP) and the HCAHPS composite measure. Round 1 also revealed that almost all participants preferred more descriptive definitions of the composites, and specifically that included a list of all the component measures making up the composite. Similarly to Round 1 findings, in Round 2 respondents were also found to be able to correctly interpret the star ratings for condition-specific quality ratings composites and the HCAHPS composite. 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, should also be easy for consumers to use. Moreover, any composite definition posted on Hospital Compare should include a list of all component measures.

2. **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 is intended to address this very issue by creating a single benchmark that enables consumers to evaluate the quality of care at a given hospital for a given condition.

3. **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.**

   Similarly to Smith et al (2005), Peters et al (2007) determined that less is more with regards 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% of respondents choose the highest quality hospital Y when only

the quality information was shown, while in the other two formats it was by selected 48% (ordered group) and 40% (unordered group). Such results reinforce the idea that a composite measure may enhance the utility of hospital quality data for consumers.

4. **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. While 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 all NQF-endorsed. However there are currently no NQF-endorsed composite measures that provide a single indication of a hospital's quality of care for HF 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 following:

1. NQF #0531 Patient Safety for Selected Indicators (AHRQ) Endorsed June 19, 2009
2. NQF #0530 Mortality for Selected Conditions (AHRQ) Endorsed June 19, 2009

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.

(for NQF staff use) **Notes on similar/related endorsed or submitted measures:**

3b. Harmonization
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.

3b C☐ P☐ M☐ N☐ NA☐

3c. Distinctive or Additive Value
3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures:

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 HF patients.

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.

3c C☐ P☐ M☐ N☐

3d. Decomposition of Composite
3d.1 Describe the information that is available from decomposing the composite into its components:

The component measures include the following information:

1. Percent of HF Patients Receiving Discharge Instructions
2. Percent of HF with Evaluation of LVS Function

3d C☐ P☐ M☐ N☐

3. Percent of HF Patients Given ACE Inhibitor or ARB for LVSD 4. Percent of HF Patients Given Smoking Cessation Advice/Counseling 5. Heart Failure (HF) 30-day Mortality 6. Heart Failure (HF) 30-day Readmission	
3e. Achieved stated purpose **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 HF 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 HF, achieving the purpose of the composite.	**3e** C☐ P☐ M☐ N☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability*?	**3**
Steering Committee: Overall, to what extent was the criterion, *Usability*, **met?** **Rationale:**	**3** C☐ P☐ M☐ N☐
4. FEASIBILITY	
Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (composite measure evaluation criteria)	**Eval**
4a. Data Generated as a Byproduct of Care Processes **4a.1 How are all the data elements that are needed to compute measure scores generated?** (*Check all that apply)* ☐ Data are generated as a 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, registry)* ☐ Survey ☐ Other *(e.g., patient experience of care surveys, provider surveys, observation)*, *Please describe*:	**4a** C☐ P☐ M☐ N☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) ☒ Yes ☐ No **4b.2 If no, specify the near-term path to achieve electronic capture by most providers.** N/A *Note: Measure stewards will be asked to specify the data elements for electronic health records at a later date*	**4b** C☐ P☐ M☐ N☐
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.** Our measures are not susceptible to inaccuracies, errors, or unintended consequences; the component outcomes are well-specified in hospital administrative data.	**4d** C☐ P☐ M☐ N☐
4e. Data Collection Strategy/Implementation **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:**	**4e** C☐ P☐ M☐ N☐

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 HF composite is therefore assured.

4.2 **Costs to implement the measure** (*costs of data collection, fees associated with proprietary measures*)**:**

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.

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

If the component measures are combined at the patient level, complete 4c.

