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 <mark>yellow highlighted</mark> 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.

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

Steering Committee: Complete all <mark>pink</mark> 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.

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

(for NQF staff use) **NQF Review #:** 0229 **NQF Project:** Cardiovascular Endorsement Maintenance 2010	
MEASURE DESCRIPTIVE INFORMATION	
De.1 Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older	
De.2 Brief description of measure*:* The measure estimates a hospital-level risk-standardized mortality rate (RSMR), defined as death from any cause within 30 days after the index admission date, for patients 18 and older discharged from the hospital with a principal diagnosis of HF.	
1.1-2 Type of Measure: Outcome **De.3 If included in a composite or paired with another measure, please identify composite or paired measure** This measure is paired with a measure of hospital-level, all-cause, 30-day, risk-standardized readmission rate (RSRR) following an HF hospitalization.	
De.4 National Priority Partners Priority Area: Safety **De.5 IOM Quality Domain**: Effectiveness, Patient-centered, Safety **De.6 Consumer Care Need:** Getting better	

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

B. The measure owner/steward verifies there is an identified responsible entity and process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every 3 years. Yes, information provided in contact section	**B** Y ☐ N ☐
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ▶**Purpose:** Public Reporting, Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	**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.1Testing:** 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, High resource use, Severity of illness, Patient/societal consequences of poor quality **1a.2** **1a.3 Summary of Evidence of High Impact:** HF incidence approaches 10 per 1000 population after 65 years of age (NHLBI 2007), and is the most common discharge diagnosis among the elderly (Jessup and Brozena 2003); prevalence of HF in the U.S. is estimated at nearly 6 million. (Lloyd-Jones 2009), and is suspected as the leading cause of death in people over age 65. Many current hospital interventions are known to decrease the risk of death within 30 days of hospital admission (Jha 2007). Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals. **1a.4 Citations for Evidence of High Impact:** Jessup M, Brozena S. Medical progress: heart failure. N Engl J Med 2003;348:2007-18. National Heart, Lung, and Blood Institute. Unpublished tabulation of NHANES, 1971-1975, 1976-1980, 1988-1994, 1999-2002, 2003-2006, and extrapolation to the U.S. population, 2007.	**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).

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

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

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

1b. Opportunity for Improvement

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

1b.2 Summary of data demonstrating performance gap (variation or overall poor performance) across providers:
Recent analyses of Medicare FFS data show substantial variation in HF RSMRs among hospitals. For the most recently reported three years of data (7/2006-6/2009) the mean hospital RSMR was 10.8% with a range of 6.6% to 18.2%. The 5th percentile was 8.4% and the 95th percentile was 13.4%. The interquartile range was 9.9% to 11.7%.

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

1b.3 Citations for data on performance gap:
The information on the performance gap is based on RSMRs calculated for HF hospitalizations among Medicare FFS patients aged 65 and over (65+) from July 1, 2006- June 30, 2009 and includes 1,096,751 hospitalizations from 4,743 hospitals. The index hospitalizations are those included in the measure and reported in the 2010 update to Hospital Compare.

1b.4 Summary of Data on disparities by population group:
The measure is a hospital-level measure and therefore CMS assessed evidence of disparities by examining hospital performance based on the proportion of African-American patients or the proportion of low-income patients served by a hospital.

The analyses analyses of Medicare FFS data examining the proportion of patients that a hospital served who are African-American show slightly better performance on RSMR for hospitals with higher proportions of African-American patients, but that the range of performance across all levels is similar. We divided hospitals into deciles based on the proportions of their patients that were African-American and looked at hospital performance on the measures across deciles. The combined lowest 5 deciles of hospitals include hospitals that have fewer than 5% African-American patients and have a median HF RSMR of 11.3% (range 6.4%- 19.4%). In comparison, hospitals in the highest decile with >25% African American patients have a median HF RSMR of 10.5% (range 6.7%-15.1%). These analyses demonstrate wide variation in hospital performance regardless of the proportion of minority patients and suggest that hospitals with large proportions of African American patients are not consistently performing at a lower or higher level than other hospitals.

