**National Quality Forum—Measure Testing (subcriteria 2a2, 2b2-2b7)**

**Measure Number** (*if previously endorsed*)**:** 0283

**Measure Title**: Asthma in Younger Adults Admission Rate (PQI 15)

**Date of Submission**: 12/14/2015

**Type of Measure:**

|  |  |
| --- | --- |
| Composite | Outcome (*including PRO-PM*) |
| Cost/resource | Process |
| Efficiency | Structure |

|  |
| --- |
| **Instructions**   * Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form. * **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.** * **For outcome and resource use measures**, section **2b4** also must be completed. * If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed. * Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed. * If you are unable to check a box, please highlight or shade the box for your response. * Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.*** * Contact NQF staff regarding questions. Check for resources at [Submitting Standards webpage](http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx). * For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment. |

|  |
| --- |
| **Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF’s evaluation criteria for testing.**  **2a2.** **Reliability testing** [**10**](#Note10) demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.  **2b2.** **Validity testing** [**11**](#Note11) demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.    **2b3.** Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [**12**](#Note12)  **AND**  If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, 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). [**13**](#Note13)  **2b4.** **For outcome measures and other measures when indicated** (e.g., resource use):   * **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [**14**](#Note14)**,**[**15**](#Note15) and has demonstrated adequate discrimination and calibration   **OR**   * rationale/data support no risk adjustment/ stratification.   **2b5.** Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [**16**](#Note16) **differences in performance**;  **OR**  there is evidence of overall less-than-optimal performance.  **2b6.** **If multiple data sources/methods are specified, there is demonstration they produce comparable results**.  **2b7.** For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data),analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.  **Notes**  **10.** Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements 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 of the measure score addresses precision of measurement (e.g., signal-to-noise).  **11.** Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.  **12.** Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.  **13.** Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.  **14.** Risk factors that influence outcomes should not be specified as exclusions  **15.** 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 percent v. 75 percent) 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 less-than-optimal performance may not demonstrate much variability across providers. |

**1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE**

*Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.*

**1.1. What type of data was used for testing**? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

|  |  |
| --- | --- |
| **Measure Specified to Use Data From:**  **(*must be consistent with data sources entered in S.23*)** | **Measure Tested with Data From:** |
| abstracted from paper record | abstracted from paper record |
| administrative claims | administrative claims |
| clinical database/registry | clinical database/registry |
| abstracted from electronic health record | abstracted from electronic health record |
| eMeasure (HQMF) implemented in EHRs | eMeasure (HQMF) implemented in EHRs |
| other: Click here to describe | other: Click here to describe |

**1.2. If an existing dataset was used, identify the specific dataset** (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

All analyses were completed using data from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID), 2009-2013. HCUP is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ)[[1]](#footnote-1). HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. The HCUP SID contain the universe of the inpatient discharge abstracts in participating States, translated into a uniform format to facilitate multi-State comparisons and analyses. All states provide data for community hospitals and together, the SID encompasses about 97 percent of all U.S. community hospital discharges. For the analyses presented here, we use 40 states representing about 89 percent of the U.S. community hospital discharges, for a total of about 30 million hospital discharges from community hospitals. As defined by the American Hospital Association, community hospitals are all non-Federal, short-term, general or other specialty hospitals, excluding hospital units of institutions. Included among community hospitals are public and academic medical centers, specialty hospitals such as obstetrics–gynecology, ear–nose–throat, orthopedic and pediatric institutions. Short-stay rehabilitation, long-term acute care hospitals are excluded from the data used for the reported analyses.

The SID data elements include ICD-9-CM coded principal and secondary diagnoses and procedures, additional detailed clinical and service information based on revenue codes, admission and discharge status, patient demographics, expected payment source (Medicare, Medicaid, private insurance as well as the uninsured), total charges and length of stay ([www.hcup-us.ahrq.gov](http://www.hcup-us.ahrq.gov)).

For additional testing of the indicators we created a county characteristic dataset which was merged with the county-level observed and risk adjusted rates from the HCUP dataset described above. Using a conceptual model for hospitalization indicators, we examined candidate predictor variables from the County Health Rankings (CHR)[[2]](#footnote-2) dataset and the American Community Survey (ACS)[[3]](#footnote-3). Candidate variables that corresponded to the conceptual model were grouped by the following categories: individual health behavior (IHB) included variables that include actions and behaviors of individuals and may be mutable, access to care (AC) included variables that reflect the structure and quality of the healthcare system in a community, socioeconomic status (SES) included variables of poverty and education levels within a community and environment (E) variables included community characteristics, such as access to food and open spaces, pollution or violent crime. County prevalence estimates were derived from the Behavioral Risk Factor Surveillance System (BRFSS) Asthma model based county estimates.

**1.3. What are the dates of the data used in testing**?

HCUP data: 2009-2013, CHR data: 2014 and ACS data: 2013.