4c. Exclusions
4c.1 **Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** ☐ No ☐ Yes ►**If yes, provide justification**

4c
H☐ M☐ L☐ N☐ NA☐

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

4

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

4
C☐ P☐ M☐ N☐

RECOMMENDATION

Steering Committee: Do you recommend for endorsement?
Comments:

Y☐ N☐ A☐

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 If different from Measure Steward**
Organization: Mathematica Policy Research
Street Address: Mathematica Policy Research **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: Mathematica Policy Research ☐ Measure Steward ☒ Measure Developer
First Name: Marian **Last Name:** Wrobel **Credentials (MD, MPH, etc.):** Ph.D.
Email: MWrobel@mathematica-mpr.com **Telephone:** 617-301-8971 ext:

Co.6 **List any additional organizations that sponsored/participated in measure development:**

ADDITIONAL INFORMATION

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of workgroup/panel member names and organizations. Describe the group's role in measure

development.

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.

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 Bodën 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 Associations and APME Chair

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.

Ad.2 If adapted, name of original measure: N/A
Ad.3 If adapted, original specifications ☐ **attachment or Ad.4 web page URL:**

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 Year the measure was first released: N/A
Ad.7 Month and Year of most recent revision: N/A
Ad.8 What is the frequency for review/update of this measure? Annually
Ad.9 When is the next scheduled review/update for this measure? 2012

Ad.10 Copyright statement/disclaimers:

Ad.11 Additional Information ☐ **attachment or web page URL:**

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. ☒

Date of Submission (*MM/DD/YY*)**: Initial:** 12/13/10 **Resubmission**: 3/15/11

The National Quality Forum
Composite Measure of Hospital Quality for HF

Appendix A
Technical Supplement

Submitted By:

Mathematica Policy Research, Inc
Marian Wrobel, Ph.D., Project Director
Bob Schmitz, Ph.D.
Mai Hubbard, Ph.D.
Jessica Ross, M.P.H.

Boston University
Jim Burgess, Ph.D.
Gary Young, Ph.D.

Prepared for:
Centers for Medicare and Medicaid Services

March 2011

TABLE OF CONTENTS

SECTION 1
BACKGROUND

1.1 Overview

The composite measure of quality of hospital care for HF 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 four individual process-of-care indicators and two outcome-of-care indicators meant to capture the quality of hospital care provided to patients with HF. 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: Acute Myocardial Infarction (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, 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	• Include the same set of process and outcome measures as Hospital Compare
Produce differences in composite values that are clinically and statistically meaningful and reflect true differences in underlying quality	• 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	• Process measures are available when the number of eligible discharges is five or more; outcome variables are available when the number of eligible discharges is 25 or more
Focus more on measures relevant to more patients	• Construct process and outcome composites using weights based on national denominators
Method is scientifically acceptable and acceptable to stakeholders	• 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 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
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	
Method is relatively simple Hospitals can duplicate results	• Use equal weighting to combine process and outcome domains • Reliability weights are a function of a hospital's number of cases and national parameters

SECTION 2
METHOD OF SCORING AND AGGREGATION

2.1 Estimation of the 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 two domain scores related to hospital inpatient quality. The first domain is comprised of four process-of-care indicators and the second domain is comprised of two outcome-of-care indicators. All of these indicators are publically reported by the CMS on *Hospital Compare* and NQF endorsed.

To construct the process-of-care domain, the process-of-care indicators were 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). More information regarding the reliability-weight adjustment is available in Section 2.2. Then, each indicator was standardized by dividing by the national mean of the indicator. Since the outcome-of-care indicators have already been risk-standardized using a hierarchical generalized linear modeling technique, the outcome-of-care indicators were not reliability-weight adjusted. Similarly to the process-of-care indicators, the outcome-of-care indicators were also was standardized by dividing by the national mean of the indicator.

Consistent with the approach used for the AHRQ measures, CMS used denominator weighting in constructing the process- and outcome-of-care domains. Denominator weighting places relatively more weight on measures that apply to relatively more patients nationally. More specifically, the process of care domain for hospital j = 1,…, J can be described as a denominator weighted average of a standardized reliability-weight adjusted process-of-care indicator k=1,…K,

$$P_j^* = \sum_{k=1}^{K} \left(\frac{\sum_{j=1}^{J} n_{jk}}{\sum_{k=1}^{K} \sum_{j=1}^{J} n_{jk}} * \frac{P_{jk}^*}{P_k^{nat}} \right) \qquad \text{(eq. 2.1.1)}$$

where P_k^{nat} is the national rate of a process-of-care indicator and n_{jk} is the total number of cases for a process-of-care indicator at hospital *j*.

