

NATIONAL QUALITY FORUM

Measure Evaluation 4.1 December 2009

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

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

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

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

Evaluation ratings of the extent to which the criteria are met

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

(for NQF staff use) NQF Review #: 0358	NQF Project: Cardiovascular Endorsement Maintenance 2010
MEASURE DESCRIPTIVE INFORMATION	
De.1 Measure Title: Congestive Heart Failure (CHF) Mortality Rate (IQI 16)	
De.2 Brief description of measure: Percent of discharges with principal diagnosis code of CHF with in-hospital mortality	
1.1-2 Type of Measure: Outcome	
De.3 If included in a composite or paired with another measure, please identify composite or paired measure Mortality for Selected Conditions composite (NQF #0530)	
De.4 National Priority Partners Priority Area: Population health, Safety	
De.5 IOM Quality Domain: Effectiveness	
De.6 Consumer Care Need: Getting better	

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

every 3 years. Yes, information provided in contact section	N <input type="checkbox"/>
C. The intended use of the measure includes <u>both</u> public reporting <u>and</u> quality improvement. ► Purpose: Public reporting , Internal quality improvement	C Y <input checked="" type="checkbox"/> N <input type="checkbox"/>
D. The requested measure submission information is complete. Generally, measures should be fully developed and tested so that all the evaluation criteria have been addressed and information needed to evaluate the measure is provided. Measures that have not been tested are only potentially eligible for a time-limited endorsement and in that case, measure owners must verify that testing will be completed within 12 months of endorsement. D.1 Testing: Yes, fully developed and tested D.2 Have NQF-endorsed measures been reviewed to identify if there are similar or related measures? Yes	D Y <input checked="" type="checkbox"/> N <input type="checkbox"/>
(for NQF staff use) Have all conditions for consideration been met? Staff Notes to Steward (if submission returned):	Met Y <input checked="" type="checkbox"/> N <input type="checkbox"/>
Staff Notes to Reviewers (issues or questions regarding any criteria): How frequent is missing data excluded? How is AHRQ QI software "easily adapted to generate 30-day mortality rates"?	
Staff Reviewer Name(s): RWinkler	

TAP/Workgroup Reviewer Name:	
Steering Committee Reviewer Name:	
1. IMPORTANCE TO MEASURE AND REPORT	
Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance. <i>Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)</i> 1a. High Impact	Eval Rati ng
(for NQF staff use) Specific NPP goal : Safety: 1) All hospitals will reduce preventable and premature hospital-level mortality rates to best-in-class. 2) All hospitals and their community partners will improve 30-day mortality rates following hospitalization for select conditions (acute myocardial infarction, heart failure, pneumonia) to best-in-class.	
1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, Leading cause of morbidity/mortality, Severity of illness, Patient/societal consequences of poor quality 1a.2 1a.3 Summary of Evidence of High Impact: Approximately 2 million persons in the United States have heart failure each year. [1] These numbers will likely increase as the population ages. The literature suggests that hospitals have improved care for heart failure patients. In a study of 29,500 elderly patients in Oregon, the 3-day mortality decreased by 41% from 1991 to 1995. [2] 1a.4 Citations for Evidence of High Impact: [1] Smith, WM. Epidemiology of congestive heart failure. Am J Cardiol 1985;55(2):3A-8A. [2] Ni H, Hershberger FE. Was the decreasing trend in hospital mortality from heart failure attributable to improved hospital care? The Oregon experience, 1991-1995. Am J Manag Care 1999;5(9):1105-15.	1a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
1b. Opportunity for Improvement	1b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/>
1b.1 Benefits (improvements in quality) envisioned by use of this measure: Congestive heart failure (CHF) is a progressive, chronic disease with substantial short-term mortality, which varies from provider to provider.	

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

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

Better processes of care may reduce short-term mortality, which represents better quality.

N

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

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

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

1b.3 Citations for data on performance gap:

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

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

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

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

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

Estimate	Standard error	Gender
27.718	0.248	Male
29.119	0.235	Female

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

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

Estimate	Standard error	Expected payment source
35.572	0.575	Private insurance
26.881	0.184	Medicare

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.