**1.4. What levels of analysis** **were tested**? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

County

|  |  |
| --- | --- |
| **Measure Specified to Measure Performance of:**  **(*must be consistent with levels entered in item S.26*)** | **Measure Tested at Level of:** |
| individual clinician | individual clinician |
| group/practice | group/practice |
| hospital/facility/agency | hospital/facility/agency |
| health plan | health plan |
| other: Population Health | other: County |

**1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

**Table 1. Reference Population Rate and Distribution of County Performance for** **PQI 15 Asthma in Younger Adults Admission Rate**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Overall Reference Population Rate** | | | | |
| **Year** | **Number Counties** | **Outcome of Interest**  **(Numerator)1** | **Population at Risk**  **(Denominator)1** | **Observed Rate**  **Per 10001** |
| 2009 | 3,135 | 52,986 | 91,475,217 | 0.5792 |
| 2010 | 3,138 | 46,476 | 91,767,953 | 0.5065 |
| 2011 | 3,141 | 43,685 | 92,184,336 | 0.4739 |
| 2012 | 3,139 | 43,746 | 90,798,464 | 0.4818 |
| 2013 | 3,140 | 34,549 | 91,667,214 | 0.3769 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Distribution of County-level Observed Rates in Reference Population Per 1000** | | | | | | | | |
| **Year** | **Number of**  **Counties** | **(p=percentile)2** | | | | | | |
| **Mean** | **SD** | **p5** | **p25** | **Median** | **p75** | **p95** |
| 2009 | 3,135 | 0.50 | 0.63 | 0.00 | 0.00 | 0.38 | 0.70 | 1.51 |
| 2010 | 3,138 | 0.46 | 0.63 | 0.00 | 0.00 | 0.34 | 0.64 | 1.38 |
| 2011 | 3,141 | 0.40 | 0.53 | 0.00 | 0.00 | 0.29 | 0.56 | 1.25 |
| 2012 | 3,139 | 0.67 | 17.85 | 0.00 | 0.00 | 0.26 | 0.51 | 1.07 |
| 2013 | 3,140 | 0.28 | 0.37 | 0.00 | 0.00 | 0.18 | 0.42 | 0.90 |

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

1The observed rate refers to the total rate for all observations included in the reference population data (numerator) divided by the total combined eligible population for the county in the reference population data (denominator). For PQI 15 this includes population ages 18-39 years.

2The distribution of area rates reports the mean and standard deviation (SD) of the observed rates for all counties included in the dataset, as well as the observed rate for counties in the 5th, 25th, 50th (median), 75th, and 95th percentile.

**1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

See 1.5 (Table 1)

**1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below**.

The reliability and performance discrimination testing was completed using 2013 HCUP data. Validity testing was completed using 2012 HCUP data. The AHRQ QI PQI 2012 reference population has 34,440,38 discharges and the AHRQ QI PQI 2013 reference population includes 29,891,024. Annual testing of rates, reliability and performance discrimination showed little change in performance between the two reference populations.

**1.8** What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We use SDS variables for two purposes: risk adjustment and validity analyses. Because the measures are applied at the county level, we use variables that describe the make-up of the county population.

Risk adjustment: Two risk models are available; one includes age and sex make-up of the county, the other also includes the percent of households falling below the federal poverty level. These data are obtained from the US Census.

Validity analyses: In addition to the risk models we used county level demographic variables in testing. Sociodemographic variables were combined using principal component analysis (PCA) into a single socioeconomic status (SES) variable, defined at the county level. Including unemployment rate, percent adults below the federal poverty line, percent of adults aged 25-44 with some post-secondary education, percent of population not proficient in English and median household income.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2a2. RELIABILITY TESTING**

***Note****: If accuracy/correctness (validity) of data elements was empirically tested*, *separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter “see section 2b2 for validity testing of data elements”; and skip 2a2.3 and 2a2.4.*

**2a2.1. What level of reliability testing was conducted**? (*may be one or both levels*)  
 **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)   
 **Performance measure score** (e.g., *signal-to-noise analysis*)  
  
**2a2.2. For each level checked above, describe the method of reliability testing and what it tests** (*describe the steps―do not just name a method; what type of error does it test; what statistical analysis was used*)

***Signal-to-noise***

The signal-to-noise ratio refers to the entire population of US counties, comparing the degree to which rates are different from county to county (the signal) to how stable the rates are within counties (the noise).   This metric is a stringent measure of reliability that takes into account the observed distribution of rates within a reference population. An indicator with a low signal-to-noise ratio may not be able to distinguish differences in performance between counties, or may identify differences inconsistently within the same time period. An indicator with a high signal-to-noise ratio will be more likely to consistently distinguish performance differences between counties (e.g. one county performs better than others).

The signal-to-noise ratio is estimated for each county.  The overall signal-to-noise estimate is an average of county-level signal to noise ratios weighted by county size. County size is calculated as the eligible population for PQI 15 (population 18-39 years). Weighting by county size reduces the impact of counties that have very small denominators (the number of patients at risk).

Because the signal-to-noise ratio quantifies the ability to consistently discriminate one county’s performance from the other counties in the population, it is sensitive to the distribution of county sizes as well as the distribution of observed rates in the reference population. If the counties in a population all have performance in a narrow range, it is more difficult to reliably distinguish between counties’ performance than when county performance is spread out over a much wider range.   For example, if all counties have nearly perfect performance, it will be impossible to distinguish between them.  As a consequence, if the distribution of county rates changes over time, the signal-to-noise ratio will also change.

There is no universally accepted threshold of “adequate” signal to noise ratio. Different methods of calculating reliability and signal-to-noise result in different distributions of reliability scores. In addition, “adequate” depends on the specific application and judgment of the user. For instance, if a complication such as mortality is very important (e.g. leads to great harm to the patient) a lower reliability may be acceptable. However, the AHRQ QI program generally considers ratios between 0.4 – 0.8 as acceptable. It is rare to achieve reliability above 0.8. To account for the uncertainty (noise) in a county’s performance due to reliability concerns stemming from low volume, smoothed rates can be calculated.