1b
C ☐
P ☐
M ☐
N ☐

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

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

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

Overall these analyses provide little compelling evidence of clinically significant disparities at the hospital level.

1b.5 Citations for data on Disparities:
The sample for the above analyses is from a similar 3-year cohort of Medicare FFS hospitalizations as the data for the performance gap analysis above (January 2006- December 2008) but limited to hospitals with at least 25 HF cases over the 3-year period, a total of 4,175 hospitals.

1c. Outcome or Evidence to Support Measure Focus

1c.1 Relationship to Outcomes (*For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population*): This measure calculates hospital-level, 30-day all-cause mortality rates after hospitalization for an HF. The goal is to directly affect patient outcomes by measuring risk-standardized rates of mortality.

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

1c.4 Summary of Evidence (*as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome*):
Numerous studies have demonstrated that appropriate and timely treatment for HF patients can reduce the risk of mortality within 30 days of hospital admission. (Hunt 2009, Jha 2007) Additionally, trials of interventions which improve patient education upon discharge have been shown to improve survival for HF patients. (Mcalister 2001) Current process-based performance measures, however, cannot capture all the ways that care within the hospital might influence outcomes. As a result, many stakeholders, including patient organizations, are interested in outcomes measures that allow patients and providers to assess relative outcomes performance for hospitals.

References:

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

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

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

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

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

1c.7 Summary of Controversy/Contradictory Evidence: Use of Hierarchical Generalized Linear Modeling Hierarchical modeling is the appropriate statistical approach for hospital outcomes measures given the structure of the data and the underlying assumption of such measures, which is that hospital quality of care influences 30-day mortality rates. However, CMS frequently receives comments and questions about this

1c
C ☐
P ☐
M ☐
N ☐

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

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

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

approach, so we are concisely reiterating the rationale for and merits of using hierarchiacal logistic regression. Patients are clustered within hospitals and, as such, have a shared exposure to the hospital quality and processes. The use of hierarchical modeling accounts for the clustering of patients within hospitals. Second, hierarchical models distinguish within-hospital variation and between-hospital variation to estimate the hospital's contribution to the risk of readmission. This allows for an estimation of the hospital's influence on patient outcomes. Finally, within hierarchical models we can account for both differences in case mix and sample size to fairly profile hospital performance. If we did not use hierarchical modeling we could overestimate variation and potentially misclassify hospitals' performance. Accurately estimating variation is an important objective for models used in public reporting and potentially used in value-based purchasing programs.

Effect of Patient Preferences Regarding End of Life Care
Some stakeholders have expressed concerns that our measure cannot adequately exclude patients who choose comfort measures or palliative care during their index hospitalization. Stakeholders are concerned that this could lead to unintended consequences, such as prolonging lives against patient wishes. To address these issues CMS has taken the following steps when applying the measure to the Medicare FFS population aged 65 years or older:

(1) We have added an exclusion for patients who are enrolled in hospice prior to, or on the day of, admission.

(2) We chose not to exclude patients who are discharged to hospice or seek a palliative care consult during admission to account for the fact that the choice of palliative/comfort care may be the result of poor care.

(3) To account for risk-factors associated with the end of life we include markers of frailty within our risk-adjustment variables, including: protein-calorie malnutrition, dementia or senility, and hemiplegia, paraplegia, paralysis and functional disability.

(4) Although CMS is confident in the current model, CMS will further consider clinical and measurement issues for patients for whom survival is not an objective.

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

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

1c.10 Clinical Practice Guideline Citation: N/A
1c.11 National Guideline Clearinghouse or other URL: N/A

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

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

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

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Importance to Measure and Report?*	1
Steering Committee: Was the threshold criterion, *Importance to Measure and Report*, met? Rationale:	1 Y☐ N☐
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	

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.

