**National Quality Forum—Measure Testing (subcriteria 2a2, 2b1-2b6)**

**Measure Number** (*if previously endorsed*)**:** 0541

**Measure Title**: Proportion of Days Covered: 3 Rates by Therapeutic Category

**Date of Submission**: 4/9/2019

**Type of Measure:**

|  |  |
| --- | --- |
| Outcome (*including PRO-PM*) | Composite – ***STOP – use composite testing form*** |
| Intermediate Clinical Outcome | Cost/resource |
| Process *(including Appropriate Use)* | 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, 2b1, 2b2, and 2b4 must be completed.** * **For outcome and resource use measures**, section **2b3** also must be completed. * If specified for **multiple data sources/sets of specifications** (e.g., claims and EHRs), section **2b5** 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 (2b1-2b6) 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 25 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 social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment. |

|  |
| --- |
| **Note:** The information provided in this form is intended to aid the Standing 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 **instrument-based measures** (including PRO-PMs) **and composite performance measures**, reliability should be demonstrated for the computed performance score.  **2b1.** **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 **instrument-based measures (including PRO-PMs) and composite performance measures**, validity should be demonstrated for the computed performance score.    **2b2.** **Exclusions** are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [**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)  **2b3.** **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 social risk 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.   **2b4.** 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.  **2b5.** **If multiple data sources/methods are specified, there is demonstration they produce comparable results**.  **2b6.** 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. The degree of consensus and any areas of disagreement must be provided/discussed.  **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.17*)** | **Measure Tested with Data From:** |
| abstracted from paper record | abstracted from paper record |
| claims | claims |
| registry | 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 Pharmacy Quality Alliance (PQA) proportion of days covered (PDC) 3 rates by therapeutic category (Diabetes, renin-angiotensin system antagonist [RASA], and Statin Medications) measure was tested within two different health plan data sources – the Medicare and the Medicaid populations.

For the Medicare population, data used for testing came from the Medicare Research Identifiable Files (RIFs) 5% national sample data. The Medicare Part D Prescription Drug Event (PDE) claims were used for the identification of prescription drugs. The claims files and Medicare Provider Analysis and Review (MedPAR) files were used to identify end-stage renal disease (ESRD) diagnoses and hospice claims. To identify demographic and eligibility information, the Medicare Beneficiaries Summary Files (MBSF) were used.

For the Medicaid population, the data used for testing came from Medicaid administrative claims in the Medicaid Analytic eXtract (MAX) data. National Medicaid sample data covering 17 states and 291 health plans were included in the testing.

*Note: Testing was conducted separately for each therapeutic category (Diabetes, renin-angiotensin system antagonist [RASA], Statins) and each line of business (Medicare, Medicaid).*

**1.3. What are the dates of the data used in testing**? 2014 and 2016

The testing for Medicare included data from January 1, 2016 to December 31, 2016. The testing for Medicaid included data from January 1, 2014 to December 31, 2014. The data from these time periods were the most recent, complete, full year data available to testers at the time of testing.

**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.20*)** | **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*)

The Medicare testing was conducted using the Medicare RIF 5% sample data – a nationally representative sample, including data from all states. Of beneficiaries aged 18 years and older as of the first day of the measurement year, the data included 491 Medicare Advantage Prescription Drug (MAPD) contracts and 63 stand-alone Prescription Drug Plans (PDPs). Of the 554 plan-contracts, the mean plan-contract size was 3,978 beneficiaries with a median size of 562 beneficiaries. (see Table 1).

For the Medicaid testing, the analysis included 291 health plans covering 17 states with beneficiaries aged 18 years or older. Of the 291 plans, 17 plans were fee-for-service (FFS), and the remaining 274 plans were Medicaid Managed Care Organizations (MCOs). There was variation in plan size, with mean plan size of 18,415 beneficiaries, and a median plan size of 3,656 beneficiaries. (see Table 1).

**Table 1. Plan/Contract Size Distribution for Medicare and Medicaid Populations**

|  |  |  |
| --- | --- | --- |
| **Statistic** | **Medicare** | **Medicaid** |
| Mean | 3,978 | 18,415 |
| Standard Deviation | 19,648 | 43,222 |
| Minimum | 30 | 30 |
| 25th Percentile | 190 | 354 |
| 50th Percentile | 562 | 3,656 |
| 75th Percentile | 2,054 | 17,560 |
| Maximum | 291,999 | 450,884 |
| Interquartile Range | 1,864 | 17,206 |

*Note: When used in performance programs, plans/contracts with <30 individuals are excluded; as such, all analyses exclude plans/contracts with <30 individuals.*

**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*)

For the Medicare testing, the initial population ages 18 years and older included 2,203,754 individuals. After applying all inclusion/exclusion criteria, the Diabetes population included 268,737 individuals, the RASA population included 775,226 individuals, and the Statins population included 872,736 individuals. For all therapeutic classes, over 80% of beneficiaries were ages 65 years and older. For all therapeutic classes, over 50% of the population was female, and between 75% (Diabetes) and 81% (Statins) was white, with a little more than 3% of the population’s race classified as “other/unknown”. In addition, 25-30% of beneficiaries had low-income subsidy (LIS) and/or dual eligibility status, with the Diabetes population having the largest percentage (30.3%). More than 10% of the population were entitled to Medicare due to disability, with the highest percentage in the Diabetes population (13.3%). (see Table 2).

**Table 2. Population Characteristics for Medicare – By Therapeutic Category**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Diabetes**  **(n = 268,737)** | | **RASA**  **(n = 775,226)** | | **Statins**  **(n = 872,736)** | |
| **Characteristic** | **n** | **%** | **n** | **%** | **n** | **%** |
| **Age Group** |  |  |  |  |  |  |
| 18-54 | 14,131 | 5.3 | 33,571 | 4.3 | 34,523 | 4.0 |
| 55-64 | 29,722 | 11.1 | 77,487 | 10.0 | 83,354 | 9.6 |
| 65-69 | 72,778 | 27.1 | 201,213 | 26.0 | 226,016 | 25.9 |
| 70-74 | 62,332 | 23.2 | 174,978 | 22.6 | 201,962 | 23.1 |
| 75-79 | 44,363 | 16.5 | 128,778 | 16.6 | 149,544 | 17.1 |
| 80+ | 45,411 | 16.9 | 159,199 | 20.5 | 177,337 | 20.3 |
| **Gender** |  |  |  |  |  |  |
| Male | 125,356 | 46.7 | 344,118 | 44.4 | 394,867 | 45.2 |
| **Race** |  |  |  |  |  |  |
| White | 200,254 | 74.5 | 618,035 | 79.7 | 710,176 | 81.4 |
| Black | 35,529 | 13.2 | 87,017 | 11.2 | 83,130 | 9.5 |
| Asian | 9,885 | 3.7 | 18,425 | 2.4 | 23,072 | 2.6 |
| Hispanic | 11,157 | 4.2 | 23,601 | 3.0 | 23,930 | 2.7 |
| North American Native | 1,072 | 0.4 | 2,798 | 0.4 | 2,661 | 0.3 |
| Other | 7,203 | 2.7 | 15,145 | 2.0 | 17,565 | 2.0 |
| Unknown | 3,637 | 1.4 | 10,205 | 1.3 | 12,202 | 1.4 |
| **LIS and/or Dual** | 81,294 | 30.3 | 200,902 | 25.9 | 219,056 | 25.1 |
| **Disability as reason for Medicare entitlement** | 35,606 | 13.3 | 88,690 | 11.4 | 93,876 | 10.8 |
| *RASA: renin-angiotensin system antagonist; LIS: low-income subsidy* | | | | | | |

For the Medicaid testing, the initial population ages 18 years and older included 5,358,811 individuals. After applying all inclusion/exclusion criteria, the Diabetes population included 234,185 individuals, the RASA population included 572,736 individuals, and the Statins population included 478,586 individuals. For all therapeutic classes, over 90% of beneficiaries were between 18 and 64 years old. For all therapeutic classes, a majority of the population was female, ranging between 59% (RASA & Statins) and 64% (Diabetes), and the proportion of the population that identified as white race was between 35% (Diabetes) and 42% (Statins). For the Medicaid population, about 14% of the population was classified as having an “other/unknown” race. (see Table 3).

**Table 3. Population Characteristics for Medicaid – By Therapeutic Category**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Diabetes**  **(n = 234,185)** | | **RASA**  **(n = 572,736)** | | **Statins**  **(n = 478,586)** | |
| **Characteristic** | **n** | **%** | **n** | **%** | **n** | **%** |
| **Age Group** |  |  |  |  |  |  |
| 18-54 | 140,882 | 60.2 | 339,519 | 59.3 | 259,195 | 54.2 |
| 55-64 | 79,185 | 33.8 | 201,415 | 35.2 | 189,561 | 39.6 |
| 65-69 | 4,820 | 2.1 | 10,081 | 1.8 | 9,953 | 2.1 |
| 70-74 | 4,156 | 1.8 | 9,007 | 1.6 | 8,688 | 1.8 |
| 75-79 | 2,935 | 1.3 | 6,790 | 1.2 | 6,298 | 1.3 |
| 80+ | 2,207 | 0.9 | 5,924 | 1.0 | 4,891 | 1.0 |
| **Gender** |  |  |  |  |  |  |
| Male | 85,145 | 36.4 | 237,688 | 41.5 | 195,843 | 40.9 |
| **Race** |  |  |  |  |  |  |
| White | 82,416 | 35.2 | 222,499 | 38.9 | 200,878 | 42.0 |
| Black | 44,398 | 19.0 | 138,581 | 24.2 | 86,970 | 18.2 |
| Asian | 14,347 | 6.1 | 24,851 | 4.3 | 30,052 | 6.3 |
| Hispanic | 56,617 | 24.2 | 103,417 | 18.1 | 84,094 | 17.6 |
| North American Native | 1,560 | 0.7 | 3,583 | 0.6 | 2,574 | 0.5 |
| Other | 11,137 | 4.8 | 22,515 | 3.9 | 22,155 | 4.6 |
| Unknown | 23,710 | 10.1 | 57,290 | 10.0 | 51,863 | 10.8 |
| *RASA: renin-angiotensin system antagonist* | | | | | | |

**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**.

All testing was conducted using the same datasets for each line of business except the following:

* empirical validity testing: conducted using publicly available national-level Centers for Medicare and Medicaid Services (CMS) Medicare data from 2016 as these measures are used in the Medicare Part D Star Ratings program,
* missing data and unadjusted reliability testing for Medicare: used the 100% CMS Medicare PDE data from 2014, and
* **additional testing:** additional unadjusted and risk-adjusted reliability testing conducted using 100% CMS Medicare PDE data from 2017.