29.834	0.885	Medicaid
57.840	1.615	Other insurance
34.378	1.437	Uninsured / self-pay / no charge
Estimate	Standard error	Hospital Ownership/control
27.378	0.197	Private, not-for-profit
28.834	0.449	Private, for-profit
33.192	0.507	Public
Estimate	Standard error	Teaching status
26.110	0.310	Teaching
29.164	0.203	Nonteaching
Estimate	Standard error	Location of hospital
25.569	0.297	Large central metropolitan
26.294	0.358	Large fringe metropolitan
25.442	0.370	Medium metropolitan
31.519	0.521	Small metropolitan
36.442	0.544	Micropolitan
48.180	0.894	Not metropolitan or micropolitan
Estimate	Standard error	Bed size of hospital
38.751	0.494	Less than 100
27.412	0.263	100 - 299
26.437	0.312	300 - 499
26.027	0.410	500 or more
1b.5 Citations for data on Disparities:		
See the following report for a complete treatment of the methodology: "Methods: Applying AHRQ Quality Indicators to Healthcare Cost and Utilization Project (HCUP) Data for the National Healthcare Quality Report" [URL: http://hcupnet.ahrq.gov/QI%20Methods.pdf?JS=Y]		
1c. Outcome or Evidence to Support Measure Focus		
1c.1 Relationship to Outcomes (<i>For non-outcome measures, briefly describe the relationship to desired outcome. For outcomes, describe why it is relevant to the target population</i>): Congestive heart failure (CHF) is a progressive, chronic disease with substantial short-term mortality, which varies from provider to provider. Better processes of care may reduce short-term mortality, which represents better quality.		
1c.2-3. Type of Evidence: Expert opinion, Systematic synthesis of research		
1c.4 Summary of Evidence (<i>as described in the criteria; for outcomes, summarize any evidence that healthcare services/care processes influence the outcome</i>): The existence of a board quality committee was associated with higher likelihoods of adopting various oversight practices and lower mortality rates for congestive heart failure measured by the Agency for Healthcare Research and Quality's Inpatient Quality Indicators and the State Inpatient Databases. [1]		
References: [1] Jiang, H. Joanna; Lockee, Carlin; Bass, Karma; Fraser, Irene; Kiely, Robert. (2008) Board engagement in quality: findings of a survey of hospital and system leaders. Journal of Healthcare Management, 53, 2, 121(15)		
1c.5 Rating of strength/quality of evidence (<i>also provide narrative description of the rating and by whom</i>): 6 Smoothing recommended Testing, rating, and review were conducted by the project team. A full report on the literature review and empirical evaluation can be found in Refinement of the HCUP Quality Indicators by the UCSF-Stanford EPC, Detailed coding information for each QI is provided in the document Prevention		
1c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>		

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

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

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

Quality Indicators Technical Specifications. Rating of performance on empirical evaluations, ranged from 0 to 26. The scores were intended as a guide for summarizing the performance of each indicator on four empirical tests of precision (signal variance, area-level share, signal ratio, and R-squared) and five tests of minimum bias (rank correlation, top and bottom decile movement, absolute change, and change over two deciles), as described in the previous section.

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

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

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

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

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

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

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

- Rank correlation coefficient of the area or hospital with (and without) risk adjustment—gives the overall impact of risk adjustment on relative provider or area performance.
- Average absolute value of change relative to mean—highlights the amount of absolute change in performance, without reference to other providers’ performance.
- Percentage of highly ranked hospitals that remain in high decile—reports the percentage of hospitals or areas that are in the highest deciles without risk adjustment that remain there after risk adjustment is performed.

<ul style="list-style-type: none"> Percentage of lowly ranked hospitals that remain in low decile—reports the percentage of hospitals or areas that are in the lowest deciles without risk adjustment that remain there after risk adjustment is performed. Percentage that change more than two deciles—identifies the percentage of hospitals whose relative rank changes by a substantial percentage (more than 20%) with and without risk adjustment. <p>Construct validity. Construct validity analyses provided information regarding the relatedness or independence of the indicators. If quality indicators do indeed measure quality, then two measures of the same construct would be expected to yield similar results. The team used factor analysis to reveal underlying patterns among large numbers of variables—in this case, to measure the degree of relatedness between indicators. In addition, they analyzed correlation matrices for indicators.</p> <p>1c.7 Summary of Controversy/Contradictory Evidence: See the following for a complete treatment of the topic: http://www.qualityindicators.ahrq.gov/downloads/iqi/iqi_guide_v31.pdf Note: The Literature Review Caveats column summarizes evidence specific to each potential concern on the link between the PQIs and quality of care, as described in step 3 above. A question mark (?) indicates that the concern is theoretical or suggested, but no specific evidence was found in the literature. A check mark indicates that the concern has been demonstrated in the literature.</p> <p>1c.8 Citations for Evidence (other than guidelines): http://www.qualityindicators.ahrq.gov/downloads/iqi/iqi_guide_v31.pdf</p> <p>1c.9 Quote the Specific guideline recommendation (including guideline number and/or page number): Not Applicable.</p> <p>1c.10 Clinical Practice Guideline Citation: Not Applicable.</p> <p>1c.11 National Guideline Clearinghouse or other URL: Not Applicable.</p> <p>1c.12 Rating of strength of recommendation (also provide narrative description of the rating and by whom): Not Applicable.</p> <p>1c.13 Method for rating strength of recommendation (If different from USPSTF system, also describe rating and how it relates to USPSTF): Not Applicable.</p> <p>1c.14 Rationale for using this guideline over others: Not Applicable.</p>	
TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Importance to Measure and Report?	1
Steering Committee: Was the threshold criterion, Importance to Measure and Report, met? Rationale:	1 Y <input type="checkbox"/> N <input type="checkbox"/>
2. SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES	
Extent to which the measure, <u>as specified</u> , produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)	Eval Rati ng
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- spe cs
<u>2a. Precisely Specified</u>	C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/>
2a.1 Numerator Statement (Brief, text description of the numerator - what is being measured about the target population, e.g. target condition, event, or outcome): Number of deaths (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator.	