Extent to which the measure, <u>as specified</u>, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ([evaluation criteria](#))	<u>Eval Ratin g</u>

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

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

2a.1 Numerator Statement (*Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome*):
This outcome measure does not have a traditional numerator and denominator like a core process measure (e.g., percentage of adult patients with diabetes aged 18-75 years receiving one or more hemoglobin A1c tests per year); thus, we are using this field to define the outcome.

The outcome for this measure is 30-day all-cause mortality. We define mortality as death from any cause within 30 days of the index admission date for patients 18 and older discharged from the hospital with a principal diagnosis of HF.

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

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

2a.4 Denominator Statement (*Brief, text description of the denominator - target population being measured*):
Note: This outcome measure does not have a traditional numerator and denominator like a core process measure; thus, we are using this field to define the patient cohort and to define exclusions to the patient cohort.
This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 18 years or older. While the measure can be applied to populations aged 18 years or older, nationally data are often only available for patients aged 65 years or older. We have explicitly tested the measure in both age groups.

The cohort includes admissions for patients discharged from the hospital with a principal diagnosis of HF (ICD-9-CM codes 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, and 428.xx) and with a complete claims history for the 12 months prior to admission. Patients who are transferred from one acute care facility to another must have a principal discharge diagnosis of HF at both hospitals. The initial hospital for a transferred patient is designated as the responsible institution for the episode.

If a patient has more than one HF admission in a year, one hospitalization is randomly selected for inclusion in the measure.

2a.5 Target population gender: Female, Male
2a.6 Target population age range: The target population is age 18 years or older

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator*):
This measure was developed with 12 months of data. Currently the measure is publicly-reported with three years of index hospitalizations.

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions*):
The denominator includes patients aged 18 and older admitted to non-federal acute care hospitals for an HF defined by a principal discharge diagnosis of (ICD-9-CM codes 402.01, 402.11, 402.91, 404.01, 404.03,

2a-spec
s
C ☐
P ☐
M ☐
N ☐

404.11, 404.13, 404.91, 404.93, and 428.xx) and with a complete claims history for the 12 months prior to admission.

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

2a.9 Denominator Exclusions (*Brief text description of exclusions from the target population*): For all cohorts, the measure excludes admissions for patients:
• who were discharged on the day of admission or the following day and did not die or get transferred (because it is less likely they had a significant HF diagnosis);
• who were transferred from another acute care hospital (because the death is attributed to the hospital where the patient was initially admitted);
• with inconsistent or unknown mortality status or other unreliable data (e.g. date of death precedes admission date);
• who were discharged alive and against medical advice (AMA) (because providers did not have the opportunity to deliver full care and prepare the patient for discharge);
• that were not the first hospitalization in the 30 days prior to a patient's death. We use this criteria to prevent attribution of a death to two admissions.

For Medicare FFS patients, the measure additionally excludes admissions for patients:
• enrolled in the Medicare Hospice program any time in the 12 months prior to the index hospitalization including the first day of the index admission (since it is likely these patients are continuing to seek comfort measures only). Although this exclusion currently applies to Medicare FFS patients, it could be expanded to include all payer data if an acceptable method for identifying hospice patients outside of Medicare becomes available.

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

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

2a.11 Stratification Details/Variables (*All information required to stratify the measure including the stratification variables, all codes, logic, and definitions*):
Results of this measure will not be stratified.

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

2a.14 Risk Adjustment Methodology/Variables (*List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method*):
Our approach to risk adjustment was tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al. 2006).