**1.8** **What were the social risk factors that were available and analyzed**? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

Variables selected for analyses for the Medicare population included beneficiary-, community- (9-digit zip code) and county-level variables. (see Table 4).

**Table 4. SDS Variables Considered for Risk Adjustment for the Medicare Population**

|  |  |
| --- | --- |
| **Variable Level** | **Variable** |
| Beneficiary-level | Age |
| Gender |
| LIS status or Dual eligibility status |
| Disability as original reason for Medicare entitlement |
| Race |
| Community-level (9-digit zip code) | Median income |
| Percent of households where residents are married |
| Percent of households where residents completed college |
| Percent of households where residents own their home |
| County-level | Federally designated primary care professional shortage area |
| Federally designated mental healthcare professional shortage area |
| *LIS: low-income subsidy; SDS: sociodemographic status* | |

No social risk factors were analyzed for the Medicaid population.

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

**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*)

Reliability testing was conducted for the unadjusted measure scores for Medicare and Medicaid, and the risk-adjusted measure scores for Medicare.

The reliability of a measurement refers to the degree to which repeated measurements of the same entity agree with each other. For the health plan/contract-level performance measures, reliability characterizes the extent to which repeated measurements of the same plans/contracts produce similar results.

For the Medicare and Medicaid unadjusted measure rates, the reliability of the computed measure scores was measured as the ratio of signal-to-noise. The signal is the proportion of the variability in measured performance that can be explained by true differences in plan (or contract) performance. Reliability scores range from 0 to 1, with a score of 0 signifying that all variation is due to measurement error. A value of 1 signifies that the variation represents true differences in performance scores between plans. A reliability score of 0.7 is the minimum threshold for reliability.

A beta-binomial model was used to calculate plan-specific reliability scores based on the method outlined by Adams.1 The reliability score is defined as the ratio of the plan-to-plan variance to the sum of the plan-to-plan variance and the plan-specific error. The plan-to-plan variance is an estimate of the variance of the true rates. The plan-specific error variance is the sampling or measurement error.

Risk adjustment was applied at the measure score level. As such, the Adams beta binomial methodology described above could not be used to assess reliability of the risk-adjusted measure scores. To assess reliability of the risk-adjusted measure scores for Medicare, we employed a split-half approach where plan-contract performance was measured using a random sample of beneficiaries, and then measured again using a second random sample. The two groups are independent samples, with each sample including half of the population within each plan-contract. This means that each plan-contract is measured twice, but each measurement is made using an entirely distinct set of beneficiaries. As a metric of agreement, we calculated the intra-class correlation coefficient (ICC)2 and assessed the values according to conventional standards.3

**References:**

1. Adams JL. The reliability of provider profiling: a tutorial. Santa Monica, CA: RAND Corporation. 2009. Retrieved from <http://www.rand.org/pubs/technical_reports/TR653>.
2. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979; 86:420-428. PMID: [27330520](https://www.ncbi.nlm.nih.gov/pubmed/27330520).
3. Koo TK, Li MY. A guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine. 2016; 15(2):155-63. PMID: [27330520](https://www.ncbi.nlm.nih.gov/pubmed/27330520).

**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*)

Unadjusted Measure Scores

Using the parameter estimates from the beta-binomial model, we computed individual plan/contract reliability scores. Table 5a shows the distribution of the plan/contract level scores for Medicare and Medicaid.

**Table 5a. Unadjusted Plan/Contract Reliability Scores – By Therapeutic Category**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Statistic** | **Medicare** | | | **Medicaid** | | |
| **Diabetes** | **RASA** | **Statins** | **Diabetes** | **RASA** | **Statins** |
| Mean | 0.8492 | 0.8953 | 0.9171 | 0.9174 | 0.9340 | 0.9305 |
| Standard Deviation | 0.1785 | 0.1579 | 0.1319 | 0.1008 | 0.0990 | 0.1005 |
| Minimum | 0.2161 | 0.2664 | 0.3321 | 0.6207 | 0.5587 | 0.5763 |
| 25th Percentile | 0.7747 | 0.8768 | 0.9025 | 0.8903 | 0.9262 | 0.9260 |
| 50th Percentile | 0.9316 | 0.9724 | 0.9793 | 0.9655 | 0.9798 | 0.9781 |
| 75th Percentile | 0.9792 | 0.9941 | 0.9951 | 0.9879 | 0.9938 | 0.9933 |
| Maximum | 0.9998 | 1.0000 | 1.0000 | 0.9992 | 0.9996 | 0.9995 |
| Interquartile Range | 0.2045 | 0.1172 | 0.0926 | 0.0976 | 0.0676 | 0.0674 |
| *RASA: renin-angiotensin system antagonist*  *Note: Reliability testing for Medicare conducted using the 2014 100% CMS PDE data.* | | | | | | |

The mean reliability score for the Medicare plan-contract (with median) was 0.8492 (0.9316) for Diabetes, 0.8953 (0.9724) for RASA, and 0.9171 (0.9793) for Statins.

The mean reliability score for the Medicaid plans (with median) was 0.9174 (0.9655) for Diabetes, 0.9340 (0.9798) for RASA, and 0.9305 (0.9781) for Statins.

**Additional Testing: Using 2017 100% CMS PDE Data for Medicare**

**Table 5b. Unadjusted Plan-Contract Reliability Scores for Medicare – By Therapeutic Category**

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistic** | **Medicare** | | |
| **Diabetes** | **RASA** | **Statins** |
| Mean | 0.8553 | 0.8774 | 0.9211 |
| Standard Deviation | 0.1811 | 0.1815 | 0.1284 |
| Minimum | 0.2346 | 0.2257 | 0.3710 |
| 25th Percentile | 0.7912 | 0.8393 | 0.9109 |
| 50th Percentile | 0.9377 | 0.9706 | 0.9834 |
| 75th Percentile | 0.9853 | 0.9945 | 0.9967 |
| Maximum | 0.9999 | 1.0000 | 1.0000 |
| Interquartile Range | 0.1941 | 0.1552 | 0.0858 |
| *RASA: renin-angiotensin system antagonist*  *Note: Additional reliability testing for Medicare conducted using the 2017 100% CMS PDE data.* | | | |

The mean reliability score for the Medicare plan-contract (with median) was 0.8553 (0.9377) for Diabetes, 0.8774 (0.9706) for RASA, and 0.9211 (0.9834) for Statins. (see Table 5b). These results were similar to the 2014 100% CMS PDE analysis in Table 5a.

Risk-Adjusted Measure Scores for Medicare

Using the 2016 5% Medicare sample

The intra-class correlation between the two risk standardized scores among the two samples was 0.3548 for Diabetes, 0.3513 for RASA and 0.5022 for Statins.

**Additional Testing: Using the 2017 100% CMS PDE data**

The intra-class correlation between the two risk standardized scores among the two samples was 0.7756 for Diabetes, 0.7260 for RASA and 0.8806 for Statins.

**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?*)

Unadjusted Measure Scores

A reliability score of 0.7 is considered the minimum threshold for reliability. Based on the mean reliability scores between 0.85 (Diabetes) and 0.92 (Statins) for Medicare and 0.92 (Diabetes) and 0.93 (RASA and Statins) for Medicaid, the measure scores for the three therapeutic categories are considered reliable.

Risk-Adjusted Measure Scores for Medicare (2016 5% sample)

The ICC scores for Diabetes and RASA are considered low, while the ICC score for Statins is considered moderate according to conventional interpretation.1 However, it is important to note that the analysis was conducted using the 5% sample of the Medicare population, and sample size is one of the drivers of reliability.

Smaller plan-contracts may negatively impact the ICC as the random split is more likely to introduce noise, since the two halves may not be equally balanced, unlike large plan-contracts where we expect both samples to be normally distributed due to the law of large numbers. Thus, we anticipate the reliability would improve when applied to the total Medicare population. This is evident in the high reliability for the unadjusted measure scores for Medicare, which was conducted using the 100% CMS PDE data.

**Additional Testing (2017 100% CMS PDE Data):**

As expected, reliability of the risk-adjusted measure scores improved significantly when applied to the 100% CMS PDE Medicare dataset, ranging from 0.73 (RASA) to 0.88 (Statins). Based on conventional interpretation,1 reliability for all three measures were good.

**References:**

1. Koo TK, Li MY. A guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine. 2016; 15(2):155-63. PMID: [27330520](https://www.ncbi.nlm.nih.gov/pubmed/27330520).

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

**2b1. VALIDITY TESTING**

**2b1.1. 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*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

**2b1.2. 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)*

We assessed convergent validity of the measure score at the plan-contract level (correlation with related measure in the same performance year) for Medicare and face validity for all lines of business.

Convergent Validity

Convergent validity was tested for each therapeutic category for the Medicare population using the publicly available CMS Medicare Part C & D Star Ratings data.1 Correlation was examined using the Pearson correlation statistics between each of the therapeutic categories and other performance measures used in the Star Ratings program.

For Diabetes, we examined whether the measure rate is correlated with the *Diabetes Care – Blood Sugar Controlled* measure used in the CMS Part C Star Ratings program. We hypothesized that organizations that perform well on the PDC Diabetes measure should perform well on the *Diabetes Care – Blood Sugar Controlled* measure as both focus on diabetes care, and adherence to anti-hyperglycemic agents can lower blood sugar and decrease complications such as visual loss and renal failure.2,3

For RASA, convergent validity was tested by exploring whether the measure rate is correlated with the *Controlling Blood Pressure* measure used in the CMS Part C Star Ratings program. We hypothesized that organizations that perform well on the PDC RASA measure should perform well on the *Controlling Blood Pressure measure*. According to the 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA hypertension guidelines and the 2018 American Diabetes Association guidelines, medication nonadherence is a major contributor to poor control of hypertension and a key barrier to reducing mortality.4,5 Moreover, studies have shown improved clinical outcomes for individuals who are adherent to their medications.6

For Statins, convergent validity was tested by exploring whether the measure rate is correlated with the *Statin Use in Persons with Diabetes* measure used in the CMS Part D Display measures. We hypothesized that organizations that perform well on the PDC Statins measure should perform well on the *Statin Use in Persons with Diabetes* measure. Atherosclerotic cardiovascular disease (ASCVD) is the leading cause of morbidity and mortality for individuals with diabetes in addition to being the largest contributor to costs of diabetes care.4 Individuals with diabetes who are 40-75 years old are at a markedly increased lifetime risk for the development of ASCVD, experience greater morbidity, and are at a decreased likelihood of survival following the onset of ASCVD.HMG-CoA reductase inhibitors, also known as statins, are recommended for management of dyslipidemia and/or primary prevention of cardiovascular disease (CVD) in several treatment guidelines.7-11 By lowering LDL cholesterol, statins decrease the risk of CVD morbidity and mortality.12

**References:**

Centers for Medicare and Medicaid Services. Part C and D Performance Data. Available at: <https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html>.

