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

**Measure Number** (*if previously endorsed*)**:** 0286

**Measure Title**: Aspirin at Arrival

**Date of Submission**: 1/31/2014

**Type of Measure:**

|  |  |
| --- | --- |
| ☐ Composite – ***STOP – use composite testing form*** | ☐ 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 specifications** (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). |

|  |
| --- |
| **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 decisionmaking) 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 that influence the measured outcome (but not factors related to disparities in care or the quality of care) 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.** 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, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.  **16.** 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*). The measure population as reported in the QIO Clinical Data Warehouse (CDW) included 114,246 cases from 2,713 short-term acute care hospitals nationwide. These 2,713 hospitals do not include military or Veterans Affairs hospitals. Additionally, Critical Access Hospitals (CAH) are excluded in this report because their participation in the Outpatient Quality Reporting program is not mandatory. The 114,246 cases were abstracted by the individual hospitals or their vendors and the data was submitted to the CDW.

**1.3. What are the dates of the data used in testing**? The measure period is January 1, 2012 to December 31, 2012.

**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*)

|  |  |
| --- | --- |
| **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: Click here to describe | ☐ other: Click here to describe |

**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*) A validation sample of these cases was selected in two stages. First, 323 hospitals were randomly selected. Second, up to 12 cases were randomly selected from each of these 323 hospitals. Critical access hospitals (CAH) were not included in the sampling of facilities. The final sample included 763 cases out of the original 114,246 cases submitted to the CDW during the measurement period.

*Chart Abstraction:* Both the original CDW dataset and the sample dataset were obtained from direct medical chart abstraction. The original population dataset was abstracted by the hospitals or their vendors. The sampled validation dataset was re-abstracted by the CMS Clinical Data Abstraction Center (CDAC) using exactly the same medical charts. CDAC is a CMS contractor center that has specialized in medical chart abstraction for the last fifteen years. The CDAC-abstracted data is considered “gold standard” for the purpose of this analysis.

See Table 1 for hospital and patient characteristics.

**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 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**. N/A

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

**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.* Per NQF comments received on 6/10/13, it is no longer necessary to report the results of the reliability testing when the results of the validity testing of individual data elements are reported.

**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*)

**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*)

**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?*)

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

**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) Validity Test*: There are 11 critical data elements for the measure. We conducted validity tests on all 11 critical data elements. For each data element, we calculated the raw agreement rate between data from the hospital chart abstractor and the CDAC re-abstractor. We reported Kappa statistics for the categorical data elements with binary Yes/No values. Kappa is a measure of inter-rater agreement that accounts for abstractors’ agreement by chance alone. It is standardized to lie on a -1 to 1 scale, where 1 is perfect agreement, 0 is exactly what would be expected by chance, and negative values indicate agreement less than chance, i.e., potential systematic disagreement between the abstractors. A common scale is used to interpret Kappa statistics: 0.01–0.20 is slight agreement; 0.21– 0.40 is fair agreement; 0.41–0.60 is moderate agreement; 0.61–0.80 is substantial agreement; 0.81–0.99 is almost perfect agreement.

**2b2.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test*) See tables at end of this document.

**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?*) Table 2 summarizes the results of the validity test of 11 data elements. Overall, the agreement rates were high. The agreement rates for all data elements except two were higher than 90%. We should point out that ECG had a high agreement rate (99%) but a low Kappa (0.25). The potential reason for the discrepancy between the agreement and Kappa is that ECG=N (an ECG was not performed) is rare event. The Kappa statistic is affected by the prevalence of the data of interest. For data of rare occurrence, very low values of kappa may not necessarily reflect low overall agreement. The kappa statistic for all other data elements reflected moderate to almost perfect agreement.

There were two types of data element for which we did not feel comfortable providing kappa statistics. Three data elements (“Discharge Status”, “E/M Code”, and “ICD-9-CM Principal Diagnosis Code”) were categorical variables with more than two response categories. For these multi-category response variables, we did not calculate the kappa statistic for three reasons. First, kappa was originally designed for a dichotomous response variable. Although it has since been extended to handle multi-category response variables (from 2 x 2 to 2 x n), one data element for this measure, “Discharge Status”, had 26 unique responses. Calculation of a kappa statistic for this data element is possible, but is likely to have less meaning than the raw agreement rate. Second, these multi-category data elements all had perfect agreement rates (100%). Any kappa statistic that was a significant departure from this raw agreement rate would suggest unreliability of the kappa statistic, not unreliability of the raw agreement rate.