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

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

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

2a.2 Numerator Time Window (*The time period in which cases are eligible for inclusion in the numerator*):
 Time window can be determined by user, but is generally a calendar year.

2a.3 Numerator Details (*All information required to collect/calculate the numerator, including all codes, logic, and definitions*):
 Number of deaths (DISP=20) among cases meeting the inclusion and exclusion rules for the denominator.

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

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

2a.7 Denominator Time Window (*The time period in which cases are eligible for inclusion in the denominator*):
 Time window can be determined by user, but is generally a calendar year.

2a.8 Denominator Details (*All information required to collect/calculate the denominator - the target population being measured - including all codes, logic, and definitions*):
 All discharges, age 18 years and older, with a principal diagnosis code of CHF.
 ICD-9-CM CHF diagnosis codes:

- 39891
- RHEUMATIC HEART FAILURE
- 40201
- MAL HYPERT HRT DIS W CHF
- 40211
- BENIGN HYP HRT DIS W CHF
- 40291
- HYPERTEN HEART DIS W CHF
- 40401
- MAL HYPER HRT/REN W CHF
- 40403
- MAL HYP HRT/REN W CHF&RF
- 40411
- BEN HYPER HRT/REN W CHF
- 40413
- BEN HYP HRT/REN W CHF&RF
- 40491
- HYPER HRT/REN NOS W CHF
- 40493
- HYP HT/REN NOS W CHF&RF
- 4280
- CONGESTIVE HEART FAILURE
- 4281
- LEFT HEART FAILURE
- 42820
- SYSTOLIC HEART FAILURE NOS OCT02-
- 42821
- AC SYSTOLIC HRT FAILURE OCT02-
- 42822
- CHR SYSTOLIC HRT FAILURE OCT02-
- 42823
- AC ON CHR SYST HRT FAIL OCT02-
- 4289
- HEART FAILURE NOS
- 42830
- DIASTOLIC HRT FAILURE NOS OCT02-

<p>42831 AC DIASTOLIC HRT FAILURE OCT02- 42832 CHR DIASTOLIC HRT FAIL OCT02- 42833 AC ON CHR DIAST HRT FAIL OCT02- 42840 SYST/DIAST HRT FAIL NOS OCT02- 42841 AC SYST/DIASTOL HRT FAIL OCT02- 42842 CHR SYST/DIASTL HRT FAIL OCT02- 42843 AC/CHR SYST/DIA HRT FAIL OCT02-</p> <p>Exclude cases:</p> <ul style="list-style-type: none"> • missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1 =missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium)
<p>2a.9 Denominator Exclusions (<i>Brief text description of exclusions from the target population</i>):</p> <ul style="list-style-type: none"> • missing discharge disposition (DISP=missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium)
<p>2a.10 Denominator Exclusion Details (<i>All information required to collect exclusions to the denominator, including all codes, logic, and definitions</i>):</p> <p>Exclude cases:</p> <ul style="list-style-type: none"> • missing discharge disposition (DISP=missing), gender (SEX=missing), age (AGE=missing), quarter (DQTR=missing), year (YEAR=missing) or principal diagnosis (DX1 =missing) • transferring to another short-term hospital (DISP=2) • MDC 14 (pregnancy, childbirth, and puerperium)
<p>2a.11 Stratification Details/Variables (<i>All information required to stratify the measure including the stratification variables, all codes, logic, and definitions</i>):</p> <p>Gender, age (5-year age groups), race / ethnicity, primary payer, custom</p>
<p>2a.12-13 Risk Adjustment Type: Risk adjustment method widely or commercially available</p>
<p>2a.14 Risk Adjustment Methodology/Variables (<i>List risk adjustment variables and describe conceptual models, statistical models, or other aspects of model or method</i>):</p> <p>The predicted value for each case is computed using a hierarchical model (logistic regression with hospital random effect) and covariates for gender, age in years (in 5-year age groups), All Patient Refined-Diagnosis Related Group (APR-DRG) and APR-DRG risk-of-mortality subclass. The reference population used in the model is the universe of discharges for states that participate in the HCUP State Inpatient Databases (SID) for the year 2007 (updated annually), a database consisting of 43 states and approximately 30 million adult discharges. The expected rate is computed as the sum of the predicted value for each case divided by the number of cases for the unit of analysis of interest (i.e., hospital, state, and region). The risk adjusted rate is computed using indirect standardization as the observed rate divided by the expected rate, multiplied by the reference population rate.</p> <p>Required data elements: Patient gender; age in years at admission; International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) principal and secondary diagnosis codes. A limited license 3M APR-DRG grouper is included with the AHRQ QI Software.</p>
<p>2a.15-17 Detailed risk model available Web page URL or attachment: URL http://qualityindicators.ahrq.gov/downloads/iqi/IOI_Risk_Adjustment_Tables_(Version_4_2).pdf</p>
<p>2a.18-19 Type of Score: Rate/proportion</p>
<p>2a.20 Interpretation of Score: Better quality = Lower score</p>