Lau DT, Nau DP. Oral antihyperglycemic medication nonadherence and subsequent hospitalization among individuals with type 2 diabetes. Diabetes Care. 2004; 27(9):2149-53. PMID: [15333476](https://www.ncbi.nlm.nih.gov/pubmed/15333476).

1. Sokol MC, McGuigan KA, Verbrugge RR, Epstein RS. Impact of medication adherence on hospitalization risk and healthcare cost. Med Care. 2005; 43(6):521-30. PMID: [15908846](https://www.ncbi.nlm.nih.gov/pubmed/15908846).
2. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Soc Hypertens. 2018; 12(8):579.e1-579.e73. PMID: [30219548](https://www.ncbi.nlm.nih.gov/pubmed/30219548).
3. American Diabetes Association. Standards of Medical Care in Diabetes – 2018. Diabetes Care. 2018; 41(Supplement 1): S1-S159. https://doi.org/10.2337/dc18-Srev01.

Sokol MC, McGuigan KA, Verbrugge RR, Epstein RS. Impact of medication adherence on hospitalization risk and healthcare cost. Med Care. 2005; 43(6):521-30. PMID: [15908846](https://www.ncbi.nlm.nih.gov/pubmed/15908846).

1. Jellinger PS, Handelsman Y, Rosenblit PD, et al. American Association of Clinical Endocrinologists and American College of Endocrinology guidelines for management of dyslipidemia and prevention of cardiovascular disease. Endocr Pract. 2017; 23(Suppl 2):1-87. PMID: [28437620](https://www.ncbi.nlm.nih.gov/pubmed/28437620).
2. Taylor F, Huffman MD, Macedo AF, et al. Statins for the primary prevention of cardiovascular disease. Cochrane Database Syst Rev. 2013; (1):CD004816. PMID: [23440795](https://www.ncbi.nlm.nih.gov/pubmed/23440795).
3. Stone NJ, Robinson JG, Lichtenstein AH, et al. American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014; 129(25 Suppl 2):S1-45. PMID: [24222016](https://www.ncbi.nlm.nih.gov/pubmed/24222016).
4. Jacobson TA, Ito MK, Maki KC, et al. National lipid association recommendations for patient-centered management of dyslipidemia: part 1—full report. J Clin Lipidol. 2015 ;9(2):129-69. PMID: [25911072](https://www.ncbi.nlm.nih.gov/pubmed/25911072).
5. Bibbins-Domingo K, Grossman DC, Curry SJ, et al. Statin Use for the Primary Prevention of Cardiovascular Disease in Adults: US Preventive Services Task Force Recommendation Statement. JAMA. 2016; 316(19):1997-2007. PMID: [27838723](https://www.ncbi.nlm.nih.gov/pubmed/27838723).
6. Cholesterol Treatment Trialists' (CTT) Collaborators, Mihaylova B, Emberson J, Blackwell L, et al. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. Lancet. 2012; 380(9841):581-90. PMID: [22607822](https://www.ncbi.nlm.nih.gov/pubmed/22607822).

Face Validity

PQA uses a systematic, transparent, consensus-based measure development, testing, and endorsement process. That process was used in 2008 to develop this measure. The measure was assessed for face validity (i.e., whether it appears to measure what it intends to measure) through review by workgroup participants that developed the measure (PQA Adherence Workgroup), the PQA Quality Metrics Expert Panel (QMEP), and PQA’s full membership.

The 2018 PQA Measure Update Panel and QMEP most recently reviewed this measure. These panels include individuals with expertise and experience in pharmacy, medicine, research, and clinical or other technical expertise related to quality improvement and measure development.

**2b1.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test*)

Results for convergent validity testing for the Medicare population is shown in Table 6. For all three therapeutic categories, there was a positive relationship with the comparator, and all three were statistically significant at alpha <0.0001.

**Table 6. Convergent Validity Testing for Medicare – By Therapeutic Category**

|  |  |  |  |
| --- | --- | --- | --- |
| **Therapeutic Category** | **Comparison** | **Correlation Coefficient** | ***p*-Value** |
| Diabetes | C15: Diabetes Care - Blood Sugar Controlled | 0.465 | <0.0001 |
| RASA | C16: Controlling Blood Pressure | 0.517 | <0.0001 |
| Statins | DMD15: Statin Use in Persons with Diabetes | 0.346 | <0.0001 |
| *RASA: renin-angiotensin system antagonists*  *Note: Correlation analyses conducted using the CMS Medicare Part C & D Star Ratings Data from January – December 2016.* | | | |

**2b1.4. 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?*)

All three therapeutic categories showed statistically significant positive correlation, which indicates that the measure demonstrates convergent validity.

Based upon the systematic, consensus-based PQA measure development process designed to assure face validity, the measure has been determined to have face validity.

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

**2b2. EXCLUSIONS ANALYSIS**

**NA**  **no exclusions — *skip to section*** [***2b3***](#section2b4)

**2b2.1. 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 examined the overall proportion of Medicare and Medicaid beneficiaries impacted as well as the proportion of individuals who would be impacted by the exclusions at the plan/contract-level.

Hospice

Individuals in hospice care are excluded from the measure because adherence to most chronic therapies does not necessarily align with the therapeutic goals and balance of risk and benefits for individuals in hospice care. In 2015, the PQA Patient & Caregiver Advisory Panel recommended that hospice should generally be an exclusion for PQA measures unless the measures are directly relevant to, and align with, the therapeutic goals for individuals in hospice care.

End-Stage Renal Disease

Individuals with end-stage renal disease (ESRD) are excluded from the measure because adherence to diabetes, hypertension, and statin medications may not be accurately reflected in pharmacy claims data due to frequent dosage and medication adjustments.

Individuals with diabetes and ESRD are at higher risk for hypoglycemia than the general population. Reasons for fluctuation in blood glucose can range from drug accumulation to more complex reasons like increased glucose utilization following the correction of anemia by erythropoietin.1 Peritoneal dialysis patients may also have glucose-containing dialysate that influences glycemic control, with alternating hyperglycemia and hypoglycemia and resultant adjustments to diabetes medications.2 Additionally, individuals with ESRD often have spontaneous hypoglycemia due to reduced renal gluconeogenesis or concurrent hepatic disease, but also experience fluctuations in insulin resistance due to the process of dialysis.1

Individuals with ESRD may require RASA medication dosage adjustments due to severe fluid imbalances resulting in high blood pressure, followed by sudden hypotension when fluids are removed with dialysis.3 A confounding factor is variable adherence to dietary regimens (e.g., fluid restrictions). Therefore, assessing adherence to oral antihypertensive medications using pharmacy claims data in individuals with ESRD is imprecise due to frequent fluctuations in blood pressure and resulting medication adjustments (e.g., they are often held, changed, discontinued, or restarted).

Individuals with ESRD may also feel generalized weakness and adherence to statins may be a lower clinical priority than interventions to manage mineral bone disease and fluid management (e.g., medication and dietary modifications).4 Although individuals with ESRD are at increased risk for cardiovascular events related to sudden cardiac arrest, left ventricular hypertrophy, vascular calcifications from hyperphosphatemia and hyperparathyroidism, and large fluid shifts5-7 there is a lack of direct evidence that statin treatment is beneficial in dialysis patients in the Kidney Disease: Improving Global Outcomes (KDIGO) guidelines.8

Insulin for Diabetes

Currently, there is not a standardized method to assess adherence to insulin using prescription claims data.9 Individuals on insulin are excluded from the Diabetes rate because insulin requires titration and frequent dosage adjustments, which in turn can result in frequent dosage adjustments of other diabetes medications.

Sacubitril/Valsartan for RASA

The RASA rate is intended to evaluate adherence to medications used for treating hypertension. Individuals receiving sacubitril/valsartan are excluded from the measure because this product, although it includes a RASA (i.e., valsartan), is only indicated for treating heart failure.

**References:**

1. Sampanis Ch. Management of hyperglycemia in patients with diabetes mellitus and chronic renal failure. Hippokratia. 2008; 12(1):22-7. PMID: [18923754](https://www.ncbi.nlm.nih.gov/pubmed/18923754).
2. Alsahli M, Gerich JE. Hypoglycemia in Patients with Diabetes and Renal Disease. J Clin Med;2015; 13;4(5):948-64. PMID: [26239457](https://www.ncbi.nlm.nih.gov/pubmed/26239457).
3. K/DOQI Workgroup. K/DOQI clinical practice guidelines for cardiovascular disease in dialysis patients. Am J Kidney Dis. 2005; 45(4 Suppl 3):S1-153. PMID: [15806502](https://www.ncbi.nlm.nih.gov/pubmed/15806502).
4. Burnier M, Pruijm M, Wuerzner G, Santschi V. Drug adherence in chronic kidney diseases and dialysis. Nephrol Dial Transplant. 2015 ;30(1):39-44. PMID: [24516224](https://www.ncbi.nlm.nih.gov/pubmed/24516224).
5. Foley RN, Parfrey PS. Cardiovascular disease and mortality in ESRD. J Nephrol.1998 ;11(5):239-45. Review. PubMed PMID: [9831236](https://www.ncbi.nlm.nih.gov/pubmed/9831236).
6. Zoccali C, Benedetto FA, Tripepi G, Mallamaci F. Cardiac consequences of hypertension in hemodialysis patients. Semin Dial. 2004 ;17(4):299-303. PMID: [15250922](https://www.ncbi.nlm.nih.gov/pubmed/15250922).
7. United States Renal Data System. 2016 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2016. Available at: https://www.usrds.org/2016/view/default.aspx
8. Wanner C, Tonelli M. KDIGO Clinical Practice Guideline for Lipid Management in CKD: summary of recommendation statements and clinical approach to the patient. Kidney Int. 2014; 85(6):1303-9. PMID: [24552851](https://www.ncbi.nlm.nih.gov/pubmed/24552851).
9. Stolpe S, Kroes MA, Webb N, Wisniewski T. A Systematic Review of Insulin Adherence Measures in Patients with Diabetes. J Manag Care Spec Pharm. 2016; 22(11):1224-1246. PMID: [27783551](https://www.ncbi.nlm.nih.gov/pubmed/27783551).

