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:** Payment Program	**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 Subcommittee. Heart disease and stroke statistics—2010 update: a report from the American Heart Association. Circulation. 2010;121:e46-e215.	**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).

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 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

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]

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).

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.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.

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

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.

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

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
•Patients with comfort measures only documented
•Reasons for no LVS function evaluation documented by a physician, advanced practice nurse, or physician assistant

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 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)*

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)*
Population : National

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

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 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

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.

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 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

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.

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:
.
http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1138115987129

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.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 implemented for performance measurement. (evaluation criteria)	**Eval Rating**
4a. Data Generated as a Byproduct of Care Processes	**4a**

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?** 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)	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:** 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	**4e** 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.

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).

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).

Process of Care Performance Measures and Long-Term Outcomes in Patients Hospitalized With Heart Failure

Mark E. Patterson, MPH, PhD, Adrian F. Hernandez, MD, MHS,†‡ Bradley G. Hammill, MS,* Gregg C. Fonarow, MD,§ Eric D. Peterson, MD, MPH,†‡ Kevin A. Schulman, MD,*‡ and Lesley H. Curtis, PhD*‡*

Background: Recent efforts to improve care for patients hospitalized with heart failure have focused on process-based performance measures. Data supporting the link between current process measures and patient outcomes are sparse.

Objective: To examine the relationship between adherence to hospital-level process measures and long-term patient-level mortality and readmission.

Research Design: Analysis of data from a national clinical registry linked to outcome data from the Centers for Medicare and Medicaid Services (CMS).

Subjects: A total of 22,750 Medicare fee-for-service beneficiaries enrolled in the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure between March 2003 and December 2004.

Measures: Mortality at 1 year; cardiovascular readmission at 1 year; and adherence to hospital-level process measures, including discharge instructions, assessment of left ventricular function, prescription of angiotensin-converting enzyme inhibitor or angiotensin receptor blocker at discharge, prescription of beta-blockers at discharge, and smoking cessation counseling for eligible patients.

Results: Hospital conformity rates ranged from 52% to 86% across the 5 process measures. Unadjusted overall 1-year mortality and cardiovascular readmission rates were 33% and 40%, respectively. In covariate-adjusted analyses, the CMS composite score was not associated with 1-year mortality (hazard ratio, 1.00; 95% confidence interval, 0.98–1.03; $P = 0.91$) or readmission (hazard ratio, 1.01; 95% confidence interval, 0.99–1.04; $P = 0.37$). Current CMS process measures were not independently associated with mortality, though prescription of beta-blockers at discharge was independently associated with lower mortality (hazard ratio, 0.94; 95% confidence interval, 0.90–098; $P = 0.004$).

Conclusion: Hospital process performance for heart failure as judged by current CMS measures is not associated with patient outcomes within 1 year of discharge, calling into question whether existing CMS metrics can accurately discriminate hospital quality of care for heart failure.

Key Words: heart failure, mortality, outcome and process assessment (health care), patient readmission

(*Med Care* 2010;48: 210–216)

Substantial variation exists in the provision of evidence-based, guideline-recommended care to patients hospitalized for heart failure in the United States.[1] Recent efforts to improve the quality of care for these patients have focused on process-based performance measures. The Centers for Medicare and Medicaid Services (CMS) and the Joint Commission have designated 4 such process measures, and the American Heart Association (AHA) and the American College of Cardiology (ACC) have designated 5 discharge measures for heart failure (the 4 CMS measures plus anticoagulation for atrial fibrillation).[2] Medicare and other payers use such measures in pay-for-performance programs and report the measures publicly on the Hospital Compare Web site to help patients select high-quality providers. Central to these programs is the implicit assumption that conformance with process measures improves patient outcomes. However, data supporting the process-outcome link are sparse.

Previous studies have examined associations between hospital-level performance and hospital-level outcomes[3–5] and associations between patient-level adherence to process measures and patient-level outcomes.[5] Hospital-level analyses have found no association between hospital-level adherence and 30-day mortality.[3] Patient-level analyses suggest that adherence to certain process measures is strongly associated with 60- to 90-day postdischarge outcomes and that adherence to other process measures is not.[5] These types of analyses do not address an important question from the

From the *Center for Clinical and Genetic Economics, †Duke Clinical Research Institute, and ‡Department of Medicine, Duke University School of Medicine, Durham, NC; and §Ahmanson-UCLA Cardiomyopathy Center, Department of Medicine, UCLA Medical Center, Los Angeles, CA.