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

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

The final set of risk-adjustment variables is:

Demographic

• Age-65 (years above 65, continuous) for 65 and over cohorts; or Age (years, continuous) for 18 and over cohorts
•   Male

Cardiovascular
•   History of PTCA
•   History of CABG
•   Congestive heart failure
•   Acute myocardial infarction
•   Unstable angina
•   Chronic atherosclerosis
•   Cardio-respiratory failure and shock
•   Valvular and rheumatic heart disease

Comorbidity
•   Hypertension
•   Stroke

- Renal failure
- Pneumonia
- Diabetes and DM complications
- Protein-calorie malnutrition
- Dementia and senility
- Hemiplegia, paraplegia, paralysis, functional disability
- Peripheral vascular disease
- Metastatic cancer, acute leukemia, and other severe cancers
- Trauma in last year
- Major psych disorders
- Chronic liver disease

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

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

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

2a.18-19 Type of Score: Rate/proportion
2a.20 Interpretation of Score: Better quality = Lower score
2a.21 Calculation Algorithm (*Describe the calculation of the measure as a flowchart or series of steps*):
The RSMR is calculated as the ratio of the number of "adjusted actual" deaths (also known as "predicted") to the number of "expected" deaths at a given hospital, multiplied by the national unadjusted mortality rate. For each hospital, the "numerator" of the ratio is the number of deaths within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the "denominator" is the number of deaths expected on the basis of the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital's performance given its case-mix to an average hospital's performance with the same case-mix. Thus a lower ratio indicates lower-than-expected mortality or better quality and a higher ratio indicates higher-than-expected mortality or worse quality.

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

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

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

2a.23 Sampling (Survey) Methodology *If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)*:
N/A –This measure is not based on a sample or survey.

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

2a.25 Data source/data collection instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*):
Two data sources were used to create the measure:
1. Medicare Part A Inpatient and Outpatient and Part B outpatient claims: This database contains claims data for fee-for service inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, and hospice care, as well as inpatient and outpatient claims for the 12 months prior to an index admission.

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

The measure was originally developed with claims data from a 1998 sample of 222,424 cases from 5,087 hospitals. The models have been maintained and re-evaluated each year since public reporting of the measures began in 2007. For details, see measure methodology and measure maintenance reports posted at http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1219069 855841

The measure was subsequently applied to California Patient Discharge Data, a large, linked all-payer database of patient hospital admissions. Records are linked by a unique patient identification number, allowing us to determine patient history from previous hospitalizations. In addition, the unique patient ID number is used to link with state vital statistics records to assess 30-day mortality.
To apply the measure to Medicare data, Medicare Part A inpatient and outpatient and Part B outpatient claims are used. To apply the measure to a non-Medicare population, inpatient claims data are used.

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

2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL N/A http://qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1219069 855841

2a.29-31 Data dictionary/code table web page URL or attachment: URL N/A Condition Category/ICD-9 Code Map available at:
http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=11 82785083979

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

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

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

TESTING/ANALYSIS

2b. Reliability testing

2b

C ☐
P ☐
M ☐
N ☐

2b.1 Data/sample (*description of data/sample and size*): The reliability of the model was tested by randomly selecting 50% of Medicare FFS patients aged 65+ in the initial one-year cohort and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients.

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.

Furthermore, in each subsequent year of measure maintenance we have re-fit the model and compared the frequencies of comorbidities and model fit across 3 years.

2b.2 Analytic Method *(type of reliability) & rationale, method for testing)*:
In measure development and testing, for all cohorts, we computed diagnostics that describe their respective performance in terms of discriminant ability, overall fit, and generated hospital RSMRs and corresponding interval estimates for the development sample.

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

2c. Validity testing

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

2c.2 Analytic Method *(type of validity) & rationale, method for testing)*:
Medical-record validation: We developed a medical record measure to compare with the administrative measure. We developed a measure cohort with the medical record data using the inclusion/exclusion criteria and risk-adjustment strategy that was consistent with the claims-based administrative measure but using chart-based risk adjusters, such as blood pressure, not available in the claims data. We then matched a sample of the same patients in the administrative data for comparison. The matched sample included 46,700 patients. We compared the output of the two measures, that is the state performance results, in the same group of patients.