**2b2.2. 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*)

For the Medicare population, overall, the percentage of individuals excluded were <1% (except for insulin for Diabetes). However, at the plan-contract level, some plan-contracts were disproportionately impacted, with the percentage of individuals in hospice who were excluded from the measure ranging from 0.0% to 3.8% for Diabetes, 0.0% - 4.8% for RASA and 0.0% - 5.6% for Statins. The ESRD exclusions among plan-contracts ranged from 0.0% to 9.6% for Diabetes, 0.0% to 16.4% for RASA and 0.0% - 21.9% for Statins.

In addition, for Diabetes, individuals on insulin are excluded from the measure. This exclusion ranged from 0.0% - 26.2% of individuals impacted within plan-contracts. For RASA, individuals are excluded if they were on sacubitril/valsartan at any time during the measurement year. This exclusion ranged from 0.0% - 3.3% of individuals impacted within plan-contracts. (see Table 7).

**Table 7. Means Distribution of Proportion of Individuals Impacted Across Plan-Contracts for Medicare**

|  |  |  |  |
| --- | --- | --- | --- |
| **Exclusion by therapeutic class**  **(n = 2,203,754)** | **N** | **%** | **Distribution across plan-contracts; min, 25th, 50th, 75th, max** |
| **Diabetes** |  |  |  |
| ESRD | 5,353 | 0.24 | (0.00, 0.00, 0.00, 0.12, 9.59) |
| Hospice | 5,013 | 0.23 | (0.00, 0.00, 0.15, 0.28, 3.77) |
| Insulin | 80,274 | 3.64 | (0.00, 2.71, 3.64, 4.74, 26.19) |
| **RASA** |  |  |  |
| ESRD | 11,787 | 0.53 | (0.00, 0.00, 0.12, 0.33, 16.44) |
| Hospice | 11,253 | 0.51 | (0.00, 0.00, 0.38, 0.60, 4.82) |
| Sacubitril/Valsartan | 9,013 | 0.41 | (0.00, 0.00, 0.30, 0.52, 3.33) |
| **Statins** |  |  |  |
| ESRD | 16,333 | 0.74 | (0.00, 0.00, 0.21, 0.45, 21.92) |
| Hospice | 12,115 | 0.55 | (0.00, 0.00, 0.42, 0.67, 5.63) |
| *ESRD: end-stage renal disease; RASA: renin-angiotensin system antagonist* | | | |

As with Medicare, the overall percentage of individuals impacted by the exclusions in the Medicaid population was <1% (except insulin for Diabetes). At the plan-level however, this impact was disproportionate, with the percentage of individuals in hospice who were excluded from the measure ranged from 0.0% to 8.0% for all three therapeutic areas. The ESRD exclusions among plans ranged from 0.0% to 1.3% for Diabetes, 0.0% to 2.3% for RASA and 0.0% - 0.6% for Statins. (see Table 8).

In addition, for Diabetes, individuals on insulin are excluded from the measure. This exclusion ranged from 0.0% - 7.1% of individuals impacted within plans. For RASA, individuals are excluded if they were on sacubitril/valsartan at any time during the measurement year. This exclusion ranged from 0.0% - 2.2% of individuals impacted within plans. (see Table 8).

**Table 8. Means Distribution of Proportion of Individuals Impacted Across Plans for Medicaid**

| **Exclusion by therapeutic class**  **(n = 5,358,811)** | **N** | **%** | **Distribution across plan-contracts; min, 25th, 50th, 75th, max** |
| --- | --- | --- | --- |
| **Diabetes** |  |  |  |
| ESRD | 1,655 | 0.03 | (0.00, 0.00, 0.01, 0.03, 1.28) |
| Hospice | 6,874 | 0.13 | (0.00, 0.00, 0.05, 0.12, 8.00) |
| Insulin | 93,797 | 1.75 | (0.00, 1.00, 1.46, 1.99, 7.14) |
| **RASA** |  |  |  |
| ESRD | 5,500 | 0.10 | (0.00, 0.00, 0.05, 0.11, 2.27) |
| Hospice | 6,874 | 0.13 | (0.00, 0.00, 0.05, 0.12, 8.00) |
| Sacubitril/Valsartan | 4,485 | 0.08 | (0.00, 0.00, 0.02, 0.06, 2.17) |
| **Statins** |  |  |  |
| ESRD | 5,052 | 0.09 | (0.00, 0.00, 0.05, 0.10, 0.58) |
| Hospice | 6,874 | 0.13 | (0.00, 0.00, 0.05, 0.12, 8.00) |
| *ESRD: end-stage renal disease; RASA: renin-angiotensin system antagonist* | | | |

**2b2.3. 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*)

Although the overall impact of these exclusions was low, the impact varied substantially at the plan/contract level. Within the Medicare population, there was significant impact of the ESRD exclusion, with up to 22% of beneficiaries in some plan-contracts impacted by this exclusion (Statins). The impact of hospice was lower, with up to 5% of beneficiaries impacted in some plan-contracts. There was also significant impact of applying the insulin requirement to the diabetes population (up to 26%) and a lesser impact of the sacubitril/valsartan exclusion to the RASA population (up to 3%). Without applying these exclusions, these beneficiaries would be included in the measure. These are significant proportions of the population that could potentially impact the measure rate at the plan-contract level.

Similar to the Medicare population, although the overall impact of the exclusions was low, the impact varied by plan with the ESRD exclusion showing the most impact. The results show that in some plans, up to 8% of the population had ESRD and would be included in the measure if ESRD was not excluded. In addition, up to 7% of the populations were impacted by the insulin exclusions. These are significant proportions of the population that could potentially impact the measure rate.

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

**2b3. 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*** [***2b4***](#section2b5)***.***

**2b3.1. What method of controlling for differences in case mix is used?**

**No risk adjustment or stratification**

**Statistical risk model with** 5 sociodemographic **risk factors**

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

**Other,** Click here to enter description

**2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.**

Within-Contract Outcome Comparisons

To ensure that beneficiary comparisons were made within plan-contracts, we employed a multivariable, random-effects logistic regression model controlling for the Medicare Part D contract for each therapeutic category rate. This approach acknowledges the variability in an outcome that is attributable to the plan-contract and only allows comparisons among beneficiaries that are in the same plan-contract.

The regression models produce odds ratios (ORs) that assess the increased or decreased odds that beneficiaries with the SDS risk factors will have an outcome, as compared to beneficiaries without those risk factors. Univariate regression models were used to explore the association between the measure outcomes and single covariates. A multivariable regression model was used to explore the association between the measure outcomes adjusted for all covariates that were statically significant based on the univariate models.

SDS Risk Factors

Using variables identified through literature review and subject matter experts, initial univariate and multivariable analyses included beneficiary-, community- (9-digit zip code), and county-level variables. Because prescription drug data have little information regarding beneficiaries’ socioeconomic status, each beneficiary’s 9-digit zip code of residence was linked to zip code-specific socioeconomic data contained in the Acxiom InfoBase Geo data procured by PQA. County-level variables from the publicly available 2015-2016 Health Resources & Services Administration Area Health Resource Files data were also linked to the dataset via beneficiary 5-digit zip code. All variables were classified as categorical variables. (see Tables 11-13).

In addition to the full model using all the variables of interest (Table 4), the same methods were used to assess the impact of a more parsimonious model, which only included the beneficiary-level risk factors that are available in CMS PDE data (i.e., age, gender, LIS/dual status, disability and race). The reduced model was used to assess whether these commonly available beneficiary-level SDS variables would yield similar or different results than the full model.

Finally, to address concerns raised by the PQA Risk Adjustment Advisory Panel (RAAP) members around the accuracy and completeness of the race variable, as well as the NQF Disparities Standing Committee concern around the use of race as a proxy for socioeconomic status, PQA also looked at the reduced model described above without race, i.e. limiting to age, gender, LIS/dual eligibility status and disability. These models were compared to the full model to determine the impact of the 9-digit zip code level characteristics and race on the outcomes.

Predictive Ability of Model and Multicollinearity

C-statistics and risk decile plots were used to assess the predictive validity and discrimination of the models. The Variance Inflation Factor (VIF) was calculated to determine if any variables were multicollinear and which variables to ultimately include in the models. A threshold of VIF greater than 10 was used to determine if variables were multicollinear, and therefore needed to be excluded from the model.

Risk-Adjusted Score Calculation

Using the variables from the most parsimonious model (i.e., age, gender, LIS/dual, and disability status), a risk-adjusted score was calculated for each Medicare Part D contract, for each of the three therapeutic categories, to determine the extent of score change after risk adjustment. For each Part D contract, the expected measure rate was calculated as the average of the patient predicted probabilities of adherence for each plan-contract based on the multivariable logistic regression model. The risk-adjusted measure score for each plan-contract was then calculated as the ratio of observed (or unadjusted) measure score to the expected score, multiplied by the aggregate unadjusted score for all Part D contracts.

Equations

Adherence to medication was modeled as:

where *Pij* is the probability of being adherent for beneficiary *i* in the plan-contract *j*, βk are model parameters; *xkij* are values of variables being adjusted for including the categorical variables for age, gender, LIS/dual and disability status, *αj* is an intercept term. The intercept term is taken to be random and different for each plan-contract.1

Coefficients

See Section 2b3.4a for coefficient estimates, odds ratios and 95% confidence intervals for the risk factors included in the final model.

Definitions for SDS Risk Factors Available and Analyzed

The beneficiary-level variables were obtained from the CMS RIF data, and defined as follows:

* Age: This is the member’s age calculated at the beginning of the measurement year using the member’s date of birth.
* Gender: Refers to the member’s gender as identified in the RIF data.
* Low-Income Subsidy (LIS) status: Refers to Medicare beneficiaries with income below 150% of the federal poverty level and limited resources who receive additional premium and cost-share assistance for prescription drugs under Medicare Part D. The LIS status is indicated in the RIF data if a beneficiary received a low-income subsidy at least one month during the measurement year.
* Dual eligibility status: Denotes beneficiaries enrolled in both Medicare and Medicaid. The dual eligible status is indicated if the beneficiary was dual eligible at least one month in the measurement year.
* Disability as reason for Medicare entitlement: Indicates whether a beneficiary was eligible for Medicare because of a disability, at the end of the measurement year.
* Race: Denotes the race of the beneficiary, as identified in the RIF data.