**2a2.3. For each level of testing checked above, what were the statistical results from reliability testing**? (e*.g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

**Table 2a. Signal-to-Noise Ratio by Size Decile for PQI 15 Asthma in Younger Adults Admission Rate**

|  |  |  |  |
| --- | --- | --- | --- |
| **Size Decile** | **Number of Counties** | **Avg. Number of Qualifying Population per County** **in Decile** | **Avg. Signal-to-Noise Ratio for Counties** **in Decile** |
| 1 | 314 | 1178.6 | 0.19649 |
| 2 | 314 | 2542 | 0.31188 |
| 3 | 314 | 3840.7 | 0.40404 |
| 4 | 314 | 5181.9 | 0.47145 |
| 5 | 314 | 6922.9 | 0.54897 |
| 6 | 314 | 9426.5 | 0.62095 |
| 7 | 314 | 12906.6 | 0.69365 |
| 8 | 314 | 19998.8 | 0.77833 |
| 9 | 314 | 38396.8 | 0.87051 |
| 10 | 314 | 191539.1 | 0.95615 |
| Overall | 3140 | 29193.4 | 0.7463 |

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. [www.hcup-us.ahrq.gov/sidoverview.jsp.](http://www.hcup-us.ahrq.gov/sidoverview.jsp) (AHRQ QI Software Version 6.0)

**Table 2b. SES Signal-to-Noise Ratio by Size Decile for PQI 15 Asthma in Younger Adults Admission Rate**

|  |  |  |  |
| --- | --- | --- | --- |
| **Size Decile** | **Number of Counties** | **Avg. Number of Qualifying Population per County** **in Decile** | **Avg. Signal-to-Noise Ratio for Counties** **in Decile** |
| 1 | 314 | 1178.6 | 0.18787 |
| 2 | 314 | 2542 | 0.29982 |
| 3 | 314 | 3840.7 | 0.39039 |
| 4 | 314 | 5181.9 | 0.45728 |
| 5 | 314 | 6922.9 | 0.53481 |
| 6 | 314 | 9426.5 | 0.60742 |
| 7 | 314 | 12906.6 | 0.68140 |
| 8 | 314 | 19998.8 | 0.76836 |
| 9 | 314 | 38396.8 | 0.86400 |
| 10 | 314 | 191539.1 | 0.95377 |
| Overall | 3140 | 29193.4 | 0.74139 |

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. [www.hcup-us.ahrq.gov/sidoverview.jsp.](http://www.hcup-us.ahrq.gov/sidoverview.jsp) (AHRQ QI Software Version 6.0)

**2a2.4 What is your interpretation of the results in terms of demonstrating reliability**? (i*.e., what do the results mean and what are the norms for the test conducted?*)

The indicator demonstrates good reliability with a signal-to-noise ratio of 0.75. Reliability meet threshold for all counties, except for the smallest counties with eligible populations under approximately 3800 individuals. Smoothed rates, which are recommended for all counties (and are implemented in the AHRQ software), address any reliability concerns for the smallest counties. When SES is added to the risk adjustment, the reliability remains adequate at 0.74.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b2. VALIDITY TESTING**

**2b2.1. What level of validity testing was conducted**? (*may be one or both levels*)  
 **Critical data elements** (*data element validity must address ALL critical data elements*)

**Performance measure score**

**Empirical validity testing** **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

**2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests** (*describe the steps―do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

***Systematic Assessment of Face Validity***

In 2008-2009, we convened clinical panels to assess the use of the PQI across a wide range of applications, one of which was comparative reporting of area level rates.[[4]](#footnote-4) We solicited nominations from national professional organization to form four clinical expert groups for a total of 73 panelists. We utilized a hybrid approach for the panel review, using two review processes, which were conducted simultaneously with information exchange between the two panels and a third and fourth panel that conducted specific reviews based on specialty. The development of this hybrid process builds from the experiences in previous panel evaluations of QI modules. The panel process that has been employed during the development of the PSIs, the PDIs and the validation of the IQIs is based on the RAND-UCLA Appropriateness Method and is termed a “nominal group” panel. The approach allowed for a wider range of input is fully described elsewhere.[[5]](#footnote-5)

Panelists rated the indicator on appropriateness of use after completing a 14 item questionnaire. The questionnaire evaluated the face validity of the indicators, the panelists’ perspectives on bias and potential for gaming, and the overall usefulness of the indicators when applied at one of three aspects of the health care system: area, payer and large provider organizations, for one of three purposes: internal quality improvement, comparative reporting (either public or not), and pay for performance.

Support was defined as follows:

* Full support for use: Median score of 7-9 without disagreement
* Some concern regarding use: Median score of 4-6.9 regardless of agreement status
* General support with some concerns regarding use due to disagreement: Median score of 7-9 with disagreement
* Major concern regarding use: Median score or 1-3.9 regardless of agreement status

***Empirical Validity***

We sought to assess the relationship of county-level hospital admission rate for COPD with county level measures of socioeconomic status (SES) and community environment, heath behaviors and individual risk factors (i.e. smoking, physical activity and obesity), and access to quality care measures (i.e. primary care physician and other primary care provider density, diabetes screening testing, uninsurance). SES and community environment variables included unemployment rate, poverty rate, some college rate, English proficiency rate, median household income, food environment index, access to exercise, violent crime rate, air pollution, severe housing problems and rural status derived from the Community Health Rankings data and the American Community Survey. Because of the high number of relevant variables, we aimed to improve the interpretability of results. Principal component analysis (PCA) was used to create three composite variables for county characteristics: SES and environment (SES/E), health behaviors and personal risk factors (HB) and access to care (AC) using. The variables included in the analyses (described earlier in this paragraph) were determined a priori based on clinical and subject matter expertise. Because the relationship between the individual variables within each factor is unlikely to be consistent across all counties, we retained multiple components within each factor that explained about 70% of the variance for that factor. We also estimated prevalence based on the CDC Behavioral Risk Factor Surveillance System (BRFSS) COPD Model-based County Prevalence.