Supported by the Agency for Healthcare Research and Quality (grant U18HS10548); and by GlaxoSmithKline. Also, supported by the American Heart Association Pharmaceutical Roundtable grant (0675060N) (to A.F.H.), by the National Institute on Aging (grant R01AG026038) (to L.H.C. and K.A.S.), and by the Ahmanson Foundation and the Corday Family Foundation (to G.C.F.).

Trial Registration clinicaltrials.gov Identifier: NCT00344513.

Reprints: Lesley H. Curtis, PhD, Center for Clinical and Genetic Economics, Duke Clinical Research Institute, PO Box 17969, Durham, NC 27715. E-mail: lesley.curtis@duke.edu.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.lww-medicalcare.com).

ISSN: 0025-7079/10/4803-0210

patient's perspective: Are hospital-level performance measures important indicators of long-term patient outcomes? That is, is receiving care at a hospital with better conformity to recommended processes of care associated with better long-term outcomes for patients with heart failure?

Using data from the Organized Program to Initiate Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF) registry linked to Medicare claims data, we examined the relationship between adherence to hospital-level process measures and patient-level mortality and readmission in the first year after discharge.

METHODS

Data Sources

Patients in this study were from the OPTIMIZE-HF registry, which has been described in detail previously.[5–7] The registry was established to collect data regarding processes of care for patients hospitalized with heart failure. The 259 participating US hospitals enrolled 48,612 patients from March 1, 2003 to December 31, 2004, and used a case ascertainment approach similar to that used by the Joint Commission.[8] Patients were eligible for the registry if (*a*) they presented with symptoms of heart failure during a hospitalization for which heart failure was the primary discharge diagnosis or (*b*) the primary reason for admission was an episode of worsening heart failure. The *International Classification of Diseases, Ninth Revision, Clinical Modification* codes used as enrollment criteria for OPTIMIZE-HF and case finding were identical to those used by CMS. Patients from all geographic regions of the United States were included and a variety of institutions participated, from community hospitals to large tertiary centers. Each center's institutional review board or a central institutional review board approved the study protocol. Hospital staff used a Web-based case report form to record patient-level information, including demographic characteristics, comorbid conditions, vital signs, and drug therapy. Automatic electronic data checks prevented out-of-range entries and duplications. In addition, an audit of the database based on predetermined criteria verified data against source documents for a 5% random sample of the first 10,000 patients.

For this study, we merged patient data from the OPTIMIZE-HF registry with Medicare Part A inpatient claims, matching by date of birth, sex, hospital, and admission and discharge dates.[9] Of 36,165 hospitalizations of patients aged 65 years or older, 29,301 (80.8%) were matched to Medicare claims, representing 25,901 distinct patients. We excluded 1218 patients who died before discharge, 1143 patients who were ineligible for any of the 4 process measures, and 790 patients in 88 hospitals with fewer than 25 eligible patients, a convention used in previous studies to improve the stability of process measure estimates.[3] The final data set contained data on 22,750 patients from 150 hospitals. In addition to the overall cohort, we created 4 separate cohorts of patients who were eligible for each of the 4 process measures of interest. These cohorts included only data from hospitals with at least 25 eligible patients for a given process measure.

Process Measures

We analyzed a total of 5 process measures. These included the 4 process measures endorsed by CMS, the Joint Commission, and the ACC/AHA: (*a*) discharge instructions that address diet, exercise, medications, and relevant follow-up care for patients discharged to home; (*b*) assessment of left ventricular function; (*c*) prescription of an angiotensin-converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB) at discharge to eligible patients with left ventricular systolic dysfunction without contraindications; and (*d*) smoking cessation counseling for patients who had smoked within 1 year of admission. In addition, we analyzed prescription of beta-blockers at discharge to eligible patients with left ventricular systolic dysfunction without contraindications. Although not endorsed by CMS, this process measure has been shown to be associated with improvement in short-term outcomes.[5,10]

We constructed the performance measures by using the numerator and denominator definitions in the Joint Commission ORYX specifications; that is, we assessed use among eligible patients without documented contraindications, intolerance, or other physician documentation.[8] Patients who died, were transferred to another acute care hospital, were discharged to hospice or a federal hospital, or left against medical advice were considered ineligible to receive any of the 5 processes of care.[8] We summarized each process measure at the hospital level by dividing the number of patients for whom the process measure was documented by the number of patients eligible for the measure. In patient-level analyses, we applied hospital-level adherence rates uniformly to all patients within a given hospital; thus, the hospital-level rates can be considered continuous measures of hospital quality.