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

$$O_j^* = \sum_{l=1}^{L} \left(\frac{\sum_{j=1}^{J} n_{jl}}{\sum_{l=1}^{L} \sum_{j=1}^{J} n_{jl}} * \frac{O_{jl}^*}{O_l^{nat}} \right) \qquad \text{(eq. 2.1.2)}$$

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

The overall composite score (C_j^*) is then estimated as a simple average of the two domains:

$$C_j^* = \frac{1}{2}(P_j^*) + \frac{1}{2}(O_j^*)$$

(eq. 2.1.3)

2.2 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*
$k = 1,\ldots K$	Process-of-care indicator
$j = 1,\ldots, J$	Hospital index

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

$$P_{jk}^* = W_{jk}P_{jk} + (1 - W_{jk})P_k^n$$

(eq. 2.2.1)

where W_{jk} is the reliability-weight:

$$W_{jk} = \frac{\sigma_{sk}^2}{\sigma_{sk}^2 + \sigma_{wjk}^2}$$

(eq. 2.2.2)

σ_{sk}^2 is the signal variance:

$$\sigma_{sk}^2 = \frac{\sum_{i=1}^{J}(P_{ik} - P_k^n)^2}{N_k} - \frac{\sum_{i=1}^{J} P_{ik}(1 - P_{ik})}{\sum_{i=1}^{J} n_{ik}}$$

(eq. 2.2.3)

and σ_{wjk}^2 is the within variance:

$$\sigma_{wjk}^2 = \frac{\sum_{i=1}^{J} P_{ik}(1 - P_{ik}) \frac{n_{ik}}{\sum_{l=1}^{J} n_{lk}}}{n_{jk}}$$

(eq. 2.2.4)

SECTION 3
PERFORMANCE DISCRIMINATION

3.1 Method for Discriminating Performance

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,…K	Process-of-care indicator
l =1,…L	Outcome-of-care indicator
j = 1,…,*J*	Hospital index

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

$$P_j^* = \sum_{k=1}^{K} \left(\frac{\sum_{j=1}^{J} n_{jk}}{\sum_{k=1}^{K} \sum_{j=1}^{J} n_{jk}} * \frac{P_{jk}^*}{P_k^{nat}} \right) \qquad \text{(eq. 2.3.1)}$$

The hospital's outcome-of-care domain composite score (O_j^*) is estimated as a denominator weighted average of the standardized risk-adjusted outcome-of-care indicator rates:

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

(eq. 2.3.2)

The composite measure (C_j) is a simple average of the normalized process-of-care and outcome-of-care sub-composites.

$$C_j^* = \frac{1}{2}\left(P_j^*\right) + \frac{1}{2}\left(O_j^*\right)$$

(eq. 2.3.3)

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

$$\mathrm{Var}\left(C_j^*\right) = \mathrm{Var}\left[\frac{1}{2}P_j^* + \frac{1}{2}O_j^*\right]$$

$$= \left(\frac{1}{2}\right)^2 \mathrm{Var}\left[\sum_{k=1}^{K}\left(\frac{\sum_{j=1}^{J} n_{jk}}{\sum_{k=1}^{K}\sum_{j=1}^{J} n_{jk}}\right)\frac{P_{jk}^*}{\widehat{\mu_{Pk}}} + \sum_{l=1}^{L}\left(\frac{\sum_{j=1}^{J} n_{jl}}{\sum_{l=1}^{L}\sum_{j=1}^{J} n_{jl}}\right)\frac{O_{jl}^*}{\widehat{\mu_{Ol}}}\right]$$

$$= \left(\frac{1}{2}\right)^2 \left\{ \frac{1}{\left(\sum_{k=1}^{K}\sum_{j=1}^{J} n_{jk}\right)^2} \sum_{\mathrm{k=1}}^{\mathrm{K}} \left[\left(\frac{\sum_{j=1}^{J} n_{jk}}{\widehat{\mu_{Pk}}}\right)^2 \frac{P_{jk}^*(1-P_{jk}^*)}{n_{jk}} \right] \right.$$

$$\left. + \frac{1}{\left(\sum_{l=1}^{L}\sum_{j=1}^{J} n_{jl}\right)^2} \sum_{\mathrm{l=1}}^{\mathrm{L}} \left[\left(\frac{\sum_{j=1}^{J} n_{jl}}{\widehat{\mu_{Ol}}}\right)^2 \mathrm{Var}\left(O_{jl}^*\right) \right] \right\}$$