**2b2.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test*)

***Face validity: Clinical Panel Review***

During the clinical panel review both the Delphi supported the measure with some concern and Nominal panel fully supported the measure. Other topics discussed by the panel included:

* Panelists endorsed restricting the indicator to patients less than 40 years of age. Panelists felt that combining the COPD and Asthma numerator for patients 40 years and older would eliminate the diagnostic uncertainly between asthma and COPD in older patients, and thus provide a cleaner measure. COPD diagnoses in cases under 40 years of age are rare, and therefore, cases of patients less than 40 years are more likely to be true cases of asthma.
* The panel generally felt this indicator reflects issues related to access to quality outpatient care, including affordability of medication and education on proper inhaler use.
* Patient adherence to treatment recommendations remains an issue as with all chronic conditions.
* As with all chronic conditions, comorbidities and disease severity are of concern. For the asthma indicator, other respiratory conditions, infectious disease and cardiovascular conditions are of particular concern. Along with risk factors such as age, race/ethnicity, socioeconomic status, and smoking rates, panelists emphasized that environmental factors may affect admissions rates for this indicator. These environmental factors include pollution levels, altitude, allergens, housing conditions, and occupational exposures from local industries.
* Panelists also generally agreed that the high cost and complicated protocols for inhaler medications present major barriers to patient adherence to treatment recommendations. They further agreed that it is within the ability of the healthcare system to mitigate these barriers, including by providing high quality education on medication needs and inhaler use.
* Panelists felt that this indicator may also reflect some amount of “social” hospital admissions. In other words, cases in which the physician determines that social support or the home environment are insufficient for recovery outside of the hospital.
* The presence of observation units may affect admission rates.

***Empirical Validity***

In a negative binomial model (see section 2b2.2), prevalence, health behaviors (HB) and SES/Environment were statistically significant predictors (p<.0001) (see Table 3a). Access to care (AC) was not significant when HB and SES/E are included in the model. However, categorizing the variables into interpretable groups (HB, AC, and SES/E) resulted in significant collinearity in the models. In particular, there was a correlation of magnitude 0.77 between IHB and one of the components for SES/E and correlations of magnitude between 0.40 and 0.50 between the components for AC and the other two components for SES/E (see Table 3b). Hence the relative importance of those factors should be interpreted with caution.

**Table 3a: P-values from negative binomial model for PQI 15 as a function of prevalence and the various groups of principal components**

|  | **Prevalence** | **Health Behaviors** | **Access to Care** | **SES/Envr** |
| --- | --- | --- | --- | --- |
| PQI 15 | 0.0070 | <.0001 | 0.31 | <.0001 |

**Table 3b: Correlation matrix of principal components (PC) used in negative binomial models.**

            HB AC\_1 AC\_2 SES\_1 SES\_2 SES\_3

Health

Behaviors(HB) 1.00   -0.20   -0.13        0.22       **-0.77**0.18

Access to

Care (AC\_1)  -0.20    1.00    0.00      **0.50**0.13        **0.40**

Access to

Care (AC\_2)  -0.13    0.00    1.00        **0.44**0.24        0.24

Socio-

Economic

Status

(SES\_1)    0.22   -0.50    0.44        1.00        0.00        0.00

Socio-

Economic

Status

(SES\_2) -0.77    0.13    0.24        0.00        1.00        0.00

Socio-

Economic

Status

(SES\_3)  0.18    0.40    0.24        0.00        0.00        1.00

Note that the groups of variables corresponding to health behaviors (HB) were reduced to one principal component, access to care (AC) to two, and SES to three. Correlations with magnitude larger than 0.4 are bolded.

**2b2.4. What is your interpretation of the results in terms of demonstrating validity?** (i.e., what do the results mean and what are the norms for the test conducted?)

The clinical panel noted that this indicator overall was useful. They identified specific actions that could improve rates, especially from a population health perspective, such as access to medications, patient education, reduction of risk factors, such as environmental exposure to pollution or allergens and smoking. However, it is important, as with many population health indicators to acknowledge the complex factors influencing the indicators.

SES explained the most variance, however moderate correlation between socioeconomic status and other factors suggest that it is difficult to fully ascertain the unique contribution of each factor on asthma hospitalizations. The disparities table (see Table 2 in the supplemental files) demonstrates that zip codes in the highest income quartile have 31% lower admission rates than those in the lowest income quartile.

From a population health perspective such disparities argue for the importance of the indicator in capturing poor outcomes for vulnerable populations. Further, taking a population health perspective, efforts to decrease individual risk factors such as obesity, smoking and limited physical exercise (the variables included in the health behaviors factor) may decrease prevalence and exacerbations of asthma.

**2b3. EXCLUSIONS ANALYSIS**

**NA**  **no exclusions — *skip to section*** [***2b4***](#section2b4)

**2b3.1. Describe the method of testing exclusions and what it tests** (*describe the steps―do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Using the 2013 data from 40 states, we examined the percent of potential denominator cases excluded by each criterion as listed in the measure specifications.