For each hospital, we constructed 2 overall scores. First, we constructed a composite score by dividing the total number of documented CMS-endorsed processes of care by the total number of opportunities to provide those processes of care, a score similar to that currently used in Medicare's Hospital Compare as a basis for pay-for-performance programs for the 4 CMS measures.[8,11] For example, a patient who received 2 of 4 processes of care for which she was eligible would contribute 2 to the numerator of the composite score and 4 to the denominator. The composite score indicates how often patients in a given hospital received the processes of care for which they were eligible. Second, we constructed a "defect-free" score to indicate the proportion of patients in the hospital who received all of the CMS-endorsed processes of care for which they were eligible. In this case, the patient from the previous example would contribute 0 to the numerator and 1 to the denominator, because she did not receive all of the processes of care.[12,13]

The main outcome measure was mortality within 1 year after hospital discharge. We also analyzed cardiovascular readmission within 1 year after discharge. We obtained dates of death from CMS data through December 31, 2006. We defined cardiovascular readmission as the first subsequent inpatient admission for a cardiovascular reason as identified in Medicare Part A claims and defined by diagnosis related group codes 104 to 112, 115 to 118, 121 to 125, 127 to 145, 476, 514

to 518, 525 to 527, 535 to 536, and 547 to 558, excluding transfers or subsequent admissions for rehabilitation.

Covariates

Baseline patient-level covariates from the OPTIMIZE-HF registry included age, race, history of acute myocardial infarction, diabetes mellitus, prior cerebrovascular disease, peripheral vascular disease, depression, hyperlipidemia, chronic obstructive pulmonary disease, and atrial arrhythmia; and mean serum creatinine, hemoglobin, systolic and diastolic blood pressure, and weight at admission. Between 1% and 6% of the patients had missing values for creatinine, hemoglobin, systolic and diastolic blood pressure, and weight. We imputed the mean values of the overall cohort for these missing values. From the CMS data, we calculated the total number of heart failure hospitalizations for each hospital and heart failure hospitalizations as a percentage of total hospital discharges and included these as hospital-level covariates.

Statistical Analysis

We calculated frequencies and means for baseline demographic characteristics, comorbid conditions, and clinical characteristics for the full sample of 22,750 patients, and hospital-level volume and performance measures for the 150 hospitals. We present Kaplan-Meier estimates of unadjusted mortality, and we calculated unadjusted cardiovascular readmission rates using the cumulative incidence function.[14] In the primary analysis, we examined the relationship between hospital-level adherence and patient-level outcomes. Specifically, we used Cox proportional hazards models to estimate the unadjusted and adjusted effects of each hospital-level process measure on mortality and cardiovascular readmission. The multivariable models included the patient-level and hospital-level covariates described above. To account for the clustering of patients within hospitals, we calculated robust standard errors.[15] We performed 2 sensitivity analyses. First, we relaxed the requirement for eligible patients per hospital from 25 to 10. Second, to assess the need for random effects, we modeled the mortality end point using a generalized linear model with a logit link and binomial variance function and specified site-level random intercepts.

To address the question of whether higher-performing hospitals have lower 1-year risk-adjusted mortality rates compared with lower-performing hospitals, we estimated the relationship between hospital-level process measures and hospital-level risk-adjusted outcomes using a bootstrap approach. For each patient, we first calculated predicted probabilities of mortality and cardiovascular readmission, based on regression models that included the baseline patient-level covariates listed above. We then drew 1000 samples (with replacement) of 22,750 patients from the data used in the main analysis. For each sample, we calculated the hospital-level conformity rates and hospital-level risk-adjusted outcome rates. Conformity rates were calculated as previously described. Risk-adjusted outcome rates were calculated by dividing the observed outcome rate by the expected outcome rate and multiplying this quantity by the observed 1-year outcome rate in the overall sample. In each sample, we regressed these hospital-level risk-adjusted mortality and readmission rates on each of the hospital-level process measures. For each outcome, the mean of all parameter estimates is reported for each process measure. To address statistical significance, we provide the 95th bootstrap percentile interval. We used SAS software version 9.1 (SAS Institute Inc, Cary, NC) for all analyses.