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

2c
C ☐
P ☐
M ☐
N ☐

2d. Exclusions Justified

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

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

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

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

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

2d
C ☐
P ☐
M ☐
N ☐
NA ☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable   11

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

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

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

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:
•supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;
AND
•a clinically appropriate exception (e.g., contraindication) to eligibility for the measure focus;
 AND
•precisely defined and specified:
–if there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

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

2e. Risk Adjustment for Outcomes/ Resource Use Measures

2e.1 Data/sample *(description of data/sample and size)*: When applied to Medicare FFS beneficiaries, the prior year of data from Medicare Part A inpatient and outpatient data and Part B outpatient data are used to identify variables for risk-adjustment. Specifically, Medicare Part A inpatient data areused to identify variables for risk adjustment in the index admission. Part A and B outpatient data are used to identify variables for risk adjustment in the 12-month period preceding the index date of admission.

Application to Medicare FFS Beneficiaries Using Inpatient Data Only for Risk Adjustment
As part of testing the model in all-payer data, we also applied the model to CMS data for Medicare FFS 65+ patients in California hospitals using only inpatient data for risk adjustment. California is a diverse state, and, with more than 37 million residents, California represents 12% of the US population. Specifically, we created a 2006 measure cohort with complete one-year history data and 30-day follow-up data (N= 24,035).

Application to Patients Aged 18 and Older
We also applied the model to all-payer data from California. The analytic sample included 60,022 cases aged 18 and older in the 2006 California Patient Discharge Data. When used in all-payer data, only admission claims data are used for risk adjustment, as the hospital discharge databases do not have outpatient claims.

2e.2 Analytic Method *(type of risk adjustment, analysis, & rationale)*:
This measure is fully risk-adjusted using a hierarchical logistic regression model to calculate hospital RSMRs accounting for differences in hospital case-mix. (See "risk adjustment methodology" for additional details.)
Approach to assessing model performance:
During measure development, we computed five summary statistics for assessing model performance (Harrell and Shih 2001) for the development and validation cohort:
(1) over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)
(2) predictive ability
(3) area under the receiver operating characteristic (ROC) curve
(4) distribution of residuals
(5) model chi-square (A test of statistical significance usually employed for categorical data to determine whether there is a good fit between the observed data and expected values; i.e., whether the differences between observed and expected values are attributable to true differences in characteristics or instead the result of chance variation).

Application to Medicare FFS Beneficiaries Using Inpatient Data Only for Risk Adjustment
To help determine whether the measure could be applied to Medicare FFS 65+ patients using only Medicare Part A data, we performed analyses to assess how the model performs when using only admission claims data for risk adjustment, as all-payer hospital discharge databases do not have outpatient claims. To assess the validity of using only admission claims data for risk adjustment, we fit the model separately using the full data and using only admission claims data and (a) compared the odds ratios (ORs) for the various risk factors; (b) conducted a reclassification analysis to compare risk prediction at the patient level; (c) compared model performance in terms of the c-statistic (discrimination); and (d) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and $R2$) to assess whether the model with only admission claims data is different from the current model in profiling hospital rates.

Application to Patients Aged 18 and Older
To help determine whether the measure could be applied to an population of patients aged 18+, we examined the interaction terms between age (18-64 vs. 65+) and each of the other risk factors. Specifically, we fit the model in all patients 18+ with and without interaction terms and (a) conducted a reclassification analysis to compare risk prediction at the patient level; (b) compared the c-statistic; and (c) compared hospital-level risk-standardized rates (scatterplot, correlation coefficient, and $R2$) to assess whether the model with interactions is different from the current model in profiling hospital rates.

--
Reference: Harrell FE, Shih YCT. Using full probability models to compute probabilities of actual interest to decision makers. Int J Technol Assess Health Care. 2001;17:17-26.

2e
C ☐
P ☐
M ☐
N ☐
NA
☐

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:
●an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care;[Error! Bookmark not defined.] OR rationale/data support no risk adjustment.

Comment [k17]: 13 Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care such as race, socioeconomic status, gender (e.g., poorer treatment outcomes of African American men with prostate cancer, inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than adjusting out differences.