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

<p>2a.21 Calculation Algorithm <i>(Describe the calculation of the measure as a flowchart or series of steps):</i> RATE: Each Inpatient Quality Indicator (IQI) expressed as a rate, is defined as outcome of interest/population at risk or numerator/denominator. The Quality Indicators software performs five steps to produce the IQI rates. 1) Discharge-level data is used to mark inpatient records containing outcomes of interest. 2) Identify populations at risk. 3) Calculate observed rates. 4) For rates that are not risk-adjusted, the risk-adjusted rate equals the observed rate. 5) Create multivariate signal extraction (MSX) smoothed rates. Shrinkage factors are applied to the risk-adjusted rates for each PQI in the MSX process. For each IQI, the shrinkage estimate reflects a reliability adjustment unique to each indicator. Full information on IQI algorithms and specification can be found at http://qualityindicators.ahrq.gov/Iqi_download.htm.</p>	
<p>2a.22 Describe the method for discriminating performance <i>(e.g., significance testing):</i> Significance testing is not prescribed by the software. Users may calculate a confidence interval for the risk-adjusted rates and a posterior probability interval for the smoothed rates at a 95% or 99% level. Users may define the relevant benchmark and the methods of discriminating performance according to their application.</p>	
<p>2a.23 Sampling (Survey) Methodology <i>If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):</i> Not applicable</p>	
<p>2a.24 Data Source <i>(Check the source(s) for which the measure is specified and tested)</i> Electronic administrative data/claims</p>	
<p>2a.25 Data source/data collection instrument <i>(Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.):</i> The data source is hospital discharge data such as the HCUP State Inpatient Databases (SID) or equivalent using UB-04 coding standards. The data collection instrument is public-use AHRQ QI software available in SAS or Windows versions.</p>	
<p>2a.26-28 Data source/data collection instrument reference web page URL or attachment: URL None http://www.qualityindicators.ahrq.gov/software.htm</p>	
<p>2a.29-31 Data dictionary/code table web page URL or attachment: URL None http://www.qualityindicators.ahrq.gov/downloads/winqi/AHRQ_QI_Windows_Software_Documentation_V41a.pdf</p>	
<p>2a.32-35 Level of Measurement/Analysis <i>(Check the level(s) for which the measure is specified and tested)</i> Facility/Agency</p>	
<p>2a.36-37 Care Settings <i>(Check the setting(s) for which the measure is specified and tested)</i> Hospital</p>	
<p>2a.38-41 Clinical Services <i>(Healthcare services being measured, check all that apply)</i> Clinicians: Physicians (MD/DO)</p>	
TESTING/ANALYSIS	
<p>2b. Reliability testing</p> <p>2b.1 Data/sample <i>(description of data/sample and size):</i> Veterans Integrated Service Networks (VISNs); and VA versus non-VA (Nationwide Inpatient Sample) using VA inpatient data (2004-2007). [1]</p> <p>A survey of hospital and system leaders (presidents/chief executive officers (CEOs)) that was conducted in the first six months of 2006 with a total of 562 respondents. Hospital-level data for these composite measures were produced by applying the IQI to the State Inpatient Databases (SID) of the Healthcare Cost and Utilization Project (HCUP) sponsored by AHRQ. The SID includes all-payer data on inpatient stays from virtually all community hospitals in each participating state. [2]</p> <p>Using 1995 to 2000 data from New York state (n = 7,021,065), analysts compared mortality risk (odds ratio) for individuals with and without Alzheimer’s disease. [3]</p>	<p style="text-align: right;">2b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

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

We restricted our analysis to 20 states (4) for which HCUP State Inpatient Databases (SID) were available. There were 1,601 nonfederal, urban, general hospitals in those 20 states. Over 300 hospitals were eliminated from the sample because of key missing variables in the American Hospital Association (AHA) Annual Survey of Hospital data, which was also used for this study, or because they had missing observations for some of the measures that we used. Thus, our sample consisted of 1,290 urban, acute-care hospitals for which complete data were available for 2001. [4]

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

VA-and VISN-level IQI observed rates, risk-adjusted rates, and observed to expected ratios (O/Es). We examined the trends in VA-and VISN-level rates using weighted linear regression, variation in VISN-level O/Es, and compared VA to non-VA trends. [1]

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

Odds Ratio. [3]

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

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

VA in-hospital mortality rates for CHF Mortality were unchanged over time. The IQIs are easily applied to VA administrative data. They can be useful to tracks rate trends over time, reveal variation between sites, and for trend comparisons with other healthcare systems. [1]

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

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

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

References

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

2c. Validity testing

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

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

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.