The second date element type for which we did not calculate kappa statistics was date/time variables that are commonly treated as continuous variables. Initially, it was thought that we could calculate the Intraclass Correlation Coefficient (ICC) for these measures. However, upon investigation it was determined not to be appropriate to calculate the ICC for the date and time variables. Because of the way the statistical software that we use (SAS) handles these date and time values, the ICC is inflated to 1, which does not provide useful information. As an alternative to the ICC for the time data elements, we calculated the median time difference between the abstractors when the time elements did not have matching values. We acknowledge this does not adjust the agreement rate for chance, but it does provide a sense of the magnitude of disagreement between abstractors when it existed. This analysis was not reported for the date data elements due to the very high agreement in these data elements.

It should be noted that the variations in the denominators of individual data elements were not due to excessive missing data. The data were collected and analyzed according to the design of measure algorithm, which may have a parent-child relationship between data elements. For example, the “Reason for No Aspirin on Arrival” data element had a parent-child relationship with other data elements. If the case received aspirin on arrival, the data would not be collected for this data element.

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

**2b3. EXCLUSIONS ANALYSIS**

**NA** ☐ **no exclusions — *skip to section*** [***2b4***](#section2b4)There is one exclusionary data element for OP-4, 0286. It is entitled *Reason for No Aspirin on Arrival* and is a yes/no data element. It does not drill down to the reason.

**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*) We conducted descriptive analysis to examine the excluded cases and the impact on OP-4 measure. We listed the total number of cases reported for this measure in the QIO clinical data warehouse for Q1, 2013. We examined the number of hospitals that submitted the cases. The percentage of excluded cases was calculated by the list of exclusion criteria. For those hospital with any excluded cases, we reported the median number of excluded cases and the inter quartile range.

**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*)   
There were 33,227 cases for OP4 measure in the QIO clinical data warehouse for Q1, 2013 from 2,967 hospitals. Out of all submitted cases, 12.63% (n=4,198) of them were excluded from the measure for one of the two reasons: (1) the cases did not have Probable Cardiac Chest Pain (n=2,861, 8.61%); (2) the cases did not have a documented Reason for No Aspirin on Arrival (n=1,337, 4.02%).

There were 1,301 (43.85%) hospitals without any excluded cases. For those hospitals with any excluded cases, the median number of excluded cases was 2 with inter quartile range of 1 - 3.

**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*)  
As indicated above, a large proportion of hospitals (43.85%) had no excluded cases. For the rest of the hospitals, the excluded cases were evenly distributed with a median of 2. There was no evidence to suggest the excluded cases unfairly distorted the performance results.

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

**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** Click here to enter number of factors **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**.   
**2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient 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 and not related to disparities*)  
**2b4.4. What were the statistical results of the analyses used to select risk factors?  
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*)

*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*)**:   
2b4.7. Statistical Risk Model Calibration Statistics** (*e.g., Hosmer-Lemeshow statistic*):   
**2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves**:  
**2b4.9. Results of Risk Stratification Analysis**:

**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*)

**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)*

*Per NQF, we reported validity test for data elements. We did not conduct the validity test for performance score.*

**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*)

*Per NQF, we reported validity test for data elements. We did not conduct the validity test for performance score.*

**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?*)

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

**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 criterion is directed 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).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

**2b6.1. Describe the method of testing conducted to demonstrate comparability of 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*)  
 **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*)  
**2b6.3. What is your interpretation of the results in terms of demonstrating comparability of 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*)  
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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*)

It should be noted that the variations in the denominators of individual data elements were not due to excessive missing data. The data were collected and analyzed according to the design of measure algorithm, which may have a parent-child relationship between data elements. For example, the “Reason for No Aspirin on Arrival” data element had a parent-child relationship with other data elements. If the case received aspirin on arrival, the data would not be collected for this data element.