**2b3.2. What were the statistical results from testing exclusions**? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

A total of 262 discharges were excluded due to diagnoses of cystic fibrosis and anomalies of the respiratory system. Removing this exclusion would increase the numerator count by 0.13%. The denominator does not change. Although discharges transferred into a hospital are excluded, these encounters are captured in the area level numerator via the originating hospitalization.

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. [www.hcup-us.ahrq.gov/sidoverview.jsp.](http://www.hcup-us.ahrq.gov/sidoverview.jsp) (AHRQ QI Software Version 6.0)

**2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results?** (*i.e., the value outweighs the burden of increased data collection and analysis.*  *Note:* ***If patient preference is an exclusion****, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The exclusion of cystic fibrosis and anomalies of the respiratory system has been retained to increase the face validity of the measure and to align the measure with the older adult and pediatric PQI/PDI for asthma/COPD admissions, although the impact on the numerator is minor for PQI15. Patients meeting the exclusion are identifiable without additional burden using the same data as is used to identify numerator qualifying discharges.

The indicator excludes patients with severe chronic respiratory diseases, because asthma in the context of these diseases differs clinically from the patients with asthma alone.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES**  
***If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section*** [***2b5***](#section2b5)***.***

**2b4.1. What method of controlling for differences in case mix is used?**

**No risk adjustment or stratification**

**Statistical risk model with** 23 **risk factors**

**Stratification by** Click here to enter number of categories **risk categories**

**Other,** Click here to enter description

**2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities**.   
Not applicable

**2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk** (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

***Sociodemographic Factors***

The risk model includes age and gender of the population.

We considered multiple income risk factors for inclusion in the model, including % of households under the federal poverty level, household income, median % of poverty level (e.g. 200% of federal poverty level). The Public Health Disparities Geocoding Project has completed extensive evaluation of alternative income variables and has demonstrated that the percent poverty variable consistently detects expected gradients in health across health outcomes, is widely available, has a low rate of missing data even at the census tract level[[6]](#footnote-6). Our team has explored alternative income variables that do not outperform the poverty level variable (data not shown).

We also considered a conceptual model that acknowledged the impact of community factors such as clean air, exposure to tobacco smoke, access to healthy foods, open space for exercise along with community norms and beliefs. However, these are factors that are difficult to measure within the framework of the AHRQ Qis, i.e., use with administrative data. These community factors can impact prevention of asthma and self-care for asthma, which in turn can impact hospitalization rates. Poverty can be a mitigating factor, inasmuch as impoverished communities are more likely to experience housing and food insecurity[[7]](#footnote-7), air pollution[[8]](#footnote-8), occupational exposure and have higher smoking rates. The relationship between environment, income and health is complex and the mechanism is not fully understood.

**2b4.4a. What were the statistical results of the analyses used to select risk factors?**

The process to select risk factors is described in the AHRQ QI Empirical Methods report. The results of the analyses are provided in the PQI Parameter Estimates document. Both documents are available to reviewers in the supporting materials. The results of the analyses are provided in the tables below as well as on the submitted excel spreadsheet.

There are several steps involved in estimating the QI risk-adjustment models.

1. Construct candidate covariates
2. Select model covariates
3. Estimate the models
4. Evaluate the models

Covariates are coded for each discharge record based on the data elements, data values, and logic described in the technical specifications and the appendices of the risk-adjustment coefficient tables. For a given covariate, if the discharge meets the technical specification for that covariate a value of “1” is assigned to the discharge level covariate data element. Otherwise a value of “0” is assigned to the discharge level covariate data element.

**Table 4a. Risk Adjustment Coefficients, for PQI 15 Asthma in Younger Adults Admission Rate**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **Label** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > Chi-Square** |
| INTERCEPT |  | 1 | -8.047392 | 0.0178685 | 202829.87 | <.0001 |
| SEX | Female | 1 | 0.837454 | 0.0213734 | 1535.2290 | <.0001 |
| AGE | Male, Age 18-24 | 1 | -0.382994 | 0.0250408 | 233.93018 | <.0001 |
| AGE | Male, Age 25-29 | 1 | -0.336554 | 0.0268152 | 157.52444 | <.0001 |
| AGE | Male, Age 30-34 | 1 | -0.226800 | 0.0262008 | 74.930256 | <.0001 |
| AGE | Male, Age 35-39 |  | Referent | . | . | . |
| AGE | Female, Age 18-24 | 1 | -0.353536 | 0.0310396 | 129.72796 | <.0001 |
| AGE | Female, Age 25-29 | 1 | -0.190547 | 0.0327403 | 33.872088 | <.0001 |
| AGE | Female, Age 30-34 | 1 | -0.042780 | 0.0314820 | 1.8465765 | 0.1742 |
| AGE | Female, Age 35-39 |  | Referent | . | . | . |
| c-statistic=0.56 | | | | | | |