2c
 C
 P
 M
 N

<p>2c.2 Analytic Method (<i>type of validity & rationale, method for testing</i>): The AHRQ IQI software was used to calculate risk-adjusted mortality rates using either (1) routine administrative data that included all the International Classification of Diseases (ICD)-9-CM codes or (2) enhanced administrative data that included only the ICD-9-CM codes representing preexisting conditions. [1]</p> <p>We derived 4-year facility-level in-hospital and 30-day observed mortality rates and observed/expected ratios (O/Es) for admissions with a principal diagnosis of acute myocardial infarction, congestive heart failure, stroke, gastrointestinal hemorrhage, hip fracture, and pneumonia. We standardized software-calculated O/Es to the VA population and compared O/Es and outlier status across sites using correlation, observed agreement, and kappas. [2]</p> <p>2c.3 Testing Results (<i>statistical results, assessment of adequacy in the context of norms for the test conducted</i>): Without using POA data, for congestive heart failure 25% of hospitals classified as low-quality hospitals using enhanced administrative data were misclassified as intermediate-quality hospitals using routine administrative data. Despite the fact that the AHRQ IQIs were primarily intended to serve as a screening tool, they are being increasingly used to publicly report hospital quality. These findings emphasized the need (which the AHRQ IQI have now adopted by incorporating POA data in the risk-adjustment) to improve the "quality" of administrative data by including a POA indicator if these data are to serve as the information infrastructure for quality reporting. [1]</p> <p>Of 119 facilities, in-hospital versus 30-day mortality O/E correlations were generally high (median: r = 0.78; range: 0.31-0.86). Examining outlier status, observed agreement was high (median: 84.7%, 80.7%-89.1%). Kappas showed at least moderate agreement (k > 0.40) for all indicators except stroke and hip fracture (k = 0.22). Across indicators, few sites changed from a high to nonoutlier or low outlier, or vice versa (median: 10, range: 7-13). The AHRQ IQI software can be easily adapted to generate 30-day mortality rates. Although 30-day mortality has better face validity as a hospital performance measure than in-hospital mortality, site assessments were similar despite the definition used. [3]</p> <p>References [1] Glance L.G.; Osler T.M.; Mukamel D.B.; Dick A.W. Impact of the present-on-admission indicator on hospital quality measurement: Experience with the Agency for Healthcare Research and Quality (AHRQ) Inpatient Quality Indicators. <i>Medical Care</i>, v. 46, no. 2, Feb. 2008, p. 112-119. DOI: 10.1097/MLR.0b013e318158aed6. [2] Borzecki AM, Christiansen CL, Chew P, Loveland S, Rosen AK. Comparison of in-hospital versus 30-day mortality assessments for selected medical conditions. <i>Med Care</i>. 2010 Dec;48(12):1117-21. PMID: 20978451</p>	
<p>2d. Exclusions Justified</p> <p>2d.1 Summary of Evidence supporting exclusion(s): Exclusions remove cases where the outcome of interest is less likely to be preventable or more likely to be preventable or with no or very low risk</p> <p>2d.2 Citations for Evidence: Refinement of the HCUP Quality Indicators (Technical Review), May 2001 http://qualityindicators.ahrq.gov/downloads/technical/qi_technical_review.zip</p> <p>2d.3 Data/sample (<i>description of data/sample and size</i>): AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges</p> <p>2d.4 Analytic Method (<i>type analysis & rationale</i>): Expert panel and descriptive analyses stratified by exclusion categories</p> <p>2d.5 Testing Results (<i>e.g., frequency, variability, sensitivity analyses</i>): Refinement of the HCUP Quality Indicators (Technical Review), May 2001 http://qualityindicators.ahrq.gov/downloads/technical/qi_technical_review.zip</p>	<p>2d C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>2e. Risk Adjustment for Outcomes/ Resource Use Measures</p>	<p>2e</p>

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

Comment [KP14]: 2d. Clinically necessary measure exclusions are identified and must be:

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

AND

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

AND

- precisely defined and specified:
 –If there is substantial variability in exclusions across providers, the measure is specified so that exclusions are computable and the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);
 if patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that it strongly impacts performance on the measure and the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

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

Comment [KP16]: 2e. For outcome measures and other measures (e.g., resource use) when indicated:

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care) and are present at start of care; Error! Bookmark not defined. OR rationale/data support no risk adjustment.