**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*)  
**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*)

Note: Aspirin at Arrival (0286) and Median Time to ECG (0289) can use the same reliability and validity data. These two measures have the same population, which includes Emergency Department acute myocardial infarction (AMI) patients or chest pain patients (with Probable Cardiac Chest Pain). The following critical data elements were reported:

|  |  |
| --- | --- |
| **Data Element Name** | **Reported for** |
| Arrival Time | OP-4, OP-5 |
| Birthdate | OP-4, OP-5 |
| Aspirin Received | OP-4 |
| Discharge Status | OP-4, OP-5 |
| E/M Code | OP-4, OP-5 |
| ECG | OP-5 |
| ECG Date | OP-5 |
| ECG Time | OP-5 |
| ICD-9-CM Principal Diagnosis Code | OP-4, OP-5 |
| Probable Cardiac Chest Pain | OP-4, OP-5 |
| Reason for No Aspirin on Arrival | OP-4 |

**Validity Testing**

*Results.* Table 1 below shows that the distributions of the patient/hospital characteristics in the sampled dataset and the original population. Patient’s characteristics in the table included age, gender, and race/ethnicity. Hospital characteristics included bed size, teaching status, and urban vs. rural location. The sampled dataset had a similar distribution of patient characteristics as the population. Compared to population data, hospitals in the sampled validation dataset tend to be smaller, non-teaching and rural.

Table 1. The distribution of patient and hospital characteristics between Sample and Population

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sample | | Population | |
|  | Frequency | Percent | Frequency | Percent |
| Patient Characteristics |  |  |  |  |
| **Gender** |  |  |  |  |
| Male | 447 | 58.58 | 68,899 | 60.31 |
| Female | 316 | 41.42 | 45,344 | 39.69 |
| Undetermined | 0 | 0.00 | 3 | 0.00 |
| **Race** |  |  |  |  |
| Caucasian | 616 | 80.73 | 92,809 | 81.24 |
| African American | 65 | 8.52 | 10,907 | 9.55 |
| Hispanic | 52 | 6.82 | 5,384 | 4.71 |
| Native American | 4 | 0.52 | 641 | 0.56 |
| Asian | 8 | 1.05 | 1,272 | 1.11 |
| Other/UTD | 18 | 2.36 | 3,233 | 2.83 |
| **Age Category** |  |  |  |  |
| Under 65 | 434 | 56.88 | 64,737 | 56.66 |
| Age 65\_74 | 159 | 20.84 | 25,607 | 22.41 |
| Age 75\_84 | 121 | 15.86 | 17,393 | 15.22 |
| Age 85 plus | 49 | 6.42 | 6,509 | 5.70 |
| **Hospital Characteristics** |  |  |  |  |
| **Bed Size** |  |  |  |  |
| 1 - 100 | 150 | 46.44 | 857 | 31.59 |
| 101 - 200 | 108 | 33.44 | 735 | 27.09 |
| 201 - 300 | 35 | 10.84 | 442 | 16.29 |
| 301 - 400 | 13 | 4.02 | 278 | 10.25 |
| 401 plus | 17 | 5.26 | 401 | 14.78 |
| **Teaching Status** |  |  |  |  |
| Yes | 48 | 14.86 | 752 | 27.72 |
| No | 275 | 85.14 | 1,961 | 72.28 |
| **Location** |  |  |  |  |
| Rural | 160 | 49.54 | 918 | 33.84 |
| Urban | 163 | 50.46 | 1,795 | 66.16 |

Table 2 summarizes the results of the validity test of 11 data elements. Overall, the agreement rates were high. The agreement rates for all data elements except two were higher than 90%.

**Table 2: Validity Test Summary Measure 0286 and 0289**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Element Name** | **Number of Eligible Cases (Denominator)** | **Number of cases in agreement** | **Agreement Rate (%)** | **Kappa Statistica** | **When times disagree, median difference** |
| Arrival Time | 656 | 621 | 94.67 | N/A | 9 minutes |
| Birthdate | 763 | 763 | 100.00 | N/A |  |
| Aspirin Received | 649 | 636 | 98.00 | 0.86a |  |
| Discharge Status | 763 | 763 | 100.00 | N/A |  |
| E/M Code | 763 | 763 | 100.00 | N/A |  |
| ECG | 649 | 643 | 99.08 | 0.25a |  |
| ECG Date | 642 | 635 | 98.91 | N/A |  |
| ECG Time | 642 | 588 | 91.59 | N/A | 11 minutes |
| ICD-9-CM Principal Diagnosis Code | 763 | 763 | 100.00 | N/A |  |
| Probable Cardiac Chest Pain | 449 | 383 | 85.30 | 0.53a |  |
| Reason for No Aspirin on Arrival | 44 | 34 | 77.27 | N/A |  |

a - Kappa Statistics

Information below uses 1Q2013 data