(eq. 3.4)

given the following assumptions:

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

A4. $\text{cov}(P^*_{jm}, O^*_{jn}) = 0$

SECTION 4
RESULTS

4.1 Results for Section 2k.3

Table 2k.3.1. Comparison of Distribution of HF Composite Measure by Weighting Method

Percentile	Equal Weighting	Differential Weighting
Min	0.65	0.64
1%	0.77	0.75
5%	0.87	0.86
10%	0.92	0.91
25%	0.98	0.97
50%	1.02	1.02
75%	1.04	1.05
90%	1.06	1.07
95%	1.07	1.08
99%	1.08	1.09
Max	1.10	1.12
Mean	1.00	1.00
N	3,586	3,586

4.2 Results for Section 2k.5

Table 2k..1. Comparison of Distribution of HF Composite Measure by Scoring Method

Percentile	Absolute Scoring Index with Reliability Weights	Absolute Scoring Index with Reliablity Weights (Old Version)
Min	0.64	71.02
1%	0.75	75.75
5%	0.86	78.54
10%	0.91	79.64
25%	0.97	81.09
50%	1.02	82.21
75%	1.05	83.11
90%	1.07	83.83
95%	1.08	84.24
99%	1.09	85.07
Max	1.12	86.86
Mean	1.00	81.91
N	3,586	3867

4.3 Results for Section 2l.3

Table 2l.3.1: Comparison of Hospitals' Rates for Hospitals the Full Sample and for Hospitals Included in the Composite Caluclation

	HF1*		HF2*		HF3*		HF4*		Survival**		Readmission**	
Percentile	All Hospitals	Included Hospitals	All Hospitals	Included Hospitals	All Hospitals	Included Hospitals	All Hospitals	Included Hospitals	All Hospitals	Included Hospitals	All Hospitals	Included Hospitals
Min	0.00	9.09	0.00	0.00	0.00	0.00	0.00	9.13	6.60	6.60	17.30	17.30
1%	24.88	50.00	0.00	13.19	0.00	30.00	2.42	37.04	8.00	7.90	20.40	20.30
5%	60.20	68.56	18.00	39.81	40.00	61.69	43.14	64.06	8.90	8.90	21.70	21.60
10%	71.57	75.29	37.76	51.94	64.29	75.31	62.93	76.84	9.40	9.40	22.40	22.40
25%	82.35	83.33	61.93	66.29	86.54	89.52	84.57	89.42	10.30	10.20	23.50	23.50
50%	89.75	89.80	77.42	79.21	96.69	96.92	94.73	95.58	11.20	11.20	24.60	24.60
75%	94.81	94.39	87.54	88.10	100.00	99.75	98.03	98.26	12.20	12.20	25.90	26.00
90%	98.71	97.46	94.45	94.66	100.00	100.00	99.33	99.33	13.20	13.20	27.20	27.30
95%	100.00	99.14	97.30	97.24	100.00	100.00	99.72	99.69	13.90	14.00	28.10	28.20
99%	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	15.30	15.30	29.90	30.00
Max	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	18.20	17.30	32.40	32.40
Mean	86.31	87.49	71.21	75.26	87.85	91.32	86.90	90.98	11.28	11.24	24.74	24.75
N	4,104	3,586	4,182	3,586	4,087	3,586	4,198	3,586	3,890	3,586	3,937	3,586

Notes:
* HF1: Percent of HF Patients that Received Discharge Instructions; HF2: Percent of HF Patients with Evaluation of LVS Function; HF3: Percent of HF Patients Given ACE Inhibitor or ARB for LVSD; HF4: Percent of HF Patients Given Smoking Cessation Advice/Counseling.
** Survival: 30-day risk-adjusted survival rate; Readmission: 30-day risk-adjusted lack of readmission.