<p>2e.1 Data/sample (<i>description of data/sample and size</i>): AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges</p> <p>2e.2 Analytic Method (<i>type of risk adjustment, analysis, & rationale</i>): Risk-adjustment models use a standard set of categories based on readily available classification systems for demographics, severity of illness and comorbidities. Within each category, covariates are initially selected based on a minimum of 30 cases in the outcome of interest. Then a stepwise regression process on a development sample is used to select a parsimonious set of covariates where $p < .05$. Model is then tested on a validation sample.</p> <p>2e.3 Testing Results (<i>risk model performance metrics</i>): c 0.787</p> <p>2e.4 If outcome or resource use measure is not risk adjusted, provide rationale: Not applicable</p>	<p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>												
<p>2f. Identification of Meaningful Differences in Performance</p> <p>2f.1 Data/sample from Testing or Current Use (<i>description of data/sample and size</i>): AHRQ 2007 State Inpatient Databases (SID) with 4,000 hospitals and 30 million adult discharges</p> <p>2f.2 Methods to identify statistically significant and practically/meaningfully differences in performance (<i>type of analysis & rationale</i>): Posterior probability distribution parameterized using the Gamma distribution</p> <p>2f.3 Provide Measure Scores from Testing or Current Use (<i>description of scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance</i>):</p> <table border="1" data-bbox="87 1003 1112 1058"> <thead> <tr> <th></th> <th>5th</th> <th>25th</th> <th>Median</th> <th>75th</th> <th>95th</th> </tr> </thead> <tbody> <tr> <td></td> <td>0.017245</td> <td>0.025607</td> <td>0.032831</td> <td>0.041305</td> <td>0.055832</td> </tr> </tbody> </table>		5th	25th	Median	75th	95th		0.017245	0.025607	0.032831	0.041305	0.055832	<p>2f</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p>
	5th	25th	Median	75th	95th								
	0.017245	0.025607	0.032831	0.041305	0.055832								
<p>2g. Comparability of Multiple Data Sources/Methods</p> <p>2g.1 Data/sample (<i>description of data/sample and size</i>): Not applicable</p> <p>2g.2 Analytic Method (<i>type of analysis & rationale</i>): Not applicable</p> <p>2g.3 Testing Results (<i>e.g., correlation statistics, comparison of rankings</i>): Not applicable</p>	<p>2g</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>												
<p>2h. Disparities in Care</p> <p>2h.1 If measure is stratified, provide stratified results (<i>scores by stratified categories/cohorts</i>): Median income of patient's ZIP code: 1) Estimate 2) Standard error 3) P-value: Relative to marked group-c 4) P-value: 2007 relative to 2006 First quartile (lowest income) 30.165 0.309 0.000 0.000 Second quartile 27.842 0.333 0.184 0.000 Third quartile 27.121 0.353 0.909 0.000 Fourth quartile (highest income)c 27.179 0.372 0.000</p> <p>[1] Although we did find overall disparities in care, we found that indicators for blacks, Hispanics, and Asians were not statistically worse than corresponding quality indicators for whites in the same hospital. Only a few hospitals provide lower quality of care to minorities than to whites.</p> <p>[1] Darrell J. Gaskin, Christine S. Spencer, Patrick Richard, Gerard F. Anderson, Neil R. Powe and Thomas A. LaVeist. Do Hospitals Provide Lower-Quality Care To Minorities Than To Whites? Health Affairs, 27, no. 2 (2008): 518-527 doi: 10.1377/hlthaff.27.2.518</p>	<p>2h</p> <p>C <input type="checkbox"/></p> <p>P <input type="checkbox"/></p> <p>M <input type="checkbox"/></p> <p>N <input type="checkbox"/></p> <p>NA <input type="checkbox"/></p>												