The community-level variables were obtained from the Acxiom InfoBase Geo datalinked to the CMS RIF data at the 9-digit zip code level. The variables are as follows:

* Median household income: Refers to the median income range of the households in the geographic area.
* Percent of households where residents are married: Denotes the number of households containing married individuals.
* Percent of households where residents completed college: Refers to the number of households where the first individual has a college degree.
* Percent of households where residents own their home: Denotes the number of households that own their home.

The county-level variables were obtained from the HRSA area resource file and linked to the PDE data at the 5-digit zip code level.2 The variables are as follows:

* Federally designated primary care professional shortage areas: Refers to beneficiaries living in areas designated by HRSA as having a shortage in primary care practitioners as reported in 2015-2016. The following criteria are used to determine primary care shortage areas:

1. The area is a rational area for the delivery of primary medical services.
2. One of the following conditions prevails within the area:
   1. The area has a population to full-time-equivalent primary care physician ratio of at least 3,500:1.
   2. The area has a population to full-time-equivalent primary care physician ratio of less than 3,500:1 but greater than 3,000:1 and has unusually high needs for primary care services or insufficient capacity of existing primary care providers.
3. Primary medical care professionals in contiguous areas are overutilized, excessively distant or inaccessible to the population of the area under consideration.

* Federally designated mental healthcare professional shortage areas: Denotes beneficiaries living in areas designated by HRSA as having a shortage in mental health practitioners as reported in 2015-2016. The following criteria are used to determine mental health shortage areas:

1. The area is a rational area for the delivery of mental health services.
2. One of the following conditions prevails within the area:
   1. The area has:
      1. population-to-core-mental-health-professional ratio greater than or equal to 6,000:1 and a population-to-psychiatrist ratio greater than or equal to 20,000:1 or
      2. a population-to-core-professional ratio greater than or equal to 9,000:1 or
      3. a population-to-psychiatrist ratio greater than or equal to 30,000:1;
   2. The area has unusually high needs for mental services, and has:
      1. population-to-core-mental-health-professional ratio greater than or equal to 4,500:1 and a population-to-psychiatrist ratio greater than or equal to 15,000:1 or
      2. a population-to-core-professional ratio greater than or equal to 6,000:1, or
      3. a population-to-psychiatrist ratio greater than or equal to 20,000:1.
3. Mental health professionals in contiguous areas are overutilized, excessively distant or inaccessible to residents of the area under consideration.

**References:**

1. Dharmarajan S, Bentley JP, Banahan BF, West-Strum DS. Measuring pharmacy performance in the area of medication adherence: addressing the issue of risk adjustment. J Manag Care Spec Pharm. 2014;20(10):1057-68. PMID: [25278328](https://www.ncbi.nlm.nih.gov/pubmed/25278328).
2. Health Resources and Services Administration. Health Professional Shortage Area (HPSA) Application and Scoring. Available at: <http://bhpr.hrsa.gov/shortage/hpsas/designationcriteria/designationcriteria.html>.

**2b3.2. If an outcome or resource use component 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**.

After careful review and consideration of all existing PQA measures, the RAAP recommended the PQA PDC 3-rates by therapeutic category measure for risk adjustment considerations as these are currently used for performance evaluation in a national payment program – the Medicare Part D Star Ratings program. To that end, PQA decided to focus on Medicare, and will evaluate SDS risk adjustment for Medicaid as these measures are incorporated into Medicaid programs (such as the Medicaid adult core set). As a result, all the risk adjustment work was focused on the Medicare population, with Medicaid risk adjustment to be considered in the future.

**2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk 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*) **Also discuss any “ordering” of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

The PQA Risk Adjustment Advisory Panel (RAAP)

To address the issue of whether to adjust performance measures for SDS, the National Quality Forum (NQF) convened an expert panel in 2014. The NQF panel recommended that performance-based measures should be risk-adjusted for sociodemographic factors if these criteria are met: (1) there is a conceptual relationship between SDS and the outcome(s), and (2) there is empirical evidence that SDS affects the outcome(s) of interest.1

In light of the NQF trial period that began in April 2015, PQA convened a Risk Adjustment Advisory Panel (RAAP), comprised of representatives with experience in healthcare disparities, risk adjustment methods, and medication adherence research (see Table 9 for list of RAAP members). The goals of the group included: 1) identifying which PQA measures may be appropriate for SDS risk adjustment; and 2) recommending a valid risk adjustment methodology for those measures, which included determining which SDS variables to use for adjustment and how to report the measure rates by plan/contract. The panel decided to focus on the PQA PDC 3-rates by therapeutic category measure (Diabetes, RASA, and Statins).

The RAAP met monthly over the course of 18 months. Through a systematic review of literature, discussion, and voting, the RAAP selected variables for risk adjustment and developed a valid risk adjustment model for the three therapeutic categories.

**Table 9. PQA Risk Adjustment Advisory Panel Members**

| **Name** | **Organization** |
| --- | --- |
| John Bentley | The University of Mississippi |
| Greg Berger | America’s Health Insurance Plans (AHIP) |
| Anton Berisha | Lexis Nexis |
| Heather Black | Merck |
| Joyce Chan | Health First |
| Rebecca Chater | Ateb, Inc. |
| Kelly Conn | St. John Fisher College |
| Jeff Cooley | Humana |
| Joseph Couto | Cigna |
| Patrick Gleason | Prime Therapeutics |
| Kelly Hollenack | ZA Pharma |
| Rita Hui | Kaiser Permanente |
| Taline Jaghasspanian | Health Net |
| Tom Kornfield | America’s Health Insurance Plans (AHIP) |
| Patrick Meek | Albany College of Pharmacy and Health Sciences |
| Brian Meissner | Bristol Myers Squibb |
| David Nau | College of Pharmacy Nova Southeastern University |
| Kyle Null | Takeda |
| Jennifer Polinski | CVS Health |
| Nathaniel Rickles (Chair) | Northeastern University |
| Rene Saucedo | University of Florida |
| Xi Tan | West Virginia University |
| Christie Teigland | Inovalon |

Variable Selection

The RAAP recommended a list of potential risk factors to examine, based on a conceptual framework related to medication adherence2 as well as a review of published literature. The underlying conceptual framework that was selected focused on older adults, given that the Medicare population is primarily 65 years and older. This framework recognizes inherent challenges attributed to age-related factors (such as declining cognitive and physical functions), but also the importance of social factors, environmental and financial constraints that may impact medication adherence.

Variables selected based on a review of published literature included risk factors that are not directly under the control of providers, as these would be most appropriate for risk adjustment. Table 10 below lists the SDS variables that negatively correlate with adherence that were identified from a meta-analysis3 and other supporting literature.4-18

**Table 10. SDS Variables and Literature on Negative Effect on Medication Adherence**

| **SDS Variable** | | **Negative Effect on Adherence3** |
| --- | --- | --- |
| Beneficiary-level | Age | * Age - older and younger age groups (vs. adults)4 * Very old age (older than 85 years)5 |
| Gender | * Male4,5 |
| Low-income subsidy status or dual eligibility status | * Low income4,6 * Poverty7,8 * Lower socioeconomic status9,10 * Financial constraints11,12 |
| Disability as original reason for Medicare entitlement | * Disability13,14 |
| Race | * Latinos (vs. Euro-Americans)10 * Hispanic patients (in the US, in TB)4 * Monolingual Spanish speakers10 * Non-white women5 |
| Community-level | Median income | * Financial constraints11,12 * Poverty7,8 * Lower socioeconomic status4,10 |
| Percent of households where residents are married | * Single or divorced (vs. married)6,15 |
| Percent of households where residents completed college | * Illiteracy4 * (inverse) education9, 11 |
| Percent of households where residents own their home | * Unstable housing16,17 |
| County-level | Federal designated healthcare provider shortage area | * Barriers to high-quality care10 * Lack of providers /caregiver availability8 * Rural settings8 * Poor access to a healthcare facility (e.g., long waiting times, inconvenient opening hours)6 * Poor follow-up by providers17,18 |
| *SDS: sociodemographic status* | | |

Statistical Methods

Univariate regression models were used to explore the association between the measure outcomes and single covariates. Variables that were statistically significant at alpha = 0.05 for any of the therapeutic categories were subsequently included in the multivariable regression model. The multivariable regression model was used to explore the association between the measure outcomes adjusted for all covariates that were statically significant based on the univariate models. Multicollinearity was not observed (VIF <10 for all variables), and no interaction terms were included in any of the models.

**References:**

1. National Quality Forum. Risk adjustment for socioeconomic status or other sociodemographic factors. 2014. Available at: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=77474.
2. Murray MD, Morrow DG, Weiner M, et al. A conceptual framework to study medication adherence in older adults. Am J Geriatr Pharmacother. 2004;2(1):36-43. [PMID: 15555477](https://www.ncbi.nlm.nih.gov/pubmed/15555477).
3. Kardas P, Lewek P, Matyjaszczyk M. Determinants of patient adherence: a review of systematic reviews. Front Pharmacol. 2013;25;4:91. PubMed [PMID: 23898295](https://www.ncbi.nlm.nih.gov/pubmed/23898295).
4. Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J. Patient adherence to tuberculosis treatment: a systematic review of qualitative research. PLoS Med. 2007; 24;4(7):e238. [PMID: 17676945](https://www.ncbi.nlm.nih.gov/pubmed/17676945).
5. Ruddy K, Mayer E, Partridge A. Patient adherence and persistence with oral anticancer treatment. CA Cancer J Clin. 2009; 59(1):56-66. [PMID: 19147869](https://www.ncbi.nlm.nih.gov/pubmed/19147869).
6. Jindal RM, Joseph JT, Morris MC, Santella RN, Baines LS. Noncompliance after kidney transplantation: a systematic review. Transplant Proc. 2003; 35(8):2868-72. PMID: [14697924](https://www.ncbi.nlm.nih.gov/pubmed/14697924).
7. Costello K, Kennedy P, Scanzillo J. Recognizing nonadherence in patients with

multiple sclerosis and maintaining treatment adherence in the long term. Medscape

J Med. 2008; 10(9):225. [PMID: 19008986](https://www.ncbi.nlm.nih.gov/pubmed/19008986).