RESULTS

The mean age of the overall cohort was 79 years, 44% were men, and 83% were white. Approximately one-quarter of the patients had a history of acute myocardial infarction or non-insulin-dependent diabetes mellitus, and almost one-third had a history of hyperlipidemia or chronic obstructive pulmonary disease (Table 1). Unadjusted overall 1-year mortality and cardiovascular readmission rates were 33% and 40%, respectively.

The median number of patients with heart failure treated annually at each hospital was 227 (interquartile range, 136–381). Mean hospital-level adherence rates for individual process measures varied considerably. On average, hospitals assessed left ventricular function in 86% of eligible patients but provided discharge instructions to only 52% of eligible patients. The mean hospital-level composite score, which indicates the proportion of CMS-endorsed care processes that were correctly provided, was 72%. The defect-free measure, which indicates the proportion of patients receiving all of the CMS-endorsed processes of care for which they were eligible, was 51% (Table 2). When applied uniformly to all patients in a

TABLE 1. Baseline Patient Characteristics (N = 22,750)

Characteristic	Patients
Age, mean ± SD, yr	79.4 ± 7.8
Male sex, number (%)	9986 (43.9)
Race, number (%)	
Black	2451 (10.8)
White	18821 (82.7)
Other	1478 (6.5)
Medical history, number (%)	
Atrial arrhythmia	8189 (36.0)
Hyperlipidemia	7577 (33.3)
Chronic obstructive pulmonary disease	6492 (28.5)
Non-insulin-dependent diabetes mellitus	5618 (24.7)
Acute myocardial infarction	5183 (22.8)
Prior cerebrovascular accident or transient ischemic attack	3930 (17.3)
Peripheral vascular disease	3390 (14.9)
Insulin-dependent diabetes mellitus	3371 (14.8)
Depression	2346 (10.3)
Clinical characteristics at admission	
Serum creatinine, mean ± SD, mg/dL	1.6 ± 1.2
Hemoglobin, mean ± SD, g/dL	12.0 ± 1.9
Systolic blood pressure, mean ± SD, mm Hg	143.3 ± 31.5
Diastolic blood pressure, mean ± SD, mm Hg	74.6 ± 17.9
Weight, median (interquartile range), kg	76.1 (63.4–88.0)

TABLE 2. Hospital-Level Process Measure Adherence

	Hospitals, n	Opportunities per Hospital, Mean ± SD	Processes per Hospital, Mean ± SD	Adherence, Mean ± SD	Adherence, Median (IQR)
Discharge instructions	139	121 (103)	65 (67)	0.52 (0.29)	0.55 (0.28–0.77)
Assessment of left ventricular function	150	149 (129)	127 (107)	0.86 (0.11)	0.88 (0.80–0.95)
ACE inhibitor or ARB at discharge	97	62 (38)	49 (32)	0.80 (0.11)	0.81 (0.70–0.89)
Smoking cessation counseling	25	40 (14)	24 (12)	0.61 (0.24)	0.59 (0.46–0.83)
Beta-blocker at discharge	101	67 (42)	54 (35)	0.81 (0.11)	0.82 (0.74–0.89)
Composite score*	150	320 (281)	232 (203)	0.72 (0.15)	0.73 (0.62–0.84)
Defect-free score*	150	152 (133)	79 (77)	0.51 (0.22)	0.51 (0.32–0.70)

*Scores include CMS-endorsed measures only (ie, discharge instructions, assessment of left ventricular function, ACE inhibitor or ARB at discharge, and smoking cessation counseling).

IQR indicates interquartile range; ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker.