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

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

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

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

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

<p>2h.2 If disparities have been reported/identified, but measure is not specified to detect disparities, provide follow-up plans: Users may stratify based on gender and race/ethnicity</p>	
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for <i>Scientific Acceptability of Measure Properties</i>?</p>	2
<p>Steering Committee: Overall, to what extent was the criterion, <i>Scientific Acceptability of Measure Properties</i>, met? Rationale:</p>	<p>2 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
3. USABILITY	
<p>Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (evaluation criteria)</p>	<p>Eval Rati ng</p>
<p><u>3a. Meaningful, Understandable, and Useful Information</u></p> <p>3a.1 Current Use: In use</p> <p>3a.2 Use in a public reporting initiative (disclosure of performance results to the public at large) (If used in a public reporting initiative, provide name of initiative(s), locations, Web page URL(s). <u>If not publicly reported</u>, state the plans to achieve public reporting within 3 years):</p> <p>Arizona (NY QIO) Why Not the Best? http://www.http://whynotthebest.org/</p> <p>California (state) Hospital Inpatient Mortality Indicators for California http://www.oshpd.ca.gov/HID/Products/PatDischargeData/AHQ/iqi-imi_overview.html</p> <p>Colorado (state hospital association) Colorado Hospital Report Card http://www.cohospitalquality.org/index.php?option=com_frontpage&Itemid=1</p> <p>Florida (state) Florida Health Finder http://www.floridahealthfinder.gov/</p> <p>Illinois (state) Illinois Hospital Report Card and Consumer Guide to Health Care http://www.healthcarereportcard.illinois.gov/</p> <p>Iowa (Iowa Healthcare Collaborative) Iowa Healthcare Collaborative http://www.ihtonline.org/asp/publicreporting/iowareport.aspx</p> <p>Kentucky (Norton Healthcare, a hospital system) Norton Healthcare Quality Report http://www.nortonhealthcare.com/body.cfm?id=157</p> <p>Kentucky (state) Health Care Information Center http://chfs.ky.gov/ohp/healthdata</p>	<p>3a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

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.

Kentucky (state hospital association)
 Kentucky Hospital Association Quality Data
<http://info.kyha.com/QualityData/IQISite/>

Maine (state)
 Maine Health Data Organization
<http://gateway.maine.gov/mhdo2008Monahrq/home.html>

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

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

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

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

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

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

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

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

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

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

The measure is also reported on HCUPnet:
http://hcupnet.ahrq.gov/HCUPnet.jsp?ld=EB57801381F71C41&Form=MAINSEL&JS=Y&Action=%3E%3ENext%3E%3E%3E_MAINSEL=AHRO%20Quality%20Indicators

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

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

University Healthcare Consortium - An alliance of 103 academic medical centers and 219 of their affiliated hospitals. Reporting the AHRQ QIs to their member hospitals. (see www.uhc.edu. Note: measure results reported to hospitals; not reported on site).

Dallas Fort Worth Hospital Council - Reporting on measure results to over 70 hospitals in Texas (see www.dfwhc.org. Note: measure results reported to hospitals; not reported on site).

Norton Healthcare - a multi-hospital system in Kentucky (see http://www.nortonhealthcare.com/about/Our_Performance/index.aspx)

Ministry Health Care - a multi-hospital system in Wisconsin (see <http://ministryhealth.org/display/router.aspx>. Note: measure results reported to hospitals; not reported on site).

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

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

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

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

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

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

A research team from the School of Public Affairs, Baruch College, under contracts with the Department of Public Health, Weill Medical College and Battelle, Inc., has developed a pair of Hospital Quality Model Reports at the request of the Agency for Healthcare Research & Quality (AHRQ). These reports are designed specifically to report comparative information on hospital performance based on the AHRQ Quality Indicators (QIs). The work was done in close collaboration with AHRQ staff and the AHRQ Quality Indicators team. The Model Reports (discussed immediately above) are based on:

- Extensive search and analysis of the literature on hospital quality measurement and reporting, as well as public reporting on health care quality more broadly;
- Interviews with quality measurement and reporting experts, purchasers, staff of purchasing coalitions, and executives of integrated health care delivery systems who are responsible for quality in their facilities;
- Two focus groups with chief medical officers of hospitals and/or systems and two focus groups with quality managers from a broad mix of hospitals;
- Four focus groups with members of the public who had recently experienced a hospital admission; and
- Four rounds of cognitive interviews (a total of 62 interviews) to test draft versions of the two Model Reports with members of the public with recent hospital experience, basic computer literacy but widely varying levels of education.

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

Given the above review of the literature and original research that was conducted, a Model report was the result that could help sponsors use the best evidence on public reports so they are most likely to have the desired effects on quality.

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

3b.1 NQF # and Title of similar or related measures:
[CMS CHF Mortality Measure](#)

(for NQF staff use) Notes on similar/related [endorsed](#) or submitted measures: #0229- 30-day, all cause risk standardized mortality rate (RSMR) following heart failure hospitalization (CMS)

<p>3b. Harmonization If this measure is related to measure(s) already endorsed by NQF (e.g., same topic, but different target population/setting/data source or different topic but same target population): 3b.2 Are the measure specifications harmonized? If not, why? The specifications are harmonized, but CMS uses 30-day mortality</p>	<p>3b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>3c. Distinctive or Additive Value 3c.1 Describe the distinctive, improved, or additive value this measure provides to existing NQF-endorsed measures: The AHRQ is all-payer (not Medicare FFS only) and uses in-hospital mortality, which is available in real-time 5.1 If this measure is similar to measure(s) already endorsed by NQF (i.e., on the same topic and the same target population), Describe why it is a more valid or efficient way to measure quality: The AHRQ measure provides a real-time indication of hospital performances, reflects the patient's experience in the hospital, and is available for all-payers</p>	<p>3c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Usability?</p>	<p>3</p>
<p>Steering Committee: Overall, to what extent was the criterion, Usability, met? Rationale:</p>	<p>3 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
4. FEASIBILITY	
<p>Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. (evaluation criteria)</p>	<p>Evaluation</p>
<p>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)</p>	<p>4a C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>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.</p>	<p>4b C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>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.</p>	<p>4c C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/> NA <input type="checkbox"/></p>
<p>4d. Susceptibility to Inaccuracies, Errors, or Unintended Consequences 4d.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measure and describe how these potential problems could be audited. If audited, provide results. Coding professionals follow detail guidelines, are subject to training and credentialing requirements, peer review and audit. Risk-adjusted measures of mortality may lead to an increase in coding of comorbidities. All in-hospital</p>	<p>4d C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>

