NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 2993

Corresponding Measures:

De.2. Measure Title: Potentially Harmful Drug-Disease Interactions in Older Adults (DDE)

Co.1.1. Measure Steward: National Committee for Quality Assurance

De.3. Brief Description of Measure: The percentage of patients 65 years of age and older who have evidence of an underlying disease, condition or health concern and who are dispensed an ambulatory prescription for a potentially harmful medication, concurrent with or after the diagnosis. Three rates are reported for this measure:

- Rate 1: The percentage of those with a history of falls that received a potentially harmful medication
- Rate 2: The percentage of those with dementia that received a potentially harmful medication
- Rate 3: The percentage of those with chronic kidney disease that received a potentially harmful medication

A lower rate represents better performance for all rates.

1b.1. Developer Rationale: Lowering the rate of potentially harmful drug-disease interactions in the older adult population should decrease morbidity and mortality associated with adverse drug reactions.

S.4. Numerator Statement: Numerator 1: Patients with a history of falls who received at least one potentially harmful medication from Table DDE-A or Table DDE-B

Numerator 2: Patients with a diagnosis of dementia who received at least one potentially harmful medication from Table DDE-D

Numerator 3: Patients with chronic kidney disease who received at least one potentially harmful medication from Table DDE-E

S.6. Denominator Statement: All patients 65 years of age and older with a history of falls, dementia or chronic kidney disease in the measurement year or the year prior to the measurement year.

S.8. Denominator Exclusions: For those who meet denominator criteria for the history of falls rate (Rate 1): exclude those with a diagnosis of psychosis, schizophrenia, schizoaffective disorder, bipolar disorder, major depressive disorder or seizure disorder.

For those who meet denominator criteria for the dementia rate (Rate 2): exclude those with a diagnosis of psychosis, schizophrenia, schizoaffective disorder or bipolar disorder.

De.1. Measure Type: Process

S.17. Data Source: Claims

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Jan 26, 2017 **Most Recent Endorsement Date:** Jan 26, 2017

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2017

- During the previous review in 2017, the developer provided evidence based on the AGS Beers Criteria recommendations against the use of potentially harmful medications in older adults with specific conditions.
- The AGS Beers Criteria identifies 12 conditions where there are potentially inappropriate medications. The time period covered by the body of evidence was 2004-14.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- For this review cycle, the developer cited updated evidence based on the *American Geriatrics Society 2019 Beers Criteria Update Expert Panel. 2019. American Geriatrics Society 2019 Updated AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. Journal of the American Geriatrics Society, 67(4): 674-94.*
- Guiding principles on which conditions would be included in the measure were also provided.
- The following changes to the 2019 Beers Criteria were applied to the DDE measure during the most recent update:
    - Added SNRIs to History of Falls rate
    - Removed H2 receptor antagonists from Dementia rate
    - Added an exclusion for major depressive disorder to History of Falls rate
    - Added Pyrilamine to the list of anticholinergics, first-generation antihistamines
    - Added Methscopolamine to the list of anticholinergics, antispasmodics

Exception to evidence

N/A

Questions for the Committee:

- Does the Committee have any concerns related to the Evidence?

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC present (Box 4)→ Quantity: high; Quality: high; Consistency: high (Box 5a)→ High

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer provided data extracted from HEDIS data collection for Medicare Advantage Health Plans (including both HMO and PPO plans). The performance data, which show room for improvement, is summarized at the health plan level and summarized by mean, standard deviation, and performance at the 10th, 25th, 50th, 75th and 90th percentile.

Disparities

- The developer noted that HEDIS data are stratified by type of insurance (e.g., Commercial, Medicaid, Medicare) and that the measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities, if the data are available to a plan.
- HEDIS includes two measures that can be used as tools for assessing race/ethnicity and language needs of a plan's population: Race/Ethnicity Diversity of Membership and the Language Diversity of Membership.
- While disparities for this measure have not been well studied, there is some evidence to suggest differences in the use of potentially inappropriate medications by gender, race, and income status. A cross-sectional study examining the prevalence of potentially inappropriate medications in community-dwelling Medicare beneficiaries in California found that use was significantly higher in women, White beneficiaries, and low-income beneficiaries (Patel et al., 2018).

- In a different study, a retrospective database analysis of HEDIS data from the Department of Veterans Affairs found that Hispanics and those with no copayments had higher rates of medications listed as potentially harmful than Whites or those with required copayments (Pugh, 2011).

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?

Preliminary rating for opportunity for improvement: ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- High rating, no concerns.
- No concerns related to evidence
- Systematic review supports disease-drug interactions
- Evidence supports
- I am not aware of any new studies. New evidence seems fully applicable.
- Evidence supports the adverse effects of these PIMs and selected diagnoses
- The evidence (Beer's Criteria 2019) is a systematic review directly related to the measure. The evidence rating is High

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- High rating, no concerns.
- Gap in performance warrants continued measurement.
- Clear differences in quality between 25th and 75th percentiles, 10% absolute, support finding of quality gaps.
- Existing performance gap
- Wide variability in all 3 measures based on summary data. racial disparities have not been directly characterized. There is indirect evidence of economic disparities.
- Performance gaps should focus on the three areas of focus, and in particular the use of NSAIDs in Black, Hispanic and Pacific Islanders which could exacerbate pre-existing risk and hasten end-stage renal disease.
- As of 2018, out of 402 health plans, history of falls and high risk medication rate range from 25.8% to 76.6&; dementia and high risk medication rate range from 23.6% to 75.0% and CKD and high risk medication rate range from 0 to 39.4%. These significant variations demonstrate opportunity for improvement. Rating is High.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of c measure are when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if 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 enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**

Evaluators: NQF Staff

Staff Review

Evaluation Summary: The appears to be reliable for all three measure rates; however, there is lower reliability in some health plans that fall well below the 0.7 threshold. Validity testing suggested that there was a significant correlation in the direction expected with a similar measure of medication safety in health plans and that there were positive correlations among the 3 measured populations. For face validity, the developer ensured that the measure was aligned with 2019 Beers criteria.

Reliability

- Reliability was tested at the performance score level.
- There are three measure rates for specific underlying conditions where a potentially harmful medication was prescribed:
    - A history of falls and a prescription for anticonvulsants, antipsychotics, benzodiazepines, nonbenzodiazepine hypnotics, or antidepressants.
    - Dementia and a prescription for antipsychotics, benzodiazepines, nonbenzodiazepine hypnotics, tricyclic antidepressants, or anticholinergic agents.
    - Chronic kidney disease and prescription for Cox-2 selective NSAIDs or non-aspirin NSAIDs**.**
- Reliability was assessed in 377 Medicare plans.
- Data was not assessed by social risk factors because that data was not available.
- Signal to noise testing was conducted, as well as SE and 95% CI

Point Estimates of Mean Signal-to-Noise Reliability, 2018

Measure	Point estimate: Mean Signal-To-Noise Reliability
Rate 1 (History of Falls)	0.872
Rate 2 (Dementia)	0.880
Rate 3 (Chronic Kidney Disease)	0.879

Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the *Potentially Harmful Drug-Disease Interactions in Older Adults* Measure by Terciles of the Denominator Size and for All Submissions, 2018

Stratification	Number of Plans	Number of Eligible Members per Plan (min - max)	Mean Signal-To-Noise Reliability	SE	95% CI
Rate 1 (History of Falls)	**402**	**32 - 46847**	**0.872**	**0.007**	**(0.857, 0.886)**
Tercile 1	133	32 – 335	0.704	0.011	(0.682, 0.726)
Tercile 2	132	336 – 1546	0.933	0.002	(0.928, 0.938)
Tercile 3	137	1549 – 46847	0.982	0.001	(0.981, 0.984)
Rate 2 (Dementia)	**407**	**30 - 36535**	**0.880**	**0.007**	**(0.867, 0.893)**
Tercile 1	134	30 – 255	0.616	0.012	(0.593, 0.639)
Tercile 2	134	259 – 1266	0.926	0.003	(0.920, 0.931)
Tercile 3	139	1268 - 36535	0.989	0.001	(0.988, 0.990)
Rate 3 (Chronic Kidney Disease)	**377**	**30 - 16878**	**0.879**	**0.007**	**(0.865, 0.893)**
Tercile 1	125	30 – 153	0.683	0.015	(0.682, 0.726)
Tercile 2	124	155 – 663	0.928	0.004	(0.682, 0.726)
Tercile 3	128	664 - 16878	0.982	0.001	(0.682, 0.726)

Distribution of Plan-Level Signal-To-Noise Reliability for the *Potentially Harmful Drug-Disease Interactions in Older Adults* Measure by Terciles of the Denominator Size and for All Submissions, 2018

Stratification	Number of Plans	Distribution of Plan Estimates of Signal-to-Noise Reliability: Min	Distribution of Plan Estimates of Signal-to-Noise Reliability: P10	Distribution of Plan Estimates of Signal-to-Noise Reliability: P25	Distribution of Plan Estimates of Signal-to-Noise Reliability: P50	Distribution of Plan Estimates of Signal-to-Noise Reliability: P75	Distribution of Plan Estimates of Signal-to-Noise Reliability: P90	Distribution of Plan Estimates of Signal-to-Noise Reliability: Max
Rate 1 (History of Falls)	402	0.378	0.630	0.821	0.934	0.976	0.991	0.999
Tercile 1	133	0.383	0.511	0.595	0.727	0.818	0.857	0.874
Tercile 2	132	0.876	0.892	0.910	0.938	0.958	0.967	0.972
Tercile 3	137	0.962	0.967	0.973	0.985	0.991	0.995	0.999
Rate 2 (Dementia)	407	0.430	0.664	0.815	0.946	0.981	0.992	0.999
Tercile 1	134	0.317	0.396	0.511	0.651	0.726	0.776	0.807
Tercile 2	134	0.850	0.870	0.903	0.938	0.951	0.961	0.967
Tercile 3	139	0.977	0.980	0.985	0.990	0.994	0.997	0.999
Rate 3 (Chronic Kidney Disease)	377	0.329	0.686	0.824	0.937	0.980	0.992	1.000
Tercile 1	125	0.271	0.431	0.580	0.716	0.802	0.860	1.000
Tercile 2	124	0.775	0.866	0.908	0.940	0.965	0.973	0.992
Tercile 3	128	0.908	0.963	0.977	0.986	0.992	0.996	0.999

Validity

- Validity testing was done at the performance measure score level.
- Empirical validity testing was performed for construct validity as compared to a similar measure Use of High-Risk Medications in Older Adults measure, which assesses the percentage of Medicare members ages 65 years and older who had at least 2 dispensing events for the same high-risk medication, and correlation between the 3 different patient populations. Correlations between the DDE measure for the three rates were all positive and varied from 0.24 to 0.63.

Health-Plan Level Pearson Correlation Coefficients Among *Potentially Harmful Drug-Disease Interactions in Older Adults* and *Use of High-Risk Medications in Older Adults* Performance Scores, 2018

Measure	Correlation Coefficient: DDE Rate 1 (History of Falls)	Correlation Coefficient: DDE Rate 2 (Dementia)	Correlation Coefficient: DDE Rate 3 (Chronic Kidney Disease)
DDE Rate 1 (History of Falls)	--	--	--
DDE Rate 2 (Dementia)	0.63	--	--
DDE Rate 3 (Chronic Kidney Disease)	0.24	0.59	--
*Use of High-Risk Medications in Older Adults**	0.61	0.53	0.24

-- cell intentionally left blank

Correlations among all rates in DDE measure and the DAE measure[1]

[Pearson Correlation Coefficients] DDE: Rate 1 (History of Falls)	[Pearson Correlation Coefficients] DDE: Rate 2 (Dementia)	[Pearson Correlation Coefficients] DDE: Rate 3 (Chronic Kidney Disease)	[Pearson Correlation Coefficients] DDE: Rate 4 (Total)	[Pearson Correlation Coefficients] DAE: High-Risk Med, 65+
--	--	--	--	--
0.694	--	--	--	--
0.155	0.585	--	--	--
0.842	0.921	0.480	--	--
0.307	0.454	0.367	0.386	--

-- cell intentionally left blank

- Face validity was performed through interactions advisory panels and NCQA staff, and public review. According to the developer, NCQA worked closely with our multi-stakeholder MAPs to re-evaluate the measure based on the latest recommendations in the American Geriatric Society's 2019 Beers Criteria. The last Beers Criteria update prior to this publication was in 2015. Based on the 2019 Beers Criteria, the primary changes to the measure were updates to medications. The measure changes were evaluated in 2019. After reviewing, the CPM recommended to send the updated measure to public comment with a majority vote in 2019. The measure was released for Public Comment in 2019 prior to publication in HEDIS. Input from advisory panels and the public comment indicate the measure has face validity.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- No concerns
- No concerns
- Reliability testing was strong. Estimates of signal-noise reliability were high across tertiles.
- Agree with moderate prelim rec
- none
- No comment
- Reliability data demonstrates moderate reliability. No concerns.

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- No
- No concerns
- No
- No
- no
- No comment
- No concerns.

2b1. Validity -Testing: Do you have any concerns with the testing results?

No

- No concerns
- No. Good correlations with related measures suggests this one is assessing a robust construct with face validity and construct validity
- No
- no
- No comment
- Face validity is sound and construct validity is done by comparing the three submeasures to one another and to the measure 0022, moderate correlation was found.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- N/A
- No concerns
- Process measure, exclusions seem well informed.
- Agree with prelim rec
- Rationale for exclusions and new exclusions is beyond my understanding.
- No
- No risk adjustment which is appropriate. Exclusion is appropriate.

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- No concerns.
- No concerns
- No
- Agree with prelim rec
- no
- This measure cannot be used to assess or make inferences about quality. The only common foci is the denominator age group. The three rates should be separate quality measures.
- No concerns.

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. **Feasibility** is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
    - Data elements are generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score).
    - All data elements are in defined fields in a combination of electronic sources.
    - The developer has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO´s ability to comply with HEDIS specifications.
    - In addition to the HEDIS Audit, NCQA provides a system to allow "real-time" feedback from measure users through their Policy Clarification Support System.

Questions for the Committee:

- Does the Standing Committee have any concerns with the measure's feasibility?

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. **Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?**

- High feasibility. No concerns.
- No concerns with feasibility
- Very feasible leveraging claims
- High feasibility
- none
- Measure is feasible.
- No concerns. Rating High.

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

Accountability program details

- HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2019, a total of 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
- STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2019, the report included results from calendar year 2018 for health plans covering a record 136 million people, or 43 percent of the U.S. population.
- QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years.

Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.

- HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2019, a total of 247 Medicare Advantage health plans were accredited using this measure among others. Health plans are scored based on performance compared to benchmarks.
- HEDIS ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care. ACO standards and guidelines incorporate whole-person care coordination throughout the health care system.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The developer notes that health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA.
- The developer publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.
- The developer also presents data at various conferences and webinars and regularly provides technical assistance on measures through its Policy Clarification Support System
- The developer evaluates the measure regularly, seeking broad input, including on performance and implementation experience. Methods of obtaining input include vetting with mutli-stakeholder advisory panels, public comment posting, and review of questions submitted to the developer.
- The developer reports that health plans have not reported significant barriers to implementation.
- Feedback obtained by the developer informed how they revised the measure specification to include clarifying text and additional examples to further support determining numerator compliance.

Additional Feedback: N/A

Questions for the Committee:

- How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The data demonstrates variation in all three rates of the measure for 2018.

- 48.7 percent of individuals with a history of falls received at least one high-risk medication.
- 45.5 percent of individuals with dementia received at least one high-risk medication.
- 10.2 percent of individuals with chronic kidney disease received at least one high-risk medication.
- Overall, rates from 2016 to 2018 showed relatively stable performance, yet the 2018 rates still suggest significant room for improvement in medication safety for older adults, particularly for the history of falls and dementia rates.
- For all rates there is a sizeable gap between the plans at the 10th percentile and 90th percentile, demonstrating a persistent gap in care between the best and worst performing health plans.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer did not identify any unintended consequences during testing or since implementation.

Potential harms

- The developer acknowledges the potential for reduced access to medications should the measure be implemented poorly, in addition to individual cases that warrant use of a potentially harmful medication based on the relative risk/benefit.

Additional Feedback:

- N/A

***Questions for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

Preliminary rating for Usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- Pass
- Drive patients towards safer delivery systems
- Public reporting of data is ongoing, and currently being used in an accountability program.
- No concerns
- ok
- Interpreting this data requires the extra analytic burden of separating the numerator rates.
- Currently being used and publicly reported on reports rating health plans. The health plans are informed on their ratings and have been given opportunity to provide feedback. Rating Pass.

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- The benefits of the measure outweigh risks.
- Again, to drive patients towards safer care delivery
- None
- Benefits > harms
- Readily usable due to 3-part focus on specific patient comorbidities.
- This is a difficult measure for a user to respond to for quality improvement due to the use of 3 different rates that do not share the same PIM categories. The measure needs to be broken down into three separate quality measures to make them usable. Otherwise it is a burden for providers to do their own analysis to separate them to create actionable QI. Separating them will provide the advantage of highlighting continued attention to dangerous drug-disease combinations that are otherwise buried in a composite measure. CKD measures need special attention considering the federal effort to delay progression to ESRD. Falls are an ongoing known concern that still isn't adequately studied.
- Has not led to much improvement, however still significant variation and I believe the overall benefits of the measure outweigh the harms. Rating high.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0022: The Use of High-Risk Medications in Older Adults (DAE)

Harmonization

- The measure specifications are not completely harmonized.
- The Use of High-Risk Medications in Older Adults (DAE) measure and NQF 2993 have a similar focus (measuring potentially inappropriate medication use in older adults) and reporting level (health plan), however they have different target populations.
- Measure 0022 targets a larger population of all older adults and assesses use of high-risk medications that have been recommended to be avoided in all older adults.
- Measure 2993 targets patients with a specific condition or disease who can experience adverse effects when combined with certain medications that are recommended to be avoided for that condition.
- Both measure 0022 and measure 2993 are being submitted for NQF re-endorsement during this Fall 2020 cycle, and together cover a significant portion of the AGS Beers Criteria recommendations for population-level medication safety assessment.

Committee Pre-evaluation Comments: Criterion 5:
Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- No concerns
- Complements (for the most part) 0022
- Yes, use of high risk medications. However, they measure different processes of care.
- 22
- 0022 competes but seems too non-specific to be as useful as this measure. Partial harmonization.
- No comment
- Harmonized with related measures.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: **01/15/2021**

- **No NQF Members have submitted support/non-support choices as of this date.**
- **No Public or NQF Member comments submitted as of this date.**

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2993

Measure Title: Potentially Harmful Drug-Disease Interactions in Older Adults (DDE)

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**
☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☐ **Facility** ☒ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   Submission document: "MIF_xxxx" document, items S.1-S.22

   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

RELIABILITY: TESTING

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?

   ☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

   Submission document: Testing approach was reasonable.

7. **Assess the results of reliability testing**

   Submission document: Testing demonstrated that average reliability scores exceeded the threshold of 0.7, however, there was variation in reliability among the health plans with some health plans demonstrating reliability below 0.7.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

☐ **Yes**

☐ **No**

☒ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

Submission document: Exclusions are conceptual and have not been empirically validated. This was done through face validity.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

Submission document: No concerns

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: N/A

15. **Please describe any concerns you have regarding missing data.**

Submission document: No missing data

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☒ No ☐ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☐ Yes ☒ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

16d. **Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

16d.5. Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

16e. **Assess the risk-adjustment approach**

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**

20. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

    ☒ **Empirical validity testing of the measure score**

    ☐ **N/A (score-level testing not conducted)**

21. **Assess the method(s) for establishing validity**

    Submission document: Approach is reasonable

22. **Assess the results(s) for establishing validity**

    Submission document: Test results should positive correlation with the 3 measure rates, as well as with a similar measure, which demonstrates validity.

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (score-level testing was not performed)

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

    NOTE that data element validation from the literature is acceptable.

    Submission document: *Testing attachment, section 2b1.*

    ☐ **Yes**

    ☐ **No**

    ☒ **Not applicable** (data element testing was not performed)

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.** Approach is reasonable, results demonstrate validity. There are no concerns.

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

Final_DDE_evidence_attachment_7.1.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2993

Measure Title: Potentially Harmful Drug-Disease Interactions in Older Adults

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission:

1a.1. This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome:

☐Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process: **Prescribing of potentially harmful drugs for older adults**

☐ Appropriate use measure:

☐ Structure:

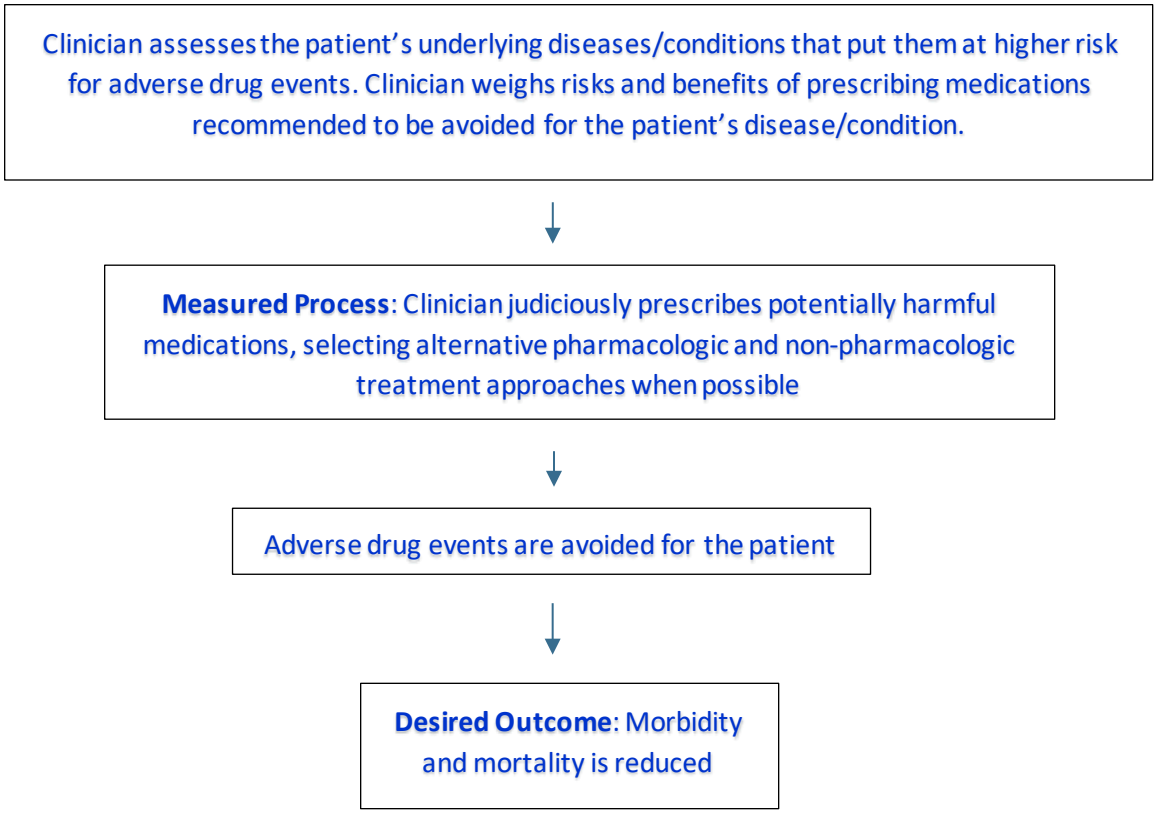
☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

2020 Submission

No changes.

2016 Submission



1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

N/A

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

N/A

1a.3. SYSTEMATIC REVIEW (SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based

on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Systematic Review

Source of Systematic Review:

- Title
- Author
- Date
- Citation, including page number
- URL

Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade

Provide all other grades and definitions from the evidence grading system

Grade assigned to the **recommendation** with definition of the grade

Provide all other grades and definitions from the recommendation grading system

Body of evidence:

- Quantity – how many studies?
- Quality – what type of studies?

Estimates of benefit and consistency across studies

What harms were identified?

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?

Evidence

2020 Submission

American Geriatrics Society 2019 Beers Criteria Update Expert Panel. 2019. American Geriatrics Society 2019 Updated AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. Journal of the American Geriatrics Society, 67(4): 674-94.

Below are the guiding principles that were developed to determine which conditions from the evidence would be included in the measure and which medications would be included in the measure.

Guiding Principles

Include conditions and medications listed in *Table 3: 2015 AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults Due to Drug-Disease or Drug-Syndrome Interactions That May Exacerbate the Disease or Syndrome*.

The following criteria were used to determine which conditions from Table 3 should be included in the measure:

1. Do not include conditions where all potentially harmful medications are already listed in *Use of High-Risk Medications in Older Adults*.
2. Do not include conditions that are rare and would not provide a sufficient denominator count for quality measurement.
3. Only include conditions where the performance rate indicates there is room for improvement (i.e., greater than minimum use of the potentially inappropriate medication).
4. Only include conditions that can be reliably identified by claims data.
5. Do not include conditions where all potentially harmful medications are primarily available over the counter.

The following criteria were used to determine which medications from Table 3 should be included in the measure:

1. Include only prescription medications.
2. Include only medications with strong recommendations to avoid.
3. When a caveat is listed in Table 3 as an appropriate use of the medication and can be identified in claims, add the medication with an exclusion for the identifiable caveat.
4. For example, anticonvulsants should be avoided except for those with seizure disorders; therefore, there is an exclusion for seizure disorders in the History of Falls rate.
5. When a caveat is listed for a medication class that cannot be identified in claims data, the medication (class) may be included in the measure if the non-identifiable caveat is considered rare.

   For example, the caveat for antipsychotics for people with dementia is that they should be avoided unless nonpharmacological options have failed and the patient is a threat to self or others. This caveat would be a rare event that cannot be identified in claims.

2016 Submission

American Geriatrics Society 2015 Beers Criteria Update Expert Panel. 2015. American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. Journal of the American Geriatrics Society, 63(11): 2227-2246. Guideline available at: http://geriatricscareonline.org/ProductAbstract/american-geriatrics-society-updated-beers-criteria-for-potentially-inappropriate-medication-use-in-older-adults/CL001

2020 Submission

Language in the table below is taken verbatim from Table 3 (pages 10-12) of the *American Geriatrics Society 2019 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults*. The following changes to the 2019 Beers Criteria were applied to the DDE measure during the most recent update:

- Added SNRIs to History of Falls rate
- Removed H2 receptor antagonists from Dementia rate
- Added an exclusion for major depressive disorder to History of Falls rate
- Added Pyrilamine to the list of anticholinergics, first-generation antihistamines
- Added Methscopolamine to the list of anticholinergics, antispasmodics

History of Falls or Fractures (page 11)

Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Antiepileptics Antipsychotics Benzodiazepines Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics - Eszopiclone - Zaleplon - Zolpidem Antidepressants - TCAs - SSRIs - SNRIs - Opioids	May cause ataxia, impaired psychomotor function, syncope, additional falls; shorter-acting benzodiazepines are not safer than long-acting ones If one of the drugs must be used, consider reducing use of other CNS-active medications that increase risk of falls and fractures (i.e., anticonvulsants, opioid receptor agonists, antipsychotics, antidepressants, benzodiazepine receptor agonists, other sedatives and hypnotics) and implement other strategies to reduce fall risk. Data for antidepressants are mixed but no compelling evidence that certain antidepressants confer less fall risk than others.	Avoid unless safer alternatives are not available Avoid antiepileptics except for seizure and mood disorders[1] Opioids: avoid, except for pain management in the setting of severe acute pain (e.g., recent fractures or joint replacement)[2]	All others: High Opioids: moderate	Strong

Dementia or cognitive impairment (page 10-11)

Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Anticholinergics (see Table 7 for full list) Benzodiazepines Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics - Eszopiclone - Zolpidem - Zaleplon Antipsychotics, chronic and as-needed use	Avoid because of adverse CNS Effects Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with greater risk of cerebrovascular accident (stroke) and mortality in persons with dementia	Avoid	Moderate	Strong

Chronic kidney disease Stages IV or higher (creatinine clearance <30 mL/min) (page 11)

Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
NSAIDs (non-COX and COX-selective, oral and Parenteral, nonacetylated salicylates)	May increase risk of acute kidney injury and further decline of renal function	Avoid	Moderate	Strong

Table 7: Drugs with Strong Anticholinergic Properties

Drugs 1:	Drugs 2:	Drugs 3:	Drugs 4:
Antihistamines (first generation) Brompheniramine Carbinoxamine Chlorpheniramine Clemastine Cyproheptadine Dexbrompheniramine Dexchlorpheniramine Dimenhydrinate Diphenhydramine (oral) Doxylamine Hydroxyzine Meclizine Clidinium-chlordiazepoxide Dicyclomine Homatropine (excludes ophthalmic) Hyoscyamine Methscopolamine Propantheline	Antiparkinsonian agents Benztropine Trihexyphenidyl	Skeletal muscle relaxants Cyclobenzaprine Orphenadrine	Antidepressants Amitriptyline Amoxapine Clomipramine Desipramine Doxepin (>6 mg) Imipramine Nortriptyline Paroxetine Protriptyline Trimipramine

Drugs 1:	Drugs 2:	Drugs 3:	Drugs 4:
Antipsychotics Chlorpromazine Clozapine Loxapine Olanzapine Perphenazine Thioridazine Trifluoperazine	Antiarrhythmic Disopyramide	Antimuscarinics (urinary incontinence) Darifenacin Fesoterodine Flavoxate Oxybutynin Solifenacin Tolterodine Trospium	Antispasmodics Atropine (excludes ophthalmic) Belladonna alkaloids Scopolamine (excludes ophthalmic)
Antiemetic Prochlorperazine Promethazine	Promethazine Pyrilamine Triprolidine		

Language in the below table is taken verbatim from Table 3 (pages 15-18) of the *American Geriatrics Society (AGS) 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults.*

History of Falls or Fractures (page 16-17)

Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Anticonvulsants Antipsychotics Benzodiazepines Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics - Eszopiclone - Zaleplon - Zolpidem - TCAs - SSRIs Opioids[1]	May cause ataxia, impaired psychomotor function, syncope, additional falls; shorter-acting benzodiazepines are not safer than long-acting ones If one of the drugs must be used, consider reducing use of other CNS-active medications that increase risk of falls and fractures (i.e., anticonvulsants, opioid receptor agonists, antipsychotics, antidepressants, benzodiazepine receptor agonists, other sedatives and hypnotics) and implement other strategies to reduce fall risk.	Avoid unless safer alternatives are not available; Avoid anticonvulsants except for seizure and mood disorders[1] Opioids: avoid, excludes pain management due to recent fractures or joint Replacement	High Opioids: moderate	Strong Opioids: strong

[1]Anticonvulsants are included in the measure because the conditions for which there is appropriate use can be reliably identified using claims data, so we can exclude those with appropriate use (see Guiding Principles under section 1a.7.1).

[2]Opioids are not included in the measure due to the caveat in the recommendation statement that opioid use for pain management due to recent fractures or joint replacement is appropriate. These uses cannot be reliably identified using claims data alone, so we cannot exclude those with appropriate use (see Guiding Principles under section 1a.7.1).

Dementia or cognitive impairment (page 15-16)

Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Anticholinergics (see Table 7 for full list) Benzodiazepines H2-receptor antagonists Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics -Eszopiclone -Zolpidem -Zaleplon Antipsychotics, chronic and as-needed use	Avoid because of adverse CNS Effects Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with greater risk of cerebrovascular accident (stroke) and mortality in persons with dementia	Avoid	Moderate	Strong

Chronic kidney disease Stages IV or less (creatinine clearance <30 mL/min) (page 18)

Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
NSAIDs (non-COX and COX-selective, oral and Parenteral)	May increase risk of acute kidney injury and further decline of renal function	Avoid	Moderate	Strong

Table 7: Drugs with Strong Anticholinergic Properties

Drugs 1:	Drugs 2:	Drugs 3:	Drugs 4:
Antihistamines Brompheniramine Carbinoxamine Chlorpheniramine Clemastine Cyproheptadine Dexbrompheniramine Dexchlorpheniramine Dimenhydrinate Diphenhydramine (oral) Doxylamine Hydroxyzine Meclizine Triprolidine	Antiparkinsonian agents Benztropine Trihexyphenidyl	Skeletal muscle relaxants Cyclobenzaprine Orphenadrine	Antidepressants Amitriptyline Amoxapine Clomipramine Desipramine Doxepin (>6 mg) Imipramine Nortriptyline Paroxetine Protriptyline Trimipramine

Drugs 1:	Drugs 2:	Drugs 3:	Drugs 4:
Antipsychotics Chlorpromazine Clozapine Loxapine Olanzapine Perphenazine Thioridazine Trifluoperazine	Antiarrhythmic Disopyramide	Antimuscarinics (urinary incontinence) Darifenacin Fesoterodine Flavoxate Oxybutynin Solifenacin Tolterodine Trospium	Antispasmodics Atropine (excludes ophthalmic) Belladonna alkaloids Clidiniumchlordiazepoxide Dicyclomine Homatropine (excludes ophthalmic) Hyoscyamine Propantheline Scopolamine (excludes ophthalmic)
Antiemetic Prochlorperazine Promethazine			

2020 Submission

The grade assigned by AGS to the quality of evidence varied by each recommendation. See the table above for the grade assigned to the evidence for each recommendation. The chart below is excerpted from the Beers Criteria article and contains the definitions for the quality of evidence ratings and the strength of recommendations.

Table 1. Designations of Quality of Evidence and Strength of Recommendations*

Quality of Evidence

Quality of evidence ratings for each criterion are based on synthetic assessment of two complementary approaches to evaluating the quality of evidence.

ACP-based approach

GRADE-based approach

High-quality evidence

"Evidence…obtained from 1 or more well-designed and well-executed randomized, controlled trials (RCTs) that yield consistent and directly applicable results. This also means that further research is very unlikely to change our confidence in the estimate of effect."

Consider the following five factors for the studies that comprise the best-available evidence for a given criterion:

1. Risk of bias: Severity of threats to studies' internal validity (e.g., randomized vs observational design, potential for confounding, bias in measurement)
2. Inconsistency: Do different studies provide similar or different estimates of effect size
3. Indirectness: How relevant are the studies to the clinical question at hand (e.g., nature of study of population, comparison group, type of outcomes measured)
4. Imprecision: Precision of estimates of effect
5. Publication bias: Risk of bias due to selective publication of results

Moderate-quality evidence

"Evidence…obtained from RCTs with important limitations…. In addition, evidence from well-designed controlled trials without randomization, well-designed cohort or case-control analytic studies, and multiple time series with or without intervention are in this category. Moderate-quality evidence also means that further research will probably have an important effect on our confidence in the estimate of effect and may change the estimate."

Low-quality evidence

"Evidence obtained from observational studies would typically be rated as low quality because of the risk for bias. Low-quality evidence means that further research is very likely to have an important effect on our confidence in the estimate of effect and will probably change the estimate. However, the quality of evidence may be rated as moderate or even high, depending on circumstances under which evidence is obtained from observational studies."

↓↓↓↓↓

Overall quality of evidence that supports a given criterion: high, moderate, low

Strength of Evidence

Strength of evidence ratings for each criterion are based on synthetic integration of the quality of evidence, the frequency and severity of potential adverse events and relationship to potential benefits, and clinical judgment.

Strong

Harms, adverse events, and risks clearly outweigh benefits.

Weak

Harms, adverse events, and risks may not outweigh benefits.

Abbreviations: ACP, American College of Physicians; GRADE, Grading of Recommendations Assessment, Development and Evaluation.

*Adapted from:

Qaseem A, Snow V, Owens DK, et al. The development of clinical practice guidelines and guidance statements of the American College of Physicians: summary of methods. Ann Intern Med. 2010;153:194–199.

Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol. 2013;66(2):151–157.

Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol. 2013;66(7):726–735.

2016 Submission

The grade assigned by AGS to the quality of evidence and the strength of the recommendations varied by each recommendation. See table under 1a.4.2 for the quality of evidence and strength of recommendation grades given to each recommendation.

Measure	Quality of Evidence
High	Evidence includes consistent results from well designed, well-conducted studies in representative populations that directly assess effects on health outcomes (≥2 consistent, higher-quality randomized controlled trials or multiple, consistent observational studies with no significant methodological flaws showing large effects)
Moderate	Evidence is sufficient to determine risks of adverse outcomes, but the number, quality, size, or consistency of included studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes (≥1 higher-quality trial with >100 participants; ≥2 higher-quality trials with some inconsistency; ≥2 consistent, lower-quality trials; or multiple, consistent observational studies with no significant methodological flaws showing at least moderate effects) limits the strength of the evidence
Measure	**Strength of Recommendation**
Strong	Benefits clearly outweigh harms, adverse events, and risks, or harms, adverse events, and risks clearly outweigh benefits

2020 Submission

N/A

2016 Submission

Measure	Quality of Evidence
Low	Evidence is insufficient to assess harms or risks in health outcomes because of limited number or power of studies, large and unexplained inconsistency between higher-quality studies, important flaws in study design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes
Measure	**Strength of Recommendation**
Weak	Benefits may not outweigh harms, adverse events, and risks
Insufficient	Evidence inadequate to determine net harms, adverse events, and risks

2020 Submission

See the table above for the grade assigned to each recommendation.

Strong: Harms, adverse events, and risks clearly outweigh benefits

2020 Submission

Weak: Harms, adverse events, and risks may not outweigh benefits

2020 Submission

Methods used for the 2019 update were similar to those used in the 2015 update of the Beers Criteria. AGS formed an expert panel to update the Beers Criteria. The panel worked from the 2015 evidence review and then reviewed any new evidence published since then to update the recommendations in the Beers Criteria. The 2019 review by the AGS 2019 Beers Criteria Update Expert Panel, which this measure is based on, included review of 67 systematic reviews and meta analyses, 29 randomized control trials (RTCs) and 281 observational studies and other types of publications. A list of studies supporting the recommendation for each drug class is not publicly available for the 2019 publication.

Overall, the quality of the evidence for each of the conditions/diseases included in quality measure is good. The dementia and chronic kidney disease section have moderate evidence and the history of falls section has high evidence. In addition to conducting a systematic review of the evidence, the AGS 2019 Beers Criteria Update Expert Panel also used technical experts and a public comment period for additional validity.

2016 Submission

The Beers Criteria were first published in 1991. Since that time the criteria have been regularly updated based off of the existing criteria and any new evidence published since the last update. The American Geriatrics Society forms an expert panel to update the Beers Criteria every few years. The panel works from the previous evidence review and then reviews any new evidence published since that last review to update the recommendations in the Beers Criteria. The 2015 review by the AGS 2015 Beers Criteria Update Expert Panel included review of 60 systematic reviews and meta analyses, 49 randomized control trials (RTCs) and 233 observational studies and other types of publications.

Overall the quality of the evidence for each of the conditions/diseases included in quality measure is good. The dementia and chronic kidney disease section have moderate evidence and the history of falls section has high evidence. In addition to conducting a systematic review of the evidence, the AGS 2015 Beers Criteria Update Expert Panel also used technical experts and a public comment period for additional validity.

History of falls: Evidence for the recommendation to avoid certain medications (anticonvulsants, antipsychotics, benzodiazepine and nonbenzodiazepine hypnotics, tricyclic antidepressants and SSRIs) for individuals with a history of falls was rated as high quality. It includes 5 systematic reviews of evidence in addition to cohort studies.

Anticonvulsants: 2 systematic review, 1 cohort

Antipsychotics: 2 systematic review; 3 cohort; 1 case-control

Benzodiazepines: 3 systematic review; 2 cohort; 3 case-control

Nonbenzodiazepine hypnotics: 2 systematic review; 1 cohort; 1 case-control

Tricyclic Antidepressants: 3 systematic review; 3 cohort; 1 case-control

SSRIs: 2 systematic review; 3 cohort; 1 case-control

Dementia or cognitive impairment: Evidence for the recommendation to avoid certain medications (anticholinergic drugs, benzodiazepines, H2-receptor antagonists, benzodiazepine and nonbenzodiazepine

hypnotics, antipsychotics) for individuals with dementia was rated as moderate quality. It includes 2 systematic reviews and 3 randomized control studies in addition to cohort studies.

Antiemetics: 1 cohort

Antipsychotics: 6 cohort

- Benzodiazepines: 5 cohort
- Nonbenzodiazepine hypnotics: 2 cohort
- Tricyclic Antidepressants: 1 systematic review; 3 cohort
- H2 Receptor Antagonists: 2 cohort

Antihistamines: 1 cohort

- Antispasmodics: 1 cohort
- Antimuscaninics (oral): 1 cohort
- Parkinson agents: 1 cohort
- Skeletal muscle relaxants: 1 cohort
- SSRIs: 1 systematic review; 2 cohort
- Antiarrhythmic: 1 cohort

Chronic kidney disease: Evidence for the recommendation to avoid NSAIDs for individuals with chronic kidney disease (stages IV or less [creatinine clearance <30 mL/min)]) was rated as moderate quality. It includes 1 randomized control study and 5 cohort studies.

- Cox-2 Selective NSAIDs: 1 randomized control study; 1 cohort
- Non-aspirin NSAIDs: 1 randomized control study; 4 cohort

2020 Submission

Recommendations in the Beers criteria are based on studies that explain the rationale for why a medication group is potentially harmful for a patient with a certain condition. Each updated study contributes to the strength of the measure by updating the medication lists. The studies consistently mention similar drugs. Since the bodies of evidence all relate to the original Beers list, they maintain consistency in process. Changes to the 2019 Beers Criteria Update improved the clarity of the recommendations and further focused the criteria on medications that are particularly problematic for older adults. Thus, the AGS Beers Criteria continue to be a useful clinical tool to improve medication safety in older adults.

2016 Submission

Recommendations in the Beers criteria are based on studies that explain the rationale for why a medication group is potentially harmful for a patient with a certain condition. Below is a summary of the number and types of studies supporting the recommendation for each drug class. Summaries of each study can be found on the American Geriatrics Society's website: http://www.americangeriatrics.org/.

Condition	Class of Drugs	Studies that support recommendation	Recommendation
Dementia or	Anticholinergics	2015 Criteria:	Avoid because of adverse CNS

Condition	Class of Drugs	Studies that support recommendation	Recommendation
cognitive impairment	(see Table 7 for full list) Benzodiazepines H2-receptor antagonists Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics Eszopiclone Zolpidem	Chavant 2011 Kalicsh Ellet 2014 From previous criteria: Boustani 2007 Hanlon2004 Finkle 2011 Frey 2011 Paterniti 2002 Rasmussen 1999 Rudolph 2008 Schneider 2005 Schneider 2006a Schneider 2006b Seitz 2011 Vigen 2011 Wright 2009	Effects Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with greater risk of cerebrovascular accident (stroke) and mortality in persons with dementia
History of falls or fractures	Anticonvulsants Antipsychotics Benzodiazepines Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics - Eszopiclone - Zaleplon - Zolpidem TCAs SSRIs	Rolita 2013 Soderberg 2013 From previous criteria: Allain 2005 Berdot 2009 Deandrea 2010 Ensrud 2003 Hartikainen 2007 Jalbert 2010 Liperoti 2007 Mets 2010 Sterke 2008 Turner 2011 van der Hooft 2008	May cause ataxia, impaired psychomotor function, syncope, additional falls; shorter-acting benzodiazepines are not safer than long-acting ones If one of the drugs must be used, consider reducing use of other CNS-active medications that increase risk of falls and fractures (i.e., anticonvulsants, opioid receptor agonists, antipsychotics, antidepressants, benzodiazepine receptor agonists, other sedatives and hypnotics) and implement other strategies to reduce fall risk

Condition	Class of Drugs	Studies that support recommendation	Recommendation
		Vestergaard 2008 Wagner 2004 Wang 2001a Wang 2001b	
Chronic Kidney disease	NSAIDs (non-COX and COX-selective, oral and parenteral)	Gooch 2007 Griffin 2000 Lafrance 2009 Murray 1995 Schneider 2006 Winkelmayer 2008	May increase risk of acute kidney injury and further decline of renal function

2020 Submission

As part of their review of the evidence, the AGS 2019 Beers Criteria Update Expert Panel identified subgroups of patients who should be exempt from the criteria and for whom listed medications may be appropriate. In addition, a patient could have a condition or comorbidity that would merit the use of a medication on the list, even if the comorbidity is not specifically listed in the criteria. The panel noted that exclusions to the criteria should not be expanded to include all adults 65 and older when only a portion of individuals may benefit from use of these medications. The criteria are designed to assist providers in the prescribing of potentially harmful medications and should not be taken as strict criteria to avoid use in all patients without weighing the harms and benefits for individual cases.

2016 Submission

As part of their review of the evidence, the AGS 2015 Beers Criteria Update Expert Panel identified subgroups of patients who should be exempt from the criteria and for whom listed medications may be appropriate. In addition, a patient could have a condition or comorbidity that would merit the use of a medication on the list, even if the comorbidity is not specifically listed in the criteria. The panel noted that exclusions to the criteria should not be expanded to include all adults 65 and older when only a portion of individuals may benefit from use of these medications. The criteria are designed to assist providers in the prescribing of potentially harmful medications, and should not be taken as strict criteria to avoid use in all patients without weighing the harms and benefits for individual cases.

2020 Submission

To our knowledge there have been no published studies since the systematic review that would impact the recommendations.

2016 Submission

To our knowledge there have been no published studies since the systematic review that would impact the recommendations.

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

Lowering the rate of potentially harmful drug-disease interactions in the older adult population should decrease morbidity and mortality associated with adverse drug reactions.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

The following data are extracted from HEDIS data collection for Medicare Advantage Health Plans (including both HMO and PPO plans). Performance data is summarized at the health plan level and summarized by mean, standard deviation, and performance at the 10th, 25th, 50th, 75th and 90th percentile. Data is stratified by year.

Rate 1 (History of Falls)^

YEAR | N | MEAN | ST DEV | MIN | 10TH (Better) | 25TH | 50TH | 75TH | 90TH (Worse) | MAX

2016 | 390 | 47.1% | 8.3 | 0.0% | 38.2% | 43.0% | 46.8% | 51.9% | 57.4% | 74.1%

2017| 411 | 48.2% | 8.3 | 0.0% | 39.1% | 43.3% | 47.6% | 52.6% | 58.7% | 75.7%

2018*| 402 | 48.7% | 7.5 | 25.8% | 39.6% | 43.8% | 48.3% | 53.6% | 58.7% | 76.6%

*For 2018 the average eligible population was 2,481, with a standard deviation of 5,368

^Note: These results are based on a previous specification of the HEDIS measure that does not include SNRI medications and the associated exclusion for major depressive disorder.

Rate 2 (Dementia)

YEAR | N | MEAN | ST DEV | MIN | 10TH (Better) | 25TH | 50TH | 75TH | 90TH (Worse) | MAX

2016 | 385 | 45.5% | 8.3 | 9.8% | 36.2% | 40.5% | 44.6% | 50.0% | 55.7% | 77.7%

2017| 408 | 46.5% | 8.5 | 19.5% | 37.3% | 40.7% | 45.2% | 51.7% | 57.2% | 75.3%

2018*| 407 | 45.5% | 8.3 | 23.6% | 36.1% | 40.0% | 44.3% | 50.5% | 56.7% | 75.0%

*For 2018 the average eligible population was 1,974, with a standard deviation of 4,073

Rate 3 (Chronic Kidney Disease)

YEAR | N | MEAN | ST DEV | MIN | 10TH (Better) | 25TH | 50TH | 75TH | 90TH (Worse) | MAX

2016 | 344 | 10.0% | 6.4 | 0.8% | 3.8% | 5.8% | 8.3% | 12.6% | 18.4% | 43.5%

2017| 369 | 10.2% | 6.5 | 0.0% | 3.5% | 5.9% | 8.6% | 12.9% | 18.8% | 42.6%

2018*| 377 | 10.2% | 6.6 | 0.0% | 3.8% | 5.9% | 8.4% | 12.7% | 18.8% | 39.4%

*For 2018 the average eligible population was 998, with a standard deviation of 2,035

The data referenced are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. In 2018, HEDIS measures covered more than 21 million Medicare enrollees. Below is a description of the denominator for this measure. It includes the number of health plans reporting the measure and the mean denominator for the measure across health plans.

Rate 1 (History of Falls)

YEAR | N Plans | Median Denominator Size per plan | Mean Denominator Size per plan

2016 | 390 | 616 | 1,919

2017 | 411 | 632 | 2,142

2018 | 402 | 719 | 2,481

Rate 2 (Dementia)

YEAR | N Plans | Median Denominator Size per plan | Mean Denominator Size per plan

2016 | 385 | 621 | 1,721

2017 | 408 | 644 | 1,830

2018 | 407 | 686 | 1,974

Rate 3 (Chronic Kidney Disease)

YEAR | N Plans | Median Denominator Size per plan | Mean Denominator Size per plan

2016 | 344 | 348 | 895

2017 | 369 | 347 | 919

2018 | 377 | 365 | 998

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Section 1b.2 references data from the most recent three years of measurement for this measure. The data in section 1b.2 includes percentiles, mean, interquartile range and standard deviation and demonstrates room for improvement.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required*

for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.

HEDIS data are stratified by type of insurance (e.g., Commercial, Medicaid, Medicare). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities if the data are available to a plan. NCQA is actively engaged with partners including the CMS Office of Minority Health in identifying feasible methods to further integrate social risk factors into health plan quality measures, with a focus on stratification. Our work is aligned with recent recommendations from MedPAC and ASPE on optimal methods for addressing social risk in quality measurement and programs.1,2 This is an NCQA wide initiative. Our intent is to implement methods to bridge data concerns in the future.

HEDIS includes two measures that can be used as tools for assessing race/ethnicity and language needs of a plan's population: Race/Ethnicity Diversity of Membership and the Language Diversity of Membership. These measures promote standardized methods for collecting these data and follow Office of Management and Budget and National Academy of Medicine guidance for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing, and using race/ethnicity and language data to assess health care disparities.

1. Medicare Payment Advisory Commission. (2020). The Medicare Advantage program: Status report. In Report to the Congress: Medicare Payment Policy (p. 397). http://medpac.gov/docs/default-source/reports/mar20_medpac_ch13_sec.pdf
2. Office of the Assistant Secretary for Planning and Evaluation, & U.S. Department of Health & Human Services. (2020). Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-basedpurchasing-programs

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

While disparities for this measure have not been well studied, there is some evidence to suggest differences in the use of potentially inappropriate medications by gender, race, and income status. A cross-sectional study examining the prevalence of potentially inappropriate medications in community-dwelling Medicare beneficiaries in California found that use was significantly higher in women, White beneficiaries, and low-income beneficiaries (Patel et al., 2018). A retrospective cohort study of 966,000 men and women treated by the Veteran's Health Administration showed that women were more likely than men to receive medications that may have harmful interactions with chronic conditions as described by the Beers Criteria (Bierman et al., 2007). In a different study, a retrospective database analysis of HEDIS data from the Department of Veterans Affairs found that Hispanics and those with no copayments had higher rates of medications listed as potentially harmful than Whites or those with required copayments (Pugh, 2011).

Bierman, A.S., M.J.V. Pugh, I. Dhalla, M. Amuan, B.G. Fincke, A. Rosen, D.R. Berlowitz. 2007. "Sex differences in inappropriate prescribing among elderly veterans." The American Journal of Geriatric Pharmacotherapy, 5(2):147-161.

Patel, R., L. Zhu, D. Sohal, E. Lenkova, N. Koshki, J. Woelfel, … and E.L. Rogan. 2018. "Use of 2015 Beers Criteria Medications by Older Medicare Beneficiaries." The Consultant Pharmacist 33(1), 48–54.

Pugh, Mary Jo V., et al. "Exposure to Potentially Harmful Drug–Disease Interactions in Older Community-Dwelling Veterans Based on the Healthcare Effectiveness Data and Information Set Quality Measure: Who Is at Risk?." Journal of the American Geriatrics Society 59.9 (2011): 1673-1678.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

De.6. Non-Condition Specific*(check all the areas that apply):*

Safety : Medication

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

Elderly, Populations at Risk, Populations at Risk : Dual eligible beneficiaries, Populations at Risk : Individuals with multiple chronic conditions

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

N/A

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment**:** 2993_DDE_Fall_2020_Value_Sets.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

Since the last endorsement, the measure name was revised to replace the term “elderly” with “older adults” to align with the language used in the American Geriatrics Society (AGS) Beers Criteria. The list of medications used in this measure has been updated to reflect the most current recommendations included in the AGS 2019 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. A new exclusion for major depressive disorder was added to the History of Falls rate with the addition of SNRIs to the measure indicator. The AGS Beers Criteria recommend avoiding SNRIs for people with a history of falls, and avoiding nearly all

antidepressants (SSRIs, tricyclics, SNRIs) is now recommended. We exclude members with a diagnosis of major depressive disorder from the rate because the benefits of using antidepressants for these individuals may outweigh the risks. We also removed Numerator 4 (total rate).

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Numerator 1: Patients with a history of falls who received at least one potentially harmful medication from Table DDE-A or Table DDE-B

Numerator 2: Patients with a diagnosis of dementia who received at least one potentially harmful medication from Table DDE-D

Numerator 3: Patients with chronic kidney disease who received at least one potentially harmful medication from Table DDE-E

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Rate 1 numerator: Dispensed an ambulatory prescription for an anticonvulsant, SSRI, or SNRI (Table DDE-A), or antipsychotic, benzodiazepine, nonbenzodiazepine hypnotic or tricyclic antidepressant (Table DDE-B) on or between the index episode start date (IESD) and December 31 of the measurement year.

Rate 2 numerator: Dispensed an ambulatory prescription for an antipsychotic, benzodiazepine, nonbenzodiazepine hypnotic or tricyclic antidepressant (Table DDE-B), or anticholinergic agent (Table DDE-D) on or between the IESD and December 31 of the measurement year.

Rate 3 numerator: Dispensed an ambulatory prescription for a Cox-2 selective NSAID or non-aspirin NSAID (Table DDE-E) on or between the IESD and December 31 of the measurement year.

Note: Do not include denied claims.

Index Episode Start Date. The earliest diagnosis, procedure or prescription between January 1 of the year prior to the measurement year and December 1 of the measurement year.

For an outpatient claim/encounter, the IESD is the date of service.

For an inpatient claim/encounter, the IESD is the discharge date.

For an acute inpatient encounter identified only by a professional claim (where the discharge date cannot be determined), the IESD is the date of service.

For dispensed prescriptions, the IESD is the dispense date.

…

Table DDE-A: Potentially Harmful Drugs – Rate 1

Anticonvulsants:

Carbamazepine, Clobazam, Divalproex sodium, Ethosuximide, Ethotoin, Ezogabine, Felbamate, Fosphenytoin, Gabapentin, Lacosamide, Lamotrigine, Levetiracetam, Methsuximide, Oxcarbazepine, Phenobarbital, Phenytoin, Pregabalin, Primidone, Rufinamide, Tiagabine HCL, Topiramate, Valproate sodium, Valproic acid, Vigabatrin, Zonisamide

SNRIs:

Desvenlafaxine, Duloxetine, Levomilnacipran, Venlafaxine

SSRIs:

Citalopram, Escitalopram, Fluoxetine, Fluvoxamine, Paroxetine, Sertraline

---

Table DDE-B: Potentially Harmful Drugs – Rate 1 (History of Falls) and Rate 2 (Dementia)

Antipsychotics:

Aripiprazole, Asenapine, Brexpiprazole, Cariprazine, Chlorpromazine, Clozapine, Fluphenazine, Haloperidol, Iloperidone, Loxapine, Lurasidone, Molindone, Olanzapine, Paliperidone, Perphenazine, Pimozide, Quetiapine, Risperidone, Thioridazine, Thiothixene, Trifluoperazine, Ziprasidone

Benzodiazepine hypnotics:

Alprazolam, Chlordiazepoxide products, Clonazepam, Clorazepate-Dipotassium, Diazepam, Estazolam, Flurazepam HCL, Lorazepam, Midazolam HCL, Oxazepam, Quazepam, Temazepam, Triazolam

Nonbenzodiazepine hypnotics:

Eszopiclone, Zaleplon, Zolpidem

Tricyclic antidepressants:

Amitriptyline, Amoxapine, Clomipramine, Desipramine, Doxepin (>6 mg), Imipramine, Nortriptyline, Protriptyline, Trimipramine

---

Table DDE-D: Potentially Harmful Drugs – Rate 2 (Dementia)

Anticholinergic agents, antiemetics:

Prochlorperazine, Promethazine

Anticholinergic agents, antihistamines:

Brompheniramine, Carbinoxamine, Chlorpheniramine, Hydroxyzine, Clemastine, Cyproheptadine, Pyrilamine, Triprolidine, Dimenhydrinate, Diphenhydramine, Meclizine, Dexbrompheniramine, Dexchlorpheniramine, Doxylamine

Anticholinergic agents, antispasmodic:

Atropine, Homatropine, Belladonna alkaloids, Dicyclomine, Hyoscyamine, Methscopolamine, Propantheline, Scopolamine, Clidinium-chlordiazepoxide

Anticholinergic agents, antimuscarinics (oral)

Darifenacin, Fesoterodine, Solifenacin, Trospium, Flavoxate, Oxybutynin, Tolterodine

Anticholinergic agents, anti-Parkinson agents

Benztropine, Trihexyphenidyl

Anticholinergic agents, skeletal muscle relaxants

Cyclobenzaprine, Orphenadrine

Anticholinergic agents, SSRIs:

Paroxetine

Anticholinergic agents, antiarrhythmic:

Disopyramide

Table DDE-E: Cox-2 Selective NSAIDs and Non-aspirin NSAIDs

Cox-2 Selective NSAIDs:

Celecoxib

Non-aspirin NSAIDs:

Diclofenac potassium, Diclofenac sodium, Etodolac, Fenoprofen, Flurbiprofen, Ibuprofen, Indomethacin, Ketoprofen, Ketorolac, Meclofenamate, Mefenamic acid, Meloxicam, Nabumetone, Naproxen, Naproxen sodium, Oxaprozin, Piroxicam, Sulindac, Tolmetin

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

All patients 65 years of age and older with a history of falls, dementia or chronic kidney disease in the measurement year or the year prior to the measurement year.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

All patients ages 67 years and older as of December 31 of the measurement year with a history of falls, dementia or chronic kidney disease. Each of the three rates in the measure has a different denominator:

Rate 1 denominator: Patients with an accidental fall or hip fracture (Note: hip fractures are used as a proxy for identifying accidental falls). Individuals with either of the following on or between January 1 of the year prior to the measurement year and December 1 of the measurement year meet criteria:

- An accidental fall (Falls Value Set).
- An acute inpatient encounter (Acute Inpatient Value Set), nonacute inpatient encounter (Nonacute Inpatient Value Set), outpatient visit (Outpatient Value Set), an observation visit (Observation Value Set) or an ED visit (ED Value Set) with a hip fracture (Hip Fractures Value Set).
- An acute or nonacute inpatient discharge with a hip fracture (Hip Fractures Value Set). To identify acute and nonacute inpatient discharges: 1) Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set). 2) Identify the discharge date for the stay. 3) Identify the index episode start date (IESD) for each patient.

Rate 2 denominator: Patients with a diagnosis of dementia (Dementia Value Set) or a dispensed dementia medication (Table DDE-C) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year. Identify the IESD for each patient.

Rate 3 denominator: Patients with chronic kidney disease as identified by a diagnosis of ESRD (ESRD Value Set), dialysis (Dialysis Procedure Value Set), stage 4 chronic kidney disease (CKD Stage 4 Value Set), nephrectomy (Nephrectomy Value Set) or kidney transplant (Kidney Transplant Value Set) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year.

Note: Patients with more than one disease or condition may appear in the measure multiple times (i.e., in each indicator for which they qualify).

Index Episode Start Date. The earliest diagnosis, procedure or prescription between January 1 of the year prior to the measurement year and December 1 of the measurement year.

For an outpatient claim/encounter, the IESD is the date of service.

For an inpatient claim/encounter, the IESD is the discharge date.

For an acute inpatient encounter identified only by a professional claim (where the discharge date cannot be determined), the IESD is the date of service.

For dispensed prescriptions, the IESD is the dispense date.

See S.2.b for all Value Sets

Table DDE-C: Prescriptions to Identify Members with Dementia

Cholinesterase inhibitors:

Donepezil, Galantamine, Rivastigmine

Miscellaneous central nervous system agents:

Memantine

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

For those who meet denominator criteria for the history of falls rate (Rate 1): exclude those with a diagnosis of psychosis, schizophrenia, schizoaffective disorder, bipolar disorder, major depressive disorder or seizure disorder.

For those who meet denominator criteria for the dementia rate (Rate 2): exclude those with a diagnosis of psychosis, schizophrenia, schizoaffective disorder or bipolar disorder