TABLE 3. Relationship Between Process Measures and 1-Year Outcomes

	Mortality				Cardiovascular Readmission			
Process Measure	Unadjusted HR (95% CI)*	*P*	Adjusted HR (95% CI)*	*P*	Unadjusted HR (95% CI)*	*P*	Adjusted HR (95% CI)*	*P*
Discharge instructions (n = 16,791)	1.00 (0.99–1.02)	0.58	1.00 (0.98–1.01)	0.71	1.00 (0.99–1.02)	0.70	1.00 (0.99–1.02)	0.75
Assessment of left ventricular function (n = 22,297)	0.99 (0.96–1.03)	0.70	0.99 (0.96–1.03)	0.78	1.03 (0.99–1.07)	0.12	1.04 (1.01–1.08)	0.02
ACE inhibitor or ARB at discharge (n = 6044)	0.95 (0.90–1.00)	0.07	0.97 (0.93–1.02)	0.21	0.96 (0.91–1.01)	0.11	0.97 (0.92–1.01)	0.14
Smoking cessation counseling (n = 1008)	0.98 (0.95–1.02)	0.35	0.96 (0.92–1.01)	0.13	0.99 (0.96–1.03)	0.67	0.99 (0.94–1.04)	0.63
Beta-blocker at discharge (n = 6597)	0.95 (0.90–1.00)	0.03	0.94 (0.90–0.98)	0.004	0.98 (0.93–1.03)	0.47	0.97 (0.92–1.02)	0.21
Composite overall score (n = 22,750)†	1.01 (0.98–1.03)	0.64	1.00 (0.98–1.03)	0.91	1.01 (0.98–1.04)	0.44	1.01 (0.99–1.04)	0.37
Defect-free score (n = 22,750)†	1.01 (0.99–1.03)	0.40	1.00 (0.98–1.02)	0.86	1.00 (0.98–1.02)	0.69	1.00 (0.99–1.02)	0.65

*Hazard ratios estimate the risk of outcome dependent upon a 10% increase in hospital-level adherence.

†Scores include CMS-endorsed measures only (ie, discharge instructions, assessment of left ventricular function, ACE inhibitor or ARB at discharge, and smoking cessation counseling).

HR indicates hazard ratio; CI, confidence interval; ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker.

given hospital, the resulting distributions of adherence rates were similar (data not shown).

Hospital-level adherence to CMS-endorsed process measures including discharge instructions, assessment of left ventricular function, prescription of an ACE inhibitor or ARB, and smoking cessation counseling was not associated with lower patient-level mortality at 1 year in the adjusted analyses. Estimated effect sizes for these process measures were small. For each 10% incremental increase in hospital-level adherence, no process measure reduced the odds of mortality by more than 4%. Hospital-level prescription of beta-blockers at discharge was significantly associated with patient-level mortality. A 10% incremental increase in hospital-level adherence was associated with 6% lower odds of mortality. Neither the CMS composite measure nor the defect-free measure was significantly related to patient-level mortality (Table 3). Similar to the mortality analyses, most of the process or composite measures were not associated with 1-year cardiovascular readmission after adjustment, with the exception of assessment of left ventricular function. A 10% increase in hospital-level adherence to the assessment of left ventricular function was associated with a 4% increase in the odds of cardiovascular readmission at 1 year.

In the first sensitivity analysis, we relaxed the requirement for eligible patients per hospital from 25 to 10. Using this criterion, the sample increased to 188 hospitals and 23,318 patients (smoking cessation at 76 hospitals; ACE inhibitor or ARB at 140 hospitals). Although most findings were unchanged, the conformity to the ACE inhibitor/ARB measure trended toward a lower adjusted mortality rate (hazard ratio, 0.96; 95% confidence interval, 0.92–1.01; $P = 0.08$). Associations between all of the process measures and cardiovascular readmission were unchanged. In a separate sensitivity analysis, we assessed the need for random effects by fitting a hierarchical model for the mortality end point. The results corresponded almost exactly with those from the proportional hazards model with robust standard errors (Table A1, online only, Supplemental Digital Content 1, available at: http://links.lww.com/MLR/A64).

Table 4 shows the results of the bootstrap analyses. None of the hospital-level individual process measure adherence rates or composite scores was found to be significantly associated with hospital-level risk-adjusted outcomes. Effect sizes were again found to be small.