1. Vreeman RC, Wiehe SE, Pearce EC, Nyandiko WM. A systematic review of pediatric adherence to antiretroviral therapy in low- and middle-income countries. Pediatr Infect Dis J. 2008 ;27(8):686-91. [PMID: 18574439](https://www.ncbi.nlm.nih.gov/pubmed/18574439).
2. Young GJ, Rickles NM, Chou CH, Raver E. Socioeconomic characteristics of enrollees appear to influence performance scores for Medicare part D contractors. Health Aff. 2014; 33(1):140-6. [PMID: 24395946](https://www.ncbi.nlm.nih.gov/pubmed/24395946).
3. Lanouette NM, Folsom DP, Sciolla A, Jeste DV. Psychotropic medication nonadherence among United States Latinos: a comprehensive literature review. Psychiatr Serv. 2009 ;60(2):157-74. [PMID: 19176409](https://www.ncbi.nlm.nih.gov/pubmed/19176409).
4. Oehl M, Hummer M, Fleischhacker WW. Compliance with antipsychotic treatment. Acta Psychiatr Scand Suppl. 2000; (407):83-6. [PMID: 11261648](https://www.ncbi.nlm.nih.gov/pubmed/11261648).
5. Mills EJ, Nachega JB, Bangsberg DR, et al. Adherence to HAART: a systematic review of developed and developing nation patient-reported barriers and facilitators. PLoS Med. 2006 ;3(11):e438. [PMID: 17121449](https://www.ncbi.nlm.nih.gov/pubmed/17121449).
6. Inovalon. An investigation of Medicare Advantage dual-eligible member level performance on CMS Five Star quality measures. Available at: <http://resources.inovalon.com/papers/an-investigation-of-medicare-advantage-dual-eligible-member-level-performance-on-cms-five-star-quality-measures>. Accessed April 2017.
7. National Academies Press. Accounting for Social Risk Factors in Medicare Payment: Identifying Social Risk Factors. 2016. Available at <http://www.nap.edu/21858>.
8. Julius RJ, Novitsky MA Jr, Dubin WR. Medication adherence: a review of the literature and implications for clinical practice. J Psychiatr Pract. 2009; 15(1):34-44. [PMID: 19182563](https://www.ncbi.nlm.nih.gov/pubmed/19182563).
9. Hirsch-Moverman Y, Daftary A, Franks J, Colson PW. Adherence to treatment for latent tuberculosis infection: systematic review of studies in the US and Canada. Int J Tuberc Lung Dis. 2008; 12(11):1235-54. [PMID: 18926033](https://www.ncbi.nlm.nih.gov/pubmed/18926033).
10. Reisner SL, Mimiaga MJ, Skeer M, Perkovich B, Johnson CV, Safren SA. A review of HIV antiretroviral adherence and intervention studies among HIV-infected youth. Top HIV Med. 2009; 17(1):14-25. [PMID: 19270345](https://www.ncbi.nlm.nih.gov/pubmed/19270345).
11. Gold DT, Alexander IM, Ettinger MP. How can osteoporosis patients benefit more from their therapy? Adherence issues with bisphosphonate therapy. Ann Pharmacother. 2006; 40(6):1143-50. [PMID: 16735667](https://www.ncbi.nlm.nih.gov/pubmed/16735667).

**2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:**

**Published literature**

**Internal data analysis**

**Other (please describe**) – subject matter experts consensus

**2b3.4a. What were the statistical results of the analyses used to select risk factors?**

Univariate random effects logistic regression models were used to explore the association between the measure outcomes and single covariates to determine which risk factors were associated with adherence. Variables that were statistically significant at alpha = 0.05 for any of the therapeutic categories were included in the multivariable random effects logistic regression models. Tables 11-13 show the coefficients, odds ratios (ORs), and 95% confidence interval (CI) for the final univariate and multivariable regression models.

*Note: The univariate models were run with one variable at a time but reported together in one table. The multivariable models were run with all variables at the same time in the model. p ≤0.001 for all variables in univariate analysis unless otherwise specified. All analyses were conducted using SAS Foundation 9.4.*

**Table 11. Univariate (Unadjusted) and Multivariable (Adjusted) Logistic Regression Results – Diabetes**

|  | **Unadjusted** | | | **Adjusted** | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | **Coeff Estimate** | **OR** | **95% CI** | **Coeff Estimate** | **OR** | **95% CI** | **p-Value** |
| Age Group (Ref = 80+) |  |  |  |  |  |  |  |
| 18-54 | -0.6429 | 0.53 | 0.50, 0.53 | -0.1491 | 0.86 | 0.79, 0.94 | 0.0004 |
| 55-64 | -0.2558 | 0.77 | 0.75, 0.80 | 0.0936 | 1.10 | 1.03, 1.17 | 0.0054 |
| 65-69 | -0.0100 | 0.99 | 0.96, 1.02 | -0.0290 | 0.97 | 0.94, 1.00 | 0.0768 |
| 70-74 | 0.0357 | 1.04 | 1.00, 1.07 | 0.0145 | 1.02 | 0.98, 1.08 | 0.3940 |
| 75-79 | 0.0447 | 1.05 | 1.01, 1.08 | 0.0287 | 1.03 | 0.99, 1.07 | 0.1180 |
| Gender (Ref = Female) |  |  |  |  |  |  |  |
| Male | -0.1403 | 0.87 | 0.85, 0.89 | -0.1329 | 0.88 | 0.86, 0.89 | <.0001 |
| LIS or Dual (Ref = Non-LIS/Non-Dual) | -0.3054 | 0.74 | 0.72, 0.75 | -0.1666 | 0.85 | 0.83, 0.87 | <.0001 |
| Disability (Ref = No Disability) | -0.5041 | 0.60 | 0.59, 0.62 | -0.4408 | 0.64 | 0.60, 0.69 | <.0001 |
| *Coeff: Coefficient; CI: confidence interval; OR: odds ratio; Ref: reference group; LIS: low-income subsidy*  *Note: For age group 65-69, the p-value in the univariate analysis was > 0.05.* | | | | | | | |

**Table 12. Univariate (Unadjusted) and Multivariable (Adjusted) Logistic Regression Results – RASA**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Unadjusted** | | | **Adjusted** | | | |
| **Characteristic** | **Coeff Estimate** | **OR** | **95% CI** | **Coeff Estimate** | **OR** | **95% CI** | **p-value** |
| Age Group (Ref = 80+) |  |  |  |  |  |  |  |
| 18-54 | -0.7257 | 0.48 | 0.47, 0.50 | -0.0942 | 0.91 | 0.86, 0.96 | 0.0008 |
| 55-64 | -0.2357 | 0.79 | 0.77, 0.81 | 0.1978 | 1.22 | 1.17, 1.27 | <.0001 |
| 65-69 | 0.0889 | 1.09 | 1.07, 1.12 | 0.0683 | 1.07 | 1.05, 1.09 | <.0001 |
| 70-74 | 0.1289 | 1.14 | 1.11, 1.16 | 0.1076 | 1.11 | 1.09, 1.14 | <.0001 |
| 75-79 | 0.0844 | 1.09 | 1.06, 1.11 | 0.0693 | 1.07 | 1.05, 1.10 | <.0001 |
| Gender (Ref = Female) |  |  |  |  |  |  |  |
| Male | -0.0096 | 0.990 | 0.98, 1.00 | -0.0030 | 1.00 | 0.98, 1.01 | 0.6583 |
| LIS or Dual (Ref = Non-LIS/Non-Dual) | -0.5205 | 0.59 | 0.58, 0.60 | -0.3902 | 0.68 | 0.67, 0.69 | <.0001 |
| Disability (Ref = No Disability) | -0.5990 | 0.55 | 0.54, 0.56 | -0.4648 | 0.63 | 0.60, 0.66 | <.0001 |
| *Coeff: Coefficient; CI: confidence interval; OR: odds ratio; Ref: reference group; LIS: low-income subsidy*  *Note: For gender, the p-value for males in the univariate analysis was >0.05.* | | | | | | | |

**Table 13. Univariate (Unadjusted) and Multivariable (Adjusted) Logistic Regression Results – Statins**

|  | **Unadjusted** | | | **Adjusted** | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | **Coeff Estimate** | **OR** | **95% CI** | **Coeff Estimate** | **OR** | **95% CI** | **p-Value** |
| Age Group (Ref = 80+) |  |  |  |  |  |  |  |
| 18-54 | -0.5927 | 0.55 | 0.54, 0.57 | -0.0631 | 0.94 | 0.90, 0.99 | 0.0105 |
| 55-64 | -0.2826 | 0.75 | 0.74, 0.77 | 0.0857 | 1.09 | 1.05, 1.13 | <.0001 |
| 65-69 | -0.1156 | 0.89 | 0.88, 0.91 | -0.1408 | 0.87 | 0.85, 0.88 | <.0001 |
| 70-74 | -0.0566 | 0.95 | 0.93, 0.96 | -0.0826 | 0.92 | 0.91, 0.94 | <.0001 |
| 75-79 | -0.0543 | 0.95 | 0.93, 0.97 | -0.0734 | 0.93 | 0.91, 0.95 | <.0001 |
| Gender (Ref = Female) |  |  |  |  |  |  |  |
| Male | -0.1490 | 0.86 | 0.85, 0.87 | -0.1473 | 0.86 | 0.85, 0.87 | <.0001 |
| LIS or Dual (Ref = Non-LIS/Non-Dual) | -0.3569 | 0.70 | 0.69, 0.71 | -0.2581 | 0.77 | 0.76, 0.78 | <.0001 |
| Disability (Ref = No Disability) | -0.4255 | 0.65 | 0.64, 0.66 | -0.4393 | 0.64 | 0.62, 0.67 | <.0001 |
| *Coeff: Coefficient; CI: confidence interval; OR: odds ratio; Ref: reference group; LIS: low-income subsidy* | | | | | | | |

**2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk 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.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Empirical Association with Outcome (Univariate Analysis)

For all three therapeutic categories, younger beneficiaries were less likely to be adherent compared to the older population (e.g., OR = 0.53, 0.48 and 0.55 for age group 18-54 years, for Diabetes, RASA, and Statins, respectively; OR = 0.99, 1.09, and 0.89 for age group 65-69 years for Diabetes, RASA, and Statins, respectively). For Diabetes and Statins, males were less likely to be adherent than females (OR = 0.87 and 0.86, respectively) while for RASA, there was no statistical difference in adherence between males and females. The LIS/Dual population was less likely to be adherent to medications for all three therapeutic categories (OR = 0.74. for Diabetes, OR = 0.59 for RASA, and OR = 0.70 for Statins). For all therapeutic categories, those persons with disability as the reason for Medicare entitlement were less likely to be adherent to their medications (OR = 0.60 for Diabetes, OR = 0.55 for RASA, and OR = 0.65 for Statins).