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

**Table 4b. SES Risk Adjustment Coefficients, for PQI 15 Asthma in Younger Adults Admission Rate**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **Label** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > Chi-Square** |
| INTERCEPT |  | 1 | -8.19743673 | 0.025177701 | 106004.424 | <.0001 |
| SEX | Female | 1 | 0.83749199 | 0.021373641 | 1535.33903 | <.0001 |
| AGE | Male, Age 18-24 | 1 | -0.39606813 | 0.025049744 | 249.996089 | <.0001 |
| AGE | Male, Age 25-29 | 1 | -0.34599236 | 0.026817633 | 166.453192 | <.0001 |
| AGE | Male, Age 30-34 | 1 | -0.23108600 | 0.026201555 | 77.7845207 | <.0001 |
| AGE | Male, Age 35-39 |  | Referent | . | . | . |
| AGE | Female, Age 18-24 | 1 | -0.35522758 | 0.031039922 | 130.970085 | <.0001 |
| AGE | Female, Age 25-29 | 1 | -0.19156186 | 0.03274053 | 34.2331361 | <.0001 |
| AGE | Female, Age 30-34 | 1 | -0.04268203 | 0.031482201 | 1.83806134 | 0.1752 |
| AGE | Female, Age 35-39 |  | Referent | . | . | . |
| POVCAT | Poverty Decile 2 | 1 | -0.00381111 | 0.026723409 | 0.02033851 | 0.8866 |
| POVCAT | Poverty Decile 3 | 1 | -0.02322692 | 0.02655037 | 0.76531867 | 0.3817 |
| POVCAT | Poverty Decile 4 | 1 | -0.01808578 | 0.026235351 | 0.47522664 | 0.4906 |
| POVCAT | Poverty Decile 5 | 1 | 0.16097745 | 0.025270753 | 40.5782915 | <.0001 |
| POVCAT | Poverty Decile 6 | 1 | 0.2223969 | 0.024676127 | 81.2275687 | <.0001 |
| POVCAT | Poverty Decile 7 | 1 | 0.20873420 | 0.024754258 | 71.1029161 | <.0001 |
| POVCAT | Poverty Decile 8 | 1 | 0.14333891 | 0.025036027 | 32.7791313 | <.0001 |
| POVCAT | Poverty Decile 9 | 1 | 0.11142012 | 0.025997461 | 18.3681488 | <.0001 |
| POVCAT | Poverty Decile10 (Highest percent poverty)1 | 1 | 0.56469829 | 0.023119412 | 596.594715 | <.0001 |
| c-statistic=0.5453 | | | | | | |

1Deciles are based on the percentage of households under the federal poverty level (FPL).

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp. (AHRQ QI Software Version 6.0)

**2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)**

See above results from validity testing in Section 2b2.3.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

**2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach** (*describe the steps―do not just name a method; what statistical analysis was used*)

This analysis evaluates how strongly the risk adjustment model is associated with the event of interest (i.e. admission for asthma). The measure of discrimination, how well the risk adjustment model distinguishes events from non-events, is the c-statistic. The c-statistic is computed by assigning each observation a predicted probability of the outcome from the risk-adjustment model based on the value of the observations covariates from the risk-adjustment model. Two copies of the dataset are sorted, first from highest to lowest predicted probability and second from lowest to highest predicted probability. This creates a set of pairs of observations. Pairs that consist of one event and one non-event (discordant pairs) are kept and concordant pairs are discarded. The c-statistic is a measure of the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk-adjustment model than the non-event. C-statistics above 0.70 and below 0.80 have moderate discrimination. Above 0.80 the discrimination is high. We did not employ common “goodness of fit” tests because these tests tend to not be informative with large samples.

We also evaluated the calibration of the risk adjustment model by evaluating how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

*Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below*.  
***If stratified, skip to*** [***2b4.9***](#question2b49)

**2b4.6. Statistical Risk Model Discrimination Statistics** (*e.g., c-statistic, R-squared*)**:**

**Table 5a. Age-sex Risk adjustment Model Discrimination and Calibration, for PQI 15 Asthma in Younger Adults Admission Rate**

| **Predicted Rate Decile** | **Number of Discharges**  **per Decile** | **Predicted** **Rate** | **Observed** **Rate** |
| --- | --- | --- | --- |
| 1 | 12,349,460 | 0.000218 | 0.000228 |
| 2 | 9,715,818 | 0.000226 | 0.00021 |
| 3 | 9,514,092 | 0.000244 | 0.000235 |
| 4 | 8,412,305 | 0.000282 | 0.000306 |
| 5 | 8,817,911 | 0.000329 | 0.000302 |
| 6 | 11,333,912 | 0.000354 | 0.000371 |
| 7 | 8,192,981 | 0.000432 | 0.000397 |
| 8 | 8,340,546 | 0.00052 | 0.000585 |
| 9 | 7,664,290 | 0.000622 | 0.00055 |
| 10 | 7,325,882 | 0.000739 | 0.000772 |
| C-Statistic | 0.56 |  |  |

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. [www.hcup-us.ahrq.gov/sidoverview.jsp.](http://www.hcup-us.ahrq.gov/sidoverview.jsp) (AHRQ QI Software Version 6.0)

**Table 5b. Age-sex and SES Risk adjustment Model Discrimination and Calibration, for PQI 15 Asthma in Younger Adults Admission Rate**

| **Predicted Rate Decile** | **Number of Discharges**  **per Decile** | **Predicted** **Rate** | **Observed** **Rate** |
| --- | --- | --- | --- |
| 1 | 10,727,468 | 0.000189 | 0.000195 |
| 2 | 10,919,256 | 0.000218 | 0.000207 |
| 3 | 10,888,415 | 0.000247 | 0.000258 |
| 4 | 9,955,771 | 0.000288 | 0.000281 |
| 5 | 8,509,855 | 0.000331 | 0.000329 |
| 6 | 10,480,717 | 0.000364 | 0.000374 |
| 7 | 8,302,290 | 0.000448 | 0.000443 |
| 8 | 7,625,628 | 0.000524 | 0.000525 |
| 9 | 7,726,889 | 0.000628 | 0.000602 |
| 10 | 6,530,908 | 0.000826 | 0.000848 |
| C-Statistic | 0.5453 |  |  |