2e.3 Testing Results *(risk model performance metrics)*:
During initial measure development, using Medicare FFS beneficiaries age 65 and over, we tested the performance of the model developed in a random selected half of the 1998 hospitalizations for HF (representing 222,424 cases discharged from 5,087 hospitals) against hospitalizations from the other half(representing 222,157 cases discharged from 5,088 hospitals). The performance was not substantively different in the validation sample (ROC area = 0.70) compared with the development sample (1998). The models appear well calibrated, with over-fitting indices of (-0.0035, 0.9928).

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

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

In subsequent years, during annual measure maintenance we looked at the distributions of comorbid conditions, hospital volume, crude rates, hospital RSMR, risk-adjusted odds ratios and 95% confidence intervals, and between-hospital variance over each subsequent year since 2005 and the and the parameters have remained consistent. For the 2005-2007 and 2006-2008 calendar year datasets, we reported each individual year results as well as the 3-year combined results. Model performance was stable over all time periods.

Model Performance in Medicare FFS Beneficiaries Using Inpatient Data Only for Risk Adjustment Using CMS data for Medicare FFS 65+ beneficiaries in California hospitals: (a) the magnitude of odds ratios for most risk factors was similar when comparing the model using full data and using only admission claims data; (b) when comparing the model with full data and with only admission claims data, the reclassification analysis demonstrated good patient-level risk prediction; (c) the c-statistic was similar (0.681 vs. 0.684); and (d) hospital-level risk-standardized rates were highly correlated (r=0.993).

Model Perfomance in Patients Aged 18 and Older
When the model was applied to all patients 18 and over (18+), overall discrimination was good (c-statistic=0.718). In addition, there was good discrimination and predictive ability in both those aged 18-64 and those aged 65+. Moreover, the distribution of Pearson residuals was comparable across the patient subgroups. When comparing the model with and without interaction terms, (a) the reclassification analysis demonstrated good patient-level risk prediction (1.9% to 25.4% vs. 2.0% to 25.1%, respectively, from the bottom decile to the top decile of the prediction values); (b) the c-statistic was nearly identical (0.720 vs. 0.718); and (c) hospital-level risk-standardized rates were highly correlated (r=1.000). Thus, the inclusion of the interactions did not substantively affect either patient-level model performance or hospital-level results.

Therefore, the measure can be applied to all payer data for patients 18 and older.

References:

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

Bernheim SM, et al. 2010 Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure and Pneumonia 30-day Risk Standardized Mortality Rate. 2010 Available at: http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic/Page/QnetTier3&cid=1163010421830

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

2f. Identification of Meaningful Differences in Performance	

2f.1 Data/sample from Testing or Current Use *(description of data/sample and size)*: The data below are based on RSMRs calculated for HF hospitalizations among Medicare FFS patients aged 65+ from July 1, 2006-June 30, 2009 and includes 1,096,751 hospitalizations from 4,743 hospitals. The index hospitalizations are those included in the measure and reported in the 2010 update to Hospital Compare.

2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance *(type of analysis & rationale)*:
For each RSMR, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national crude mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare Web site as "better than the U.S. national rate" or "worse than the U.S. national rate." If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or "the difference is uncertain." CMS does not classify performance for hospitals that have fewer than 25 HF cases in the three-year period.

2f.3 Provide Measure Scores from Testing or Current Use *(description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
 Recent analyses of Medicare FFS data show substantial variation in HF RSMRs among hospitals. For the most recently reported three years of data (7/2006-6/2009) the mean hospital RSMR was 10.8% with a range of 6.6% to 18.2%. The 5th percentile was 8.4% and the 95th percentile was 13.4%. The interquartile range was 9.9% to 11.7%.

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

2f
C ☐
P ☐
M ☐
N ☐

2g. Comparability of Multiple Data Sources/Methods	

2g.1 Data/sample *(description of data/sample and size)*: The measure performs well in both Medicare FFS data and all-payer data.