4.4 Results for Section 2h.2

Table 2h.2.1. Comparison of Distribution of Composite Measure, by Bed Size

	Bed Size			
Percentile	0-49	50-199	200-399	400+
Min	0.64	0.68	0.74	0.83
1%	0.69	0.77	0.90	0.90
5%	0.76	0.87	0.95	0.95
10%	0.80	0.92	0.97	0.97
25%	0.90	0.97	1.01	1.00
50%	0.97	1.01	1.03	1.03
75%	1.02	1.05	1.05	1.06
90%	1.05	1.07	1.07	1.07
95%	1.06	1.08	1.08	1.08
99%	1.08	1.09	1.09	1.09
Max	1.09	1.12	1.11	1.10
Mean	0.95	1.00	1.03	1.03
N	664	1,539	868	437

Table 2h.2.2. Comparison of Distribution of Composite Measure, by Ownership Type

	Ownership		
Percentile	Government	Not for Profit	For Profit
Min	0.64	0.64	0.64
1%	0.70	0.78	0.75
5%	0.77	0.89	0.86
10%	0.83	0.93	0.92
25%	0.93	0.98	0.98
50%	0.99	1.02	1.02
75%	1.03	1.05	1.05
90%	1.05	1.07	1.07
95%	1.07	1.08	1.08
99%	1.08	1.09	1.10
Max	1.10	1.11	1.12
Mean	0.96	1.01	1.00
N	659	2,257	592

Table 2h.2.3. Comparison of Distribution of Composite Measure, by Teaching Hospital Status

	Teaching Hospital	
Percentile	Yes	No
Min	0.88	0.64
1%	0.90	0.74
5%	0.97	0.85
10%	0.98	0.91
25%	1.01	0.97
50%	1.03	1.01
75%	1.06	1.05
90%	1.07	1.07
95%	1.08	1.08
99%	1.09	1.09
Max	1.10	1.12
Mean	1.03	1.00
N	268	3,240

Table 2h.2.4. Comparison of Distribution of Composite Measure, by Census Region

	Census Region			
Percentile	Northeast	South	Midwest	West
Min	0.72	0.64	0.71	0.68
1%	0.77	0.73	0.75	0.74
5%	0.93	0.83	0.86	0.86
10%	0.97	0.89	0.91	0.91
25%	1.00	0.96	0.98	0.96
50%	1.03	1.01	1.02	1.01
75%	1.05	1.04	1.05	1.05
90%	1.07	1.07	1.07	1.07
95%	1.08	1.08	1.08	1.08
99%	1.09	1.09	1.09	1.09
Max	1.11	1.12	1.10	1.11
Mean	1.02	0.99	1.00	1.00
N	547	1,424	920	587

Table 2h.2.5. Comparison of Distribution of Composite Measure, by Percentage of Patients that are Black

	Percentage of Black Patients*			
Percentile	0	>0 and ≤15	>15 and ≤30	>30
Min	0.64	0.64	0.69	0.69
1%	0.72	0.76	0.74	0.75
5%	0.78	0.88	0.87	0.86
10%	0.85	0.93	0.92	0.91
25%	0.93	0.98	0.98	0.98
50%	0.99	1.02	1.02	1.02
75%	1.03	1.05	1.05	1.04
90%	1.06	1.07	1.07	1.06
95%	1.07	1.08	1.08	1.07
99%	1.08	1.09	1.09	1.09
Max	1.09	1.11	1.12	1.09
Mean	0.97	1.01	1.00	1.00
N	602	1,960	488	536

The National Quality Forum
Composite Measure of Hospital Quality for HF

Appendix B

Submitted By:

Boston University
Jim Burgess, Ph.D.
Gary Young, Ph.D.

Prepared for:
Centers for Medicare and Medicaid Services

March 2011

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.