CONCLUSIONS

In this analysis of 22,750 Medicare beneficiaries hospitalized with heart failure at 150 US hospitals, we found

TABLE 4. Relationship Between Hospital-Level Process Measures and Hospital-Level Risk-Adjusted Outcomes at 1 Year

Measure	Absolute Percentage Change in Outcome Due to 10% Change in Process Measure (95% Bootstrap Percentile Interval)	
	Mortality	Cardiovascular Readmission
Discharge instructions	0.0 (−0.3 to 0.3)	−0.2 (−0.4 to 0.1)
Assessment of left ventricular function	−0.2 (−1.1 to 0.7)	0.0 (−0.8 to 0.8)
ACE inhibitor or ARB at discharge	0.1 (−0.7 to 0.8)	0.0 (−0.6 to 0.7)
Smoking cessation counseling	−0.2 (−0.8 to 0.3)	0.2 (−0.2 to 0.6)
Beta-blocker at discharge	−0.3 (−1.1 to 0.4)	−0.4 (−1.1 to 0.2)
Composite score*	0.1 (−0.5 to 0.7)	−0.1 (−0.6 to 0.3)
Defect-free score*	0.2 (−0.3 to 0.6)	−0.1 (−0.4 to 0.2)

*Scores include CMS-endorsed measures only (ie, discharge instructions, assessment of left ventricular function, ACE inhibitor or ARB at discharge, and smoking cessation counseling).

substantial variation in hospital adherence to the 4 CMS process measures. Yet, with the exception of the positive association between hospital-level conformity to the assessment of left ventricular function and cardiovascular readmission, there were no associations between the CMS hospital performance measures or the composite measures and patient-level mortality or cardiovascular readmission rates at 1 year. However, we did find a significant association between hospital-level adherence to prescription of beta-blockers at discharge and lower mortality at 1 year. To explore these associations with risk-adjusted hospital-level outcomes, we conducted bootstrap analyses and found the results to be generally consistent with the primary analysis.

These findings are generally consistent with a previous analysis examining patient-level predictors and outcomes of 5791 patients from the 91 hospitals who participated in OPTIMIZE-HF. In that study, only conformity with a measure for prescription of a beta-blocker for left ventricular systolic dysfunction was significantly associated with a lower risk of 60- to 90-day mortality after propensity adjustment and risk adjustment.[5] The findings are also consistent with a study using an administrative data source to examine associations between hospital-level processes of care and hospital-level outcomes in 3657 acute care hospitals, which found that assessment of left ventricular function and prescription of an ACE inhibitor at discharge were not significantly associated with improved survival at 1 year.[3] The absolute risk reduction in risk-adjusted mortality between hospitals performing in the 25th percentile compared with those performing in the 75th percentile was 0.002 ($P = 0.05$) for assessment of left ventricular function, −0.003 ($P = 0.04$) for ACE inhibitor use, and 0.002 ($P = 0.08$) for 1-year mortality. In contrast, a study of 2958 patients drawn from a 20-hospital health care system in a single community reported an association between CMS process measures at discharge and 1-year survival, though multiple known confounders were not included in the multivariable models and nurse case managers continued to be involved in the care of patients after discharge.[16]

The present analysis expands upon findings from previous studies in 2 key ways. First, this study links Medicare administrative data to a detailed clinical source to allow for both longitudinal outcome assessment and rigorous risk adjustment. Thus, we were able to determine whether CMS process measures for heart failure had measurable effects up to 1 year after discharge in a broad cohort of patients from all regions of the United States. In addition, the analysis examines how overall hospital adherence levels are related to patient-level mortality and cardiovascular readmission, thereby addressing the question of whether patients who are treated at hospitals that score higher on process measures have better outcomes. This analytic approach addresses whether receiving care at a hospital with better conformity to recommended processes of care is associated with improvements in long-term outcomes for patients with heart failure. Previous research from CRUSADE (Can Rapid Risk Stratification of Unstable Angina Patients Suppress Adverse Outcomes with Early Implementation of the ACC/AHA Guidelines) has also addressed the associations between hospital-level predictors and patient-level outcomes, but for patients hospitalized with acute coronary syndromes.[17–19] Although hospital profit status[17] and the presence of an inpatient cardiology service[18] were not significantly associated with inpatient outcomes, hospital participation in clinical trials[19] was significantly related to patient-level mortality.