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

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

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

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

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

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

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

<p>mortality measures may encourage earlier post-operative discharge, and thereby shift deaths to skilled nursing facilities or outpatient settings. However, Rosenthal et al. found no evidence that hospitals with lower in-hospital standardized mortality had higher (or lower) early post-discharge mortality. [1]</p> <p>Coding professionals follow detailed guidelines, are subject to training and credentialing requirements, peer review and audit.</p> <p>References: [1] Rosenthal GE, Baker DW, Norris DG, et al. Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. <i>Health Serv Res</i> 2000;34(7):1449-68.</p>	
<p>4e. Data Collection Strategy/Implementation</p> <p>4e.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data/missing data, timing/frequency of data collection, patient confidentiality, time/cost of data collection, other feasibility/ implementation issues: Relative to other indicators, a lower percentage of the variation occurs at the provider level rather than the discharge level. The signal ratio (i.e., the proportion of the total variation across providers that is truly related to systematic differences in provider performance rather than random variation) is moderate, at 53.5%, indicating that some of the observed differences in provider performance likely do not represent true differences.</p> <p>4e.2 Costs to implement the measure (costs of data collection, fees associated with proprietary measures): All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm</p> <p>4e.3 Evidence for costs: All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm</p> <p>4e.4 Business case documentation: All data necessary to calculate this measure are routinely collected for hospital administrative purposes. The software for calculating the measure is available for free at: http://www.qualityindicators.ahrq.gov/software.htm</p>	<p>4e C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>TAP/Workgroup: What are the strengths and weaknesses in relation to the subcriteria for Feasibility?</p>	<p>4</p>
<p>Steering Committee: Overall, to what extent was the criterion, <i>Feasibility</i>, met? Rationale:</p>	<p>4 C <input type="checkbox"/> P <input type="checkbox"/> M <input type="checkbox"/> N <input type="checkbox"/></p>
<p>RECOMMENDATION</p>	
<p>(for NQF staff use) Check if measure is untested and only eligible for time-limited endorsement.</p>	<p>Time - limit ed <input type="checkbox"/></p>
<p>Steering Committee: Do you recommend for endorsement? Comments:</p>	<p>Y <input type="checkbox"/> N <input type="checkbox"/> A <input type="checkbox"/></p>
<p>CONTACT INFORMATION</p>	
<p>Co.1 Measure Steward (Intellectual Property Owner) Co.1 <u>Organization</u> Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850</p>	

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

<p>Co.2 Point of Contact John, Bott, MSSW, MBA, david.atkins@ahrq.hhs.gov, 301-427-1608-</p>
<p>Measure Developer If different from Measure Steward Co.3 Organization Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, Maryland, 20850</p>
<p>Co.4 Point of Contact John, Bott, MSSW, MBA, david.atkins@ahrq.hhs.gov, 301-427-1608-</p>
<p>Co.5 Submitter If different from Measure Steward POC John, Bott, MSSW, MBA, david.atkins@ahrq.hhs.gov, 301-427-1608-, Agency for Healthcare Research and Quality</p>
<p>Co.6 Additional organizations that sponsored/participated in measure development UC Davis, Stanford University, Battelle Memorial Institute</p>
<p>ADDITIONAL INFORMATION</p>
<p>Workgroup/Expert Panel involved in measure development Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development. None</p>
<p>Ad.2 If adapted, provide name of original measure: None Ad.3-5 If adapted, provide original specifications URL or attachment</p>
<p>Measure Developer/Steward Updates and Ongoing Maintenance Ad.6 Year the measure was first released: 2001 Ad.7 Month and Year of most recent revision: 10, 2010 Ad.8 What is your frequency for review/update of this measure? Annual Ad.9 When is the next scheduled review/update for this measure? 05, 2011</p>
<p>Ad.10 Copyright statement/disclaimers: The AHRQ QI software is publicly available; no copyright disclaimers</p>
<p>Ad.11 -13 Additional Information web page URL or attachment:</p>
<p>Date of Submission (MM/DD/YY): 02/01/2011</p>

1c. The measure focus is:

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

OR

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