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

2g.3 Testing Results *(e.g., correlation statistics, comparison of rankings)*:
See above

2g
C ☐
P ☐
M ☐
N ☐
NA
☐

2h. Disparities in Care	

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

2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans:
The analyses performed by CMS (described in section 1b) demonstrate that hospitals have similar and overlapping performance on the measure regardless of the proportion of patients of low socioeconomic status or of African-American race. Importantly, the analyses show that hospitals with high proportions of low socioeconomic status patients or high proportions of African-American patients are able to perform well on the measure. For this reason CMS does not plan to stratify the measure.

2h
C ☐
P ☐
M ☐
N ☐
NA
☐

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable   14

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

Comment [k19]: 14 With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% v. 75%) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall poor performance may not demonstrate much variability across providers.

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

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

TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Scientific Acceptability of Measure Properties*?	2
Steering Committee: Overall, to what extent was the criterion, *Scientific Acceptability of Measure Properties*, met? Rationale:	*2* C☐ P☐ M☐ N☐

3. USABILITY

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. ([evaluation criteria](#))	Eval Rating
3a. Meaningful, Understandable, and Useful Information **3a.1** Current Use: In use **3a.2** Use in a public reporting initiative (disclosure of performance results to the public at large) (*If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). If not publicly reported, state the plans to achieve public reporting within 3 years*): The measure has been publicly reported on Hospital Compare (www.hospitalcompare.hhs.gov) since June 2007 and is used in CMS' Hospital Inpatient Quality Reporting Program (Formerly RHQDAPU). **3a.3** If used in other programs/initiatives (*If used in quality improvement or other programs/initiatives, name of initiative(s), locations, Web page URL(s). If not used for QI, state the plans to achieve use for QI within 3 years*): Testing of Interpretability (*Testing that demonstrates the results are understood by the potential users for public reporting and quality improvement*) **3a.4** Data/sample (*description of data/sample and size*): **3a.5** Methods (*e.g., focus group, survey, QI project*): This measure was NQF endorsed in 2007. Prior to public reporting in 2007, CMS conducted a dry run in December 2006 to provide hospitals and the public with an opportunity to preview the measure methodology, proposed information for public reporting and hospital-specific information Additionally, CMS has also conducted consumer testing of the language on Hospital Compare to ensure clarity and ease of interpretation of the information to be posted publicly. **3a.6** Results (*qualitative and/or quantitative results and conclusions*):	3a C☐ P☐ M☐ N☐
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? Yes, the risk-adjustment strategy is similar.	3b C☐ P☐ M☐ N☐ NA ☐
3c. Distinctive or Additive Value **3c.1** Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures: The measure looks at a different condition, HF, than the AMI and pneumonia measures listed in 3b.1.	3c C☐ P☐ M☐

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

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

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

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).

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable   15

5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality: NQF #0358 Congestive Heart Failure Mortality (IQI 16). Inpatient mortality rates can be influenced by hospital length of stay, thus 30-day measures, that establish a standard follow-up period are more appropriate for profiling a diverse group of hospitals.	N ☐ NA ☐
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for *Usability?*	3
Steering Committee: Overall, to what extent was the criterion, *Usability*, met? **Rationale:**	3 C ☐ P ☐ M ☐ N ☐