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. [www.hcup-us.ahrq.gov/sidoverview.jsp.](http://www.hcup-us.ahrq.gov/sidoverview.jsp) (AHRQ QI Software Version 6.0)

**2b4.7. Statistical Risk Model Calibration Statistics** (*e.g., Hosmer-Lemeshow statistic*):   
See Table 5 in 2b4.6

**2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves**:  
See Table 5 in 2b4.6

**2b4.9. Results of Risk Stratification Analysis**:

Not applicable

**2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)?** (i*.e., what do the results mean and what are the norms for the test conducted*)

A model that is well calibrated will have observed values similar to predicted values across the predicted value deciles. This indicator is well calibrated, as the observed to predicted values across the deciles range between 0.88– 1.12. The discrimination is low with a c-statistic of 0.56, presumably due to the limited predictors included. Addition of SES to the model slightly improves the calibration, with observed to predicted values ranging across the deciles range between 0.95 – 1.03. The c-statistic does not change substantially.

**2b4.11.** **Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE**

**2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified** (*describe the steps―do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

This analysis assesses the probability that a county is higher or lower than a benchmark or threshold, given county size. It reflects whether the indicator can discriminate the best performing counties from the lower performing counties.

For this analysis, “benchmark” refers to the smoothed indicator rate based on the 20th percentile of the reference population (i.e., 20% of counties have a lower admission rate or better performance). “Threshold” refers to the indicator rate based on the 80th percentile (i.e., 80% have lower mortality or better performance).

The analysis is reported by size decile, based on the denominator cases, demonstrating performance across counties of various sizes. Each county is assumed to have an underlying distribution of smoothed rates that follows a Gamma distribution. The parameters of a Gamma distribution are shape and scale. For each county the shape is calculated as ((smoothed rate)2/ smoothed rate variance), and the scale is calculated as (smoothed rate variance / smoothed rate). The smoothed rate variance (aka posterior variance) is calculated as the signal variance – (reliability weight \* signal variance). The reliability weight is calculated as (signal variance / (signal variance + noise variance)). Counties are ranked by size and grouped into 10 equal categories of size (deciles). The Benchmark and Threshold are compared to the Gamma distribution of the smoothed rates for each county to determine if the county rate is better or worse than the Benchmark and Threshold rates with 95% probability. This provides a 95% confidence interval for the Benchmark and Threshold rate.

Table 6 reports the proportion of counties above (better than) and below (worse than) the Benchmark and Threshold rates and the proportion not classified as either above or below. The proportion of counties not classified as either better or worse have rates that fall within the 95% confidence interval.

**2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities?** (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

**Table 6a. Performance Categories by County Size Decile, for PQI 15 Asthma in Younger Adults Admission Rate**

|  |  |  | **Benchmark** | | | | **Threshold** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Size Decile** | **Number of Counties** | **Average Number of Denominator**  **Discharges Per County** | **Proportion Better** | **Proportion Worse** | **Proportion Unclassified** | | **Proportion Better** | **Proportion Worse** | **Proportion Unclassified** |
| 1 | 314 | 1,178.60 | 0.0000 | 0.0669 | | 0.9331 | 0.0000 | 0.0000 | 1.0000 |
| 2 | 314 | 2,542.00 | 0.0000 | 0.1497 | | 0.8503 | 0.0000 | 0.0127 | 0.9873 |
| 3 | 314 | 3,840.70 | 0.0000 | 0.2229 | | 0.7771 | 0.0000 | 0.0159 | 0.9841 |
| 4 | 314 | 5,181.90 | 0.0000 | 0.1561 | | 0.8439 | 0.1019 | 0.0064 | 0.8917 |
| 5 | 314 | 6,922.90 | 0.0000 | 0.3089 | | 0.6911 | 0.3567 | 0.0350 | 0.6083 |
| 6 | 314 | 9,426.50 | 0.0000 | 0.2675 | | 0.7325 | 0.4395 | 0.0255 | 0.5350 |
| 7 | 314 | 12,906.60 | 0.0000 | 0.3535 | | 0.6465 | 0.5605 | 0.0287 | 0.4108 |
| 8 | 314 | 19,998.80 | 0.0000 | 0.4268 | | 0.5732 | 0.5955 | 0.0541 | 0.3503 |
| 9 | 314 | 38,396.80 | 0.0000 | 0.5287 | | 0.4713 | 0.6592 | 0.0510 | 0.2898 |
| 10 | 314 | 191,539.10 | 0.0637 | 0.7070 | | 0.2293 | 0.6879 | 0.1656 | 0.1465 |
| Overall | 3,140 | 29,193.40 | 0.0064 | 0.3188 | | 0.6748 | 0.3401 | 0.0395 | 0.6204 |

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. [www.hcup-us.ahrq.gov/sidoverview.jsp.](http://www.hcup-us.ahrq.gov/sidoverview.jsp) (AHRQ QI Software Version 6.0)