Low vs. High Performing Plan-Contracts (Multivariable Analysis)

Several members of the RAAP expressed concerns that SDS risk adjustment may mask real disparities and create lower standards of performance for beneficiaries in disadvantaged populations. To assess the impact of risk adjustment on the highest and lowest performing plan-contracts, plan-contracts were ranked based on unadjusted and risk-adjusted scores. (see Figures 1-3). For all three therapeutic categories, the lowest performing plan-contracts prior to SDS risk adjustment continued to perform poorly after risk adjustment. This trend also held true for the highest performing plan-contracts.

This showed that very poor performers and top performers stay the same after risk adjustment, with most of the movement occurring in the middle performing plan-contracts. Thus, risk adjustment provided a more accurate reflection of the relative risk of the population of the plan-contract, but still showed which plan-contracts were performing worse compared to plan-contracts serving similar beneficiaries.

*Note: The higher the ranking, the better the plan-contract performance*

**Figure 1. Unadjusted vs. Risk-Adjusted Rankings – Diabetes**

**Figure 2. Unadjusted vs. Risk-Adjusted Rankings – RASA**

**Figure 3. Unadjusted vs. Risk-Adjusted Rankings – Statins**

**2b3.5. 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*)

Risk adjustment is important in making fair comparisons among plan-contracts. However, it may introduce an uncertainty when alternative statistical methodologies do not agree on which plan-contracts are identified as high- and low-quality.1-3 Based on literature review4-6 and discussions with subject matter experts, the RAAP identified the multivariable, random effects logistic regression model as the suitable method for SDS risk adjustment. To validate this statistical methodology, we considered an alternative approach, using the classical logistic regression model, which assumes that individuals are independent observations in the dataset, and does not account for the nesting of contracts within plans.

We hypothesized that risk-adjusted measure scores would show better overall agreement with each other than with the unadjusted compared to each risk-adjusted score. We compared the two methods and looked at the weighted kappa as the measure of agreement.

**References:**

1. Iezzoni LI. The risks of risk adjustment. JAMA. 1997;278(19):1600-7. [PMID: 9370507](https://www.ncbi.nlm.nih.gov/pubmed/9370507).
2. DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S, Mark DB. Comparing risk-adjustment methods for provider profiling. Stat Med. 1997;16(23):2645-64. [PMID: 9421867](https://www.ncbi.nlm.nih.gov/pubmed/9421867).
3. Li Y, Dick AW, Glance LG, et al. Misspecification Issues in Risk Adjustment and Construction of Outcome-Based Quality Indicators.’ Health Services and Outcomes Research Methodology 2007;7: 39–56.
4. Dharmarajan S, Bentley JP, Banahan Iii BF, West-Strum DS. Measuring pharmacy performance in the area of medication adherence: addressing the issue of risk adjustment. J Manag Care Spec Pharm. 2014;20(10):1057-68. [PMID: 25278328](https://www.ncbi.nlm.nih.gov/pubmed/25278328).
5. Young GJ, Rickles NM, Chou CH, Raver E. Socioeconomic characteristics of enrollees appear to influence performance scores for Medicare part D contractors. Health Aff. 2014;33(1):140-6. [PMID: 24395946](https://www.ncbi.nlm.nih.gov/pubmed/24395946).
6. Li Y, Cai X, Glance LG, Spector WD, Mukamel DB. National release of the nursing home quality report cards: implications of statistical methodology for risk adjustment. Health Serv Res. 2009;44(1):79-102. [PMID: 19146565](https://www.ncbi.nlm.nih.gov/pubmed/19146565).

*Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below*.  
***If stratified, skip to*** [***2b3.9***](#question2b49)

**2b3.6. Statistical Risk Model Discrimination Statistics** (*e.g., c-statistic, R-squared*)**:**Table 14 shows the area under the receiver operating characteristic (ROC) curve (the c-statistic) for the primary method used for risk adjustment (random effects logistic regression models) for each therapeutic category.

**Table 14. C-statistic for Random Effects Logistic Regression Models – By Therapeutic Category**

|  |  |
| --- | --- |
| **Therapeutic Category** | **c-statistic** |
| Diabetes | 0.583 |
| RASA | 0.597 |
| Statins | 0.591 |

We also show the agreement between the unadjusted and the primary method (random effects), as well as the primary method and the alternative method (classical logistic regression). (see Table 15).

**Table 15. Agreement (Weighted Kappa) Between Measure Scores – By Therapeutic Category**

|  |  |  |  |
| --- | --- | --- | --- |
| **Comparison** | **Diabetes** | **RASA** | **Statins** |
| Unadjusted vs. Random effects model | 0.803 (0.775, 0.831) | 0.693 (0.658, 0.729) | 0.748 (0.718, 0.777) |
| Random effects vs. Classical logistic regression | 0.989 (0.984, 0.995) | 0.977 (0.969, 0.984) | 0.975 (0.967, 0.983) |

**2b3.7. Statistical Risk Model Calibration Statistics** (*e.g., Hosmer-Lemeshow statistic*):

The generalized linear mixed (GLIMMIX) procedure with random intercept in SAS was used for risk adjustment. While this method is well suited for binary measures, there is no formal statistic to assess the goodness of fit (GOF) for these models. We used an extension of the Hosmer-Lemeshow (H-L) GOF test developed and validated by Li et al1 to determine the model calibration.

In large datasets, small/unimportant deviations from good calibration can still lead to large H-L test statistic or small p-value, and therefore may not be useful. As such we provide risk decile plots to measure predictive ability.

**References:**

1. Li R, Su Z, Mendelsohn A, Gemmen E. Extension of the Hosmer-Lemeshow Goodness of Fit statistic to linear models with repeated measurements. Value in Health 16(2013) A1-A298.

**2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves**:

The risk decile plot provides a graphical representation of the deciles calculated to measure predictive ability as calculated above in 2b3.7. Figures 4-6 show the distribution of the observed and expected PDC deciles for each of the therapeutic categories.

**Figure 4. Risk Decile Plot for Diabetes**

**Figure 5. Risk Decile Plot for RASA**

**Figure 6. Risk Decile Plot for Statins**

**2b3.9. Results of Risk Stratification Analysis**:

N/A

**2b3.10. 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 comparison of agreement between the two risk adjustment methodologies showed almost perfect agreement, with kappa = 0.99 for Diabetes, 0.98 for RASA and 0.97 for Statins. This was much higher than the unadjusted vs. random effects models (0.69-0.80). This showed that SDS risk adjustment was more appropriate for evaluating performance scores compared to the unadjusted measure scores. In addition, as shown in Tables 19-21, we observed significant shifts in deciles post risk-adjustment, with over 50% of plan-contracts changing deciles.

The c-statistic is used to assess model discrimination, and ranges from 0.5 to 1.0 with 0.5 indicating the model is no better than random prediction and 1.0 showing perfect prediction. In research, a c-statistic of 0.7 or greater indicate acceptable discrimination. However, with performance measurement, the purpose of risk adjustment is to reduce bias due to patient characteristics present at the start of care, not to completely explain variations in outcomes, and therefore does not include variables related to quality of care.

It is important to note that the variables included in this analysis have been found to have an impact on outcome measures in other studies. This suggests that although the covariates used for risk adjustment in this study are important, there may be other covariates that could improve the model, such as clinical variables for diagnoses, disease severity, etc. As with any risk adjustment modeling, the model can only account for measurable and available covariates. Therefore, if any unmeasured factors are not randomly distributed within contracts, the risk adjustment methodology may not adequately mitigate the impact of these unmeasured factors.

Finally, the risk decile plots show that the higher deciles of the predicted outcomes were associated with higher observed outcomes. In addition, within each decile, there is no meaningful discrepancy between the observed PDC score in a decile and that predicted by the model, which shows good discrimination and predictive ability of the models.

**2b3.11.** **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*)

In 2016, PQA contracted with CMS to examine the PDC 3-rates measure (Diabetes, RASA, and Statins) to determine if SDS risk adjustment was needed. As part of this work, PQA received the 2014 CMS PDE data containing information (including age, gender, LIS/dual eligibility status, race, disability as reason for Medicare entitlement, flag for numerator inclusion, etc.) about individuals eligible for inclusion in the measure calculations for each of the three therapeutic categories. PQA, working with the RAAP conducted the study to examine SDS risk adjustment for the Medicare population, and based on the results of the study, recommended that all three therapeutic categories should be risk-adjusted for age, gender, LIS/dual eligibility status and disability status.

Part of the study included examining not just the beneficiary-level risk factors available in the CMS PDE data, but community- and county-level variables. A review of the three models (full – with all beneficiary, community and county; reduced with race, which included all beneficiary-level variables; and reduced without race, which included all beneficiary-level variables except race) showed similar results in terms of magnitude and direction of the odds ratios, with one exception in the PDC Diabetes model, where the odds ratio for the 65-69 years age group changed from 1.02 in the reduced model with race to 0.97 in the reduced model without race. In addition, the movement of plan-contracts post risk adjustment was similar for all three models. (see Tables 16-18).

**Table 16. Risk-Adjusted Decile Rankings for Full vs. Reduced Models – Diabetes**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Full Model** | **Reduced**  **(w/ race)** | **Reduced**  **(w/o race)** |
| Contracts adjusted to a higher decile | 25.8% | 25.0% | 20.2% |
| contracts adjusted to a lower decile | 44.6% | 43.4% | 40.3% |
| contracts adjusted to the same decile | 29.6% | 31.5% | 39.5% |
| Average (absolute) change in decile | 1.6 | 1.4 | 0.9 |
| Average decile change for contracts adjusted to a higher decile | 2.5 | 2.3 | 2.2 |
| Average decile change for contracts adjusted to a lower decile | -2.0 | -1.6 | -1.1 |

**Table 17. Risk-Adjusted Decile Rankings for Full vs. Reduced Models – RASA**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Full Model** | **Reduced**  **(w/ race)** | **Reduced**  **(w/o race)** |
| Contracts adjusted to a higher decile | 26.7% | 25.7% | 24.4% |
| contracts adjusted to a lower decile | 46.4% | 47.6% | 46.8% |
| contracts adjusted to the same decile | 26.9% | 26.7% | 28.8% |
| Average (absolute) change in decile | 1.3 | 1.3 | 1.1 |
| Average decile change for contracts adjusted to a higher decile | 2.5 | 2.5 | 2.3 |
| Average decile change for contracts adjusted to a lower decile | -1.4 | -1.4 | -1.2 |

**Table 18. Risk-Adjusted Decile Rankings for Full vs. Reduced Models – Statins**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Full Model** | **Reduced**  **(w/ race)** | **Reduced**  **(w/o race)** |
| Contracts adjusted to a higher decile | 25.2% | 24.5% | 20.6% |
| contracts adjusted to a lower decile | 43.6% | 42.8% | 36.9% |
| contracts adjusted to the same decile | 31.2% | 32.7% | 42.5% |
| Average (absolute) change in decile | 1.1 | 1.1 | 0.8 |
| Average decile change for contracts adjusted to a higher decile | 2.2 | 2.2 | 1.9 |
| Average decile change for contracts adjusted to a lower decile | -1.3 | -1.2 | -1.1 |

The results of the study using the 100% CMS PDE data provide additional validation of the risk adjustment models and results reported in section 2b3 above using the 5% Medicare sample. A comparison of the plan-contract movement at the decile-level post risk adjustment using the final list of recommended SDS risk factors (age, gender, LIS/dual eligibility, and disability status) was similar using the 2014 100% PDE data and the 2016 5% sample. (see Tables 19-21).

**Table 19. Risk-Adjusted Decile Rankings for 100% vs.5% Sample PDE Datasets – Diabetes**

|  |  |  |
| --- | --- | --- |
|  | **2016 5% Sample** | **2014 100% Data** |
| Contracts adjusted to a higher decile | 20.6% | 20.2% |
| contracts adjusted to a lower decile | 29.7% | 40.3% |
| contracts adjusted to the same decile | 49.8% | 39.5% |
| Average (absolute) change in decile | 0.6 | 0.9 |
| Average decile change for contracts adjusted to a higher decile | 1.6 | 2.2 |
| Average decile change for contracts adjusted to a lower decile | -1.1 | -1.1 |

**Table 20. Risk-Adjusted Decile Rankings for 100% vs.5% Sample PDE Datasets – RASA**

|  |  |  |
| --- | --- | --- |
|  | **2016 5% Sample** | **2014 100% Data** |
| Contracts adjusted to a higher decile | 22.7% | 24.4% |
| contracts adjusted to a lower decile | 39.4% | 46.8% |
| contracts adjusted to the same decile | 37.9% | 28.8% |
| Average (absolute) change in decile | 1.0 | 1.1 |
| Average decile change for contracts adjusted to a higher decile | 2.2 | 2.3 |
| Average decile change for contracts adjusted to a lower decile | -1.3 | -1.2 |

**Table 21. Risk-Adjusted Decile Rankings for 100% vs.5% Sample PDE Datasets – Statins**

|  |  |  |
| --- | --- | --- |
|  | **2016 5% Sample** | **2014 100% Data** |
| Contracts adjusted to a higher decile | 22.9% | 20.6% |
| contracts adjusted to a lower decile | 35.4% | 36.9% |
| contracts adjusted to the same decile | 41.7% | 42.5% |
| Average (absolute) change in decile | 0.8 | 0.8 |
| Average decile change for contracts adjusted to a higher decile | 1.8 | 1.9 |
| Average decile change for contracts adjusted to a lower decile | -1.2 | -1.1 |

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

**2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE**

**2b4.1. 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)*

To assess significant differences in measure rates, unadjusted measure rates for Medicare and Medicaid, as well as risk-adjusted rates for Medicare were used to calculate the mean, median, standard deviation, and interquartile range. In addition, the rates were divided into quartiles, and a Student’s t-test was used to compare the rates of the plans/contracts in the 25th percentile to the rates of the plans/contracts in the 75th percentile. Finally, to assess impact of risk adjustment on measure rates, plan-contracts were placed into deciles using the unadjusted as well as the risk adjusted rates and assessed to determine the rate of decile shifts post risk adjustment.

**2b4.2. 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*)

For the Medicare population, for Diabetes, the mean rate (with standard deviation [SD]) was 82.1% (5.5%), for RASA the mean rate (SD) was 85.7% (5.0%) and for Statins the mean rate (SD) was 80.6% (5.9%). (see Table 22).

**Table 22. Means Distribution of Unadjusted Measure Rates for Medicare – By Therapeutic Category**

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistic** | **Diabetes** | **RASA** | **Statins** |
| Mean | 82.1% | 85.7% | 80.6% |
| Std. Deviation (Mean) | 5.5% | 5.0% | 5.9% |
| Minimum | 62.5% | 63.9% | 59.3% |
| 25th Percentile | 78.8% | 83.0% | 77.3% |
| 50th Percentile | 82.5% | 86.7% | 81.7% |
| 75th Percentile | 85.8% | 88.9% | 84.2% |
| Maximum | 96.8% | 97.2% | 97.1% |
| Interquartile Range | 7.0% | 5.9% | 7.0% |
| Student’s t-test p-value | <.0001 | <.0001 | <.0001 |
| *RASA: renin-angiotensin system antagonist* | | | |

For the Medicaid population, for Diabetes, the mean rate (SD) was 59.6% (11.4%), for RASA the mean rate (SD) was 62.2% (10.7%) and for Statins the mean rate (SD) was 58.7% (11.0%). (see Table 23).

**Table 23. Means Distribution of Measure Rates for Medicaid – By Therapeutic Category**

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistic** | **Diabetes** | **RASA** | **Statins** |
| Mean | 59.6% | 62.2% | 58.7% |
| Std. Deviation (Mean) | 11.4% | 10.7% | 11.0% |
| Minimum | 32.1% | 34.1% | 31.9% |
| 25th Percentile | 53.0% | 55.6% | 52.2% |
| 50th Percentile | 60.9% | 63.0% | 59.6% |
| 75th Percentile | 67.1% | 68.7% | 65.6% |
| Maximum | 85.0% | 85.4% | 86.6% |
| Interquartile Range | 14.1% | 13.1% | 13.4% |
| Student’s t-test p-value | <.0001 | <.0001 | <.0001 |
| *RASA: renin-angiotensin system antagonist* | | | |

Table 24 shows the means distribution of the risk-adjusted measure rates for Medicare, with Tables 19-21 above highlighting the impact of risk adjustment on decile rankings for Medicare plan-contracts.

**Table 24. Means Distribution of Risk-Adjusted Measure Rates for Medicare – By Therapeutic Category**

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistic** | **Diabetes** | **RASA** | **Statins** |
| Mean | 81.8% | 85.5% | 80.1% |
| Std. Deviation (Mean) | 5.1% | 4.3% | 5.4% |
| Minimum | 61.1% | 61.0% | 55.9% |
| 25th Percentile | 79.0% | 83.8% | 77.9% |
| 50th Percentile | 81.9% | 85.9% | 80.6% |
| 75th Percentile | 84.7% | 88.0% | 83.4% |
| Maximum | 96.3% | 96.4% | 97.3% |
| Interquartile Range | 5.6% | 4.1% | 5.5% |
| Student’s t-test p-value | <.0001 | <.0001 | <.0001 |
| *RASA: renin-angiotensin system antagonist* | | | |

**2b4.3. 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?*)

As expected, given the importance of these three PDC measure rates in the CMS Part D Star Ratings program (each is triple weighted), the variations were lower in the unadjusted measure rates for Medicare than in the Medicaid population, as health plans work to improve their measure scores to maximize overall Star Ratings and payment. However, there was some variation in Medicare, with a standard deviation ranging from 5.0% (RASA) to 5.9% (Statins) and an interquartile range from 5.9% (RASA) to 7.0% (Diabetes & Statins). There is also a statistically significant difference in measure rates between the top and bottom quartile of the plan-contracts included in the testing (P< .0001 at alpha = 0.05) for all three therapeutic categories in both adjusted and unadjusted measure scores. This variation shows that there are statistically significant and meaningful differences in rates across plan-contracts.

Within Medicaid, the variation was even more pronounced, with a standard deviation ranging from 10.7% (RASA) to 11.4% (Diabetes) and an interquartile range from 13.1% (RASA) to 14.1% (Diabetes). There is a statistically significant difference in measure rates between the top and bottom quartile of the plan-contracts included in the testing (P< .0001 at alpha = 0.05) for all three therapeutic categories. This variation shows that there are statistically significant and meaningful differences in rates across plans.

In addition, a review of the unadjusted and risk-adjusted rates showed that between 21% (Diabetes) and 23% (RASA & Statins) of plan-contracts were adjusted to a higher decile while 30% (Diabetes), 35% (RASA) and 39% (Statins) were adjusted to a lower decile after risk adjustment, showing that risk adjustment had an impact on measure rates at the decile-level. (see Tables 19-21). However, it is unclear what impact SDS risk adjustment will have on the Star Ratings of plan-contracts in the Medicare program.

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

**2b5. 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 social risk 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 social risk 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.***

**2b5.1. 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*)  
N/A

**2b5.2. 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*)  
N/A

**2b5.3. 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*)

N/A  
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

**2b6.1. 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*)  
   
With the use of prescription claims data as the data source for this measure, the dispensing information (including medication, days’ supply, quantity dispensed, and dosage) is available for each beneficiary. Since each of these data elements are available via prescription claims data, it is not expected—nor was it found—that missing data would result.

The final risk adjustment model included only variables available in the CMS PDE data, and were considered to be reliable as many of these variables, including age, LIS status, disability as reason for Medicare entitlement are all important for determining eligibility for enrollment and payment of services.

In addition, race was excluded from the SDS risk adjustment in part because of RAAP concerns about the lack of completeness of the race variable, where about 20% of individuals’ race in the 2014 100% CMS PDE was classified as “unknown”.

**2b6.2. 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*)

Race was found to be “unknown” for about 20% of the population in the 2014 100% CMS PDE and was therefore not included in the final model. As discussed in section 2b3.11, the models were run with and without race to determine impact of race on the models.

**2b6.3. 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*)

As discussed in 2b3.11, a comparison of the models with and without race show that the models were similar for all three rates with respect to the coefficient estimate direction and odds ratios of the included covariates with one exception in the Diabetes model, where the odds ratio for the 65-69 years age group changed from 1.02 in the reduced model with race to 0.97 in the reduced model without race.