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)	**Eval Rating**
4a. Data Generated as a Byproduct of Care Processes **4a.1-2 How are the data elements that are needed to compute measure scores generated?** Coding/abstraction performed by someone other than person obtaining original information (E.g., DRG, ICD-9 codes on claims, chart abstraction for quality measure or registry)	4a C ☐ P ☐ M ☐ N ☐
4b. Electronic Sources **4b.1 Are all the data elements available electronically?** (*elements that are needed to compute measure scores are in defined, computer-readable fields, e.g., electronic health record, electronic claims*) Yes **4b.2 If not, specify the near-term path to achieve electronic capture by most providers.**	4b C ☐ P ☐ M ☐ N ☐
4c. Exclusions **4c.1 Do the specified exclusions require additional data sources beyond what is required for the numerator and denominator specifications?** No **4c.2 If yes, provide justification.**	4c C ☐ P ☐ M ☐ N ☐ NA ☐
4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences **4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results.** Using administrative claims variables for risk adjustment: This measure uses variables from claims data submitted by hospitals for payment as clinical risk adjusters. Our analyses have demonstrated that administrative claims data can be used to develop a risk-adjusted outcome measure for mortality following admission for HF and that the model produced estimates of RSMRs that are very similar to rates estimated by models based on chart data. This high level of agreement in the results based on the two different approaches supports the use of the claims-based model for public reporting. The model has also demonstrated consistent performance across years of claims data. The approach to gathering risk factors for patients also mitigates the potential limitations of claims data. Because not every diagnosis is coded at every visit, we use inpatient, outpatient, and physician claims data for the year prior to admission, and diagnosis codes during the index admission, for risk adjustment when the measure is used in Medicare FFS data. When the measure is used in all-payer data, only admission claims data (from the index hospitalization and prior year) are used for risk adjustment; however, model testing demonstrated both strong patient-level model performance and consistent hospital-level results when using	4d C ☐ P ☐ M ☐ N ☐

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

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

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

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

only admission claims data. The 1-year time frame provides a more comprehensive view of patients' medical histories than is provided by the secondary diagnosis codes from the index hospitalization alone. If a diagnosis appears in some visits and not others, it is included, minimizing the effect of incomplete coding. We were careful, however, to include information about each patient's status at admission and not to adjust for possible complications of the admission. Although some codes, by definition, represent conditions that are present before admission (e.g. cancer), other codes and conditions cannot be differentiated from complications during the hospitalization (e.g. infection or shock). If these are secondary diagnoses from the index admission, then they are not adjusted for in the analysis.

4e. Data Collection Strategy/Implementation

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

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

4e.3 Evidence for costs:
N/A

4e.4 Business case documentation: N/A

	4e
	C ☐
	P ☐
	M ☐
	N ☐

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

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

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

RECOMMENDATION	

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

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

CONTACT INFORMATION

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

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

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

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

Rating: C=Completely; P=Partially; M=Minimally; N=Not at all; NA=Not applicable   17

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

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

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The working group involved in the initial measure development is detailed in the original technical report available at www.qualitynet.org

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

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.6 Year the measure was first released: 2007
Ad.7 Month and Year of most recent revision: 04, 2011
Ad.8 What is your frequency for review/update of this measure? Yearly
Ad.9 When is the next scheduled review/update for this measure? 08, 2011

Ad.10 Copyright statement/disclaimers: N/A

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

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

1c. The measure focus is:

- an outcome (e.g., morbidity, mortality, function, health-related quality of life) that is relevant to, or associated with, a national health goal/priority, the condition, population, and/or care being addressed;

OR

- if an intermediate outcome, process, structure, etc., there is **evidence** that supports the specific measure focus as follows:
    o Intermediate outcome - evidence that the measured intermediate outcome (e.g., blood pressure, Hba1c) leads to improved health/avoidance of harm or cost/benefit.
    o Process - evidence that the measured clinical or administrative process leads to improved health/avoidance of harm and
    if the measure focus is on one step in a multi-step care process, it measures the step that has the greatest effect on improving the specified desired outcome(s).
    o Structure - evidence that the measured structure supports the consistent delivery of effective processes or access that lead to improved health/avoidance of harm or cost/benefit.
    o Patient experience - evidence that an association exists between the measure of patient experience of health care and the outcomes, values and preferences of individuals/ the public.
    o Access - evidence that an association exists between access to a health service and the outcomes of, or experience with, care.
    o Efficiency - demonstration of an association between the measured resource use and level of performance with respect to one or more of the other five IOM aims of quality.