There are several potential explanations for the lack of associations in this study. First, the processes of care selected for the performance measures may truly not be associated with outcomes. Evidence of associations between discharge instructions, assessment of left ventricular function, and smoking cessation counseling are based on expert opinion rather than randomized clinical trials. Furthermore, outcomes after hospital discharge likely reflect a combination of many domains of care and may be dominated by postdischarge care processes, frequency of follow-up, and the underlying disease process. For example, being discharged with an ACE inhibitor or ARB does not ensure that a patient will remain on therapy or that an effective dose has been prescribed, nor does it ensure that the clinical effects will be observable within 1 year. However, the significant relationship observed between beta-blockers at discharge and mortality at 1 year demonstrates that associations can be detected when they exist. Second, hospital documentation of process measures may not reflect actual care. For example, patient education may be documented in the medical record even if it was completed at discharge in a rushed or superficial manner. Conversely, physicians or nurses may instruct a patient about medications, diet, symptoms of worsening heart failure, and daily weight monitoring but may not record this in the patient's medical record. Third, the self-reported nature of the process measure forms carries the risk that hospitals purposely underreport eligible patients to inflate the process measure adherence score, a violation that was suspected but not confirmed in a study of process measure adherence in family practices in the United Kingdom.[20] Finally, studies

examining effects of system-level exposures on individual-level outcomes may be limited by the inability to control for unobserved system-level characteristics, which could result in null associations.

Other findings in this study warrant comment. First, we found a small but significant association between assessment of left ventricular function and greater risk of cardiovascular readmission. The reason for this finding is unclear; we suspect it may reflect residual confounding in which patients who are sicker in ways we did not measure may have been more likely to undergo assessment of left ventricular function and be hospitalized as compared with healthier patients. Second, the demographic characteristics of the sample are comparable to another study estimating trends in mortality among hospitalized Medicare beneficiaries with heart failure,[21] providing some evidence of how the results of the current study are generalizable to Medicare fee-for-service beneficiaries. Third, the high mortality and cardiovascular readmission rates found in this patient population indicate that this is a high-risk population that would likely benefit from improved process measure conformity in measures with a strong process-outcome link.

Our study has some limitations. First, the process-outcome association may be confounded by socioeconomic factors or other unmeasured confounders related to both health status and hospital adherence level. Second, to the extent that Medicare beneficiaries enrolled in OPTIMIZE-HF are not representative of all Medicare beneficiaries with heart failure, the results may not be generalizable. Evidence suggests, however, that Medicare beneficiaries enrolled in OPTIMIZE-HF are similar to Medicare fee-for-service beneficiaries hospitalized with heart failure in terms of baseline characteristics, survival, and all-cause readmission.[22] Third, the generalizability of the results may be further limited if participating hospitals differ from nonparticipating hospitals in ways not reflected inpatient demographic characteristics, core measures, or postdischarge outcomes. Fourth, patient eligibility for a performance measure was based on documentation in the medical record, which may not always be accurate. For example, some patients may have had undocumented contraindications or intolerances, leading to an overestimation of the number of patients eligible for the performance measure. Finally, the cross-sectional nature of the data did not allow us to assess changes in performance measure conformity and clinical outcomes over time.

Performance measures are used for public reporting of the quality of cardiovascular care at the hospital level, affecting financial payments to medical centers and individual physicians. Thus, it is essential that measures be prioritized to include those that are known to be closely associated with patient outcomes. Given the lack of associations between individual measures and a composite measure and postdischarge clinical outcomes, the use of the CMS heart failure performance measures in their current form in pay-for-performance programs may not be the most efficacious way to assess and reward quality. Although clearly stated methods have been used to develop and implement heart failure performance measures, these measures are not fulfilling their stated purpose. Consequently, additional measures with stronger process-outcome links after hospital discharge should be considered. If a documentation process at the hospital does not accurately capture the most important elements of care provided, it may be unreasonable to expect that incentives for these process measures would improve outcomes.

To our knowledge, this analysis is the first to examine how overall hospital conformity to the 4 current CMS heart failure-specific process measures is associated with individual-level, long-term outcomes in a broad cohort of patients from all regions of the United States. To build upon these results, future research is needed to refine how performance measures are created and selected. Consideration should be given to prospective validation and testing of measures, rather than the selection of measures by expert panels. Before implementing pay-for-performance broadly across all systems, the limitations of current performance measures and the differences in measure reliability across disease types, provider settings, and patient populations need to be better recognized. In addition, a minimally important difference needs to be defined before policy makers decide to implement new process measures, especially given the small effect sizes.[4] The small effect sizes may not be sufficient to justify broad policy changes, especially if the cost of such changes would not justify changes that were not clinically significant. It is essential that new process of care measures for heart failure be developed and implemented so that the quality of care can be more accurately measured and outcomes of this high-risk patient population can be improved.

REFERENCES

1. Fonarow GC, Yancy CW, Heywood JT. Adherence to heart failure quality-of-care indicators in US hospitals: analysis of the ADHERE Registry. *Arch Intern Med*. 2005;165:1469–1477.
2. Bonow RO, Bennett S, Casey DE Jr, 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) endorsed by the Heart Failure Society of America. *J Am Coll Cardiol*. 2005;46:1144–1178.
3. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA*. 2006;296:2694–2702.
4. Glickman SW, Ou FS, DeLong ER, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*. 2007;297:2373–2380.
5. Fonarow GC, Abraham WT, Albert NM, et al. Association between performance measures and clinical outcomes for patients hospitalized with heart failure. *JAMA*. 2007;297:61–70.
6. Fonarow GC, Abraham WT, Albert NM, et al. Organized program to initiate lifesaving treatment in hospitalized patients with heart failure (OPTIMIZE-HF): rationale and design. *Am Heart J*. 2004;148:43–51.
7. Fonarow GC, Abraham WT, Albert NM, et al. Carvedilol use at discharge in patients hospitalized for heart failure is associated with improved survival: an analysis from organized program to initiate lifesaving treatment in hospitalized patients with heart failure (OPTIMIZE-HF). *Am Heart J*. 2007;153:82.e1–82.e11.
8. Joint Commission on Accreditation of Healthcare Organizations. *Specification Manual for National Implementation of Hospital Core Measures*. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations; 2004.
9. Hammill BG, Hernandez AF, Peterson ED, et al. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. *Am Heart J*. 2009;157:995–1000.
10. Fonarow GC, Abraham WT, Albert NM, et al. Prospective evaluation of

beta-blocker use at the time of hospital discharge as a heart failure performance measure: results from OPTIMIZE-HF. *J Card Fail*. 2007; 13:722–731.
11. Hospital Compare. Available at: http://www.hospitalcompare.hhs.gov. Accessed April 15, 2008.
12. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA*. 2006;295:1168–1170.
13. Hiratzka LF, Eagle KA, Liang L, et al. ; Get With the Guidelines Steering Committee. Atherosclerosis secondary prevention performance measures after coronary bypass graft surgery compared with percutaneous catheter intervention and nonintervention patients in the Get With the Guidelines database. *Circulation*. 2007;116(suppl 11):I207–I212.
14. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York, NY: Wiley; 1980.
15. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *J Am Stat Assoc*. 1989;84:1074–1078.
16. Kfoury AG, French TK, Horne BD, et al. Incremental survival benefit with adherence to standardized. *J Card Fail*. 2008;14:95–102.
17. Shah BR, Glickman SW, Liang L, et al. The impact of for-profit health status on the care and outcomes of patients with non-ST segment elevation myocardial infarction: results from the CRUSADE initiative. *J Am Coll Cardiol*. 2007;50:1462–1468.
18. Roe MT, Chen AY, Mehta RH, et al. Influence of inpatient service specialty on care processes and outcomes for patients with non ST-segment elevation acute coronary syndromes. *Circulation*. 2007;116: 1153–1161.
19. Majumdar SR, Roe MT, Peterson ED, et al. Better outcomes for patients treated at hospitals that participate in clinical trials. *Arch Intern Med*. 2008;168:657–662.
20. Doran T, Fullwood C, Gravelle H, et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med*. 2006;355: 375–384.
21. Curtis LH, Greiner MA, Hammill BG, et al. Early and long-term outcomes of heart failure in elderly persons, 2001–2005. *Arch Intern Med*. 2008;168:2481–2488.
22. Curtis LH, Greiner MA, Hammill BG, et al. Representativeness of a national heart failure quality-of-care registry: comparison of Medicare patients in the OPTIMIZE-HF registry with non-OPTIMIZE-HF Medicare patients. *Circ Cardiovasc Qual Outcomes*. 2009;2:377–384.