**Table 6b. Performance Categories by County Size Decile, for PQI 15 Asthma in Younger Adults Admission Rate**

|  |  |  | **Benchmark** | | | **Threshold** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Size Decile** | **Number of Counties** | **Average Number of Denominator**  **Discharges Per County** | **Proportion Better** | **Proportion Worse** | **Proportion Unclassified** | **Proportion Better** | **Proportion Worse** | **Proportion Unclassified** |
| 1 | 314 | 1,178.60 | 0.0000 | 0.0541 | 0.9459 | 0.0000 | 0.0000 | 1.0000 |
| 2 | 314 | 2,542.00 | 0.0000 | 0.1497 | 0.8503 | 0.0000 | 0.0127 | 0.9873 |
| 3 | 314 | 3,840.70 | 0.0000 | 0.2070 | 0.7930 | 0.0000 | 0.0159 | 0.9841 |
| 4 | 314 | 5,181.90 | 0.0000 | 0.1433 | 0.8567 | 0.0892 | 0.0096 | 0.9013 |
| 5 | 314 | 6,922.90 | 0.0000 | 0.3057 | 0.6943 | 0.3408 | 0.0350 | 0.6242 |
| 6 | 314 | 9,426.50 | 0.0000 | 0.2611 | 0.7389 | 0.4299 | 0.0287 | 0.5414 |
| 7 | 314 | 12,906.60 | 0.0000 | 0.3439 | 0.6561 | 0.5350 | 0.0287 | 0.4363 |
| 8 | 314 | 19,998.80 | 0.0000 | 0.4140 | 0.5860 | 0.5924 | 0.0541 | 0.3535 |
| 9 | 314 | 38,396.80 | 0.0000 | 0.5127 | 0.4873 | 0.6561 | 0.0510 | 0.2930 |
| 10 | 314 | 191,539.10 | 0.0732 | 0.6975 | 0.2293 | 0.6879 | 0.1656 | 0.1465 |
| Overall | 3,140 | 29,193.40 | 0.0073 | 0.3089 | 0.6838 | 0.3331 | 0.0401 | 0.6268 |

Source: HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2013. Agency for Healthcare Research and Quality, Rockville, MD. [www.hcup-us.ahrq.gov/sidoverview.jsp.](http://www.hcup-us.ahrq.gov/sidoverview.jsp) (AHRQ QI Software Version 6.0)

**2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities?** (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

This indicator has moderate discrimination to identify low performing counties for most counties; 38% of counties can be classified as better or worse than the threshold (the percentage classified as either above or below the threshold). The indicator has moderate discrimination to identify high performing counties; 32% of counties can be classified as better or worse than the benchmark.Performance discrimination remains moderate when adding SES to risk adjustment.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS**

***If only one set of specifications, this section can be skipped.***

**Note***: This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

Not applicable

**2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications** (*describe the steps―do not just name a method; what statistical analysis was used*)  
Not applicable

**2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications?** (*e.g., correlation, rank order*)  
Not applicable

**2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications?** (i*.e., what do the results mean and what are the norms for the test conducted*)  
Not applicable

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

**2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps―do not just name a method; what statistical analysis was used*)

The AHRQ QIs use frequently reported administrative data variables. PQI 15 excludes cases with missing discharge disposition, age, sex, discharge quarter, discharge year, and principal diagnosis. These variables are required for indicator construction and are required of all hospital discharge records. The rate of missing data for each variable is available by state and year from the AHRQ HCUP website (<http://www.hcup-us.ahrq.gov/cdstats/cdstats_search.jsp>).

**2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data?** (*e.g.,**results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

For these variables, rates of missing data are typically less than 1% of the state database. It is unlikely the bias would occur from such a low rate of missing data.

**2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i*.e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Exclusion of cases for missing data is appropriate.

1. HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). 2009-2013. Agency for Healthcare Research and Quality, Rockville, MD. [www.hcup-us.ahrq.gov/sidoverview.jsp](http://www.hcup-us.ahrq.gov/sidoverview.jsp). (AHRQ QI Software Version 6.0) [↑](#footnote-ref-1)
2. “County Health Rankings & Roadmaps”. University of Wisconsin Population Health Institute, 2014. http://www.countyhealthrankings.org. Accessed 26 Jan. 2015. [↑](#footnote-ref-2)
3. "American Community Survey (ACS)." *https://www.census.gov/programs-surveys/acs/data.html*. Accessed 2015. [↑](#footnote-ref-3)
4. Davies SM, McDonald KM, Schmidt E, Schultz E, Geppert J, Romano, PS. 'Expanding the Uses of AHRQ's Prevention Quality Indicators: Validity from the Clinician Perspective'. *Medical Care.* 2011 Aut;49(8):479-85. [↑](#footnote-ref-4)
5. Davies S, Romano PS, Schmidt EM, Schultz E, Geppert JJ, McDonald KM. Assessment of a Novel Hybrid Delphi and Nominal Groups Technique to Evaluate Quality Indicators. *Health services research.* 2011;46(6 Pt 1):2005-2018. [↑](#footnote-ref-5)
6. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian S. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures-the public health disparities geocoding project. *American journal of public health.* 2003;93(10):1655-1671. [↑](#footnote-ref-6)
7. Larson NI, Story MT, Nelson MC. Neighborhood environments: disparities in access to healthy foods in the US. American journal of preventive medicine. 2009;36(1):74-81. e10. [↑](#footnote-ref-7)
8. Bell ML, Ebisu K. Environmental inequality in exposures to airborne particulate matter components in the United States. *Environmental health perspectives.* 2012;120(2):1699-1704. [↑](#footnote-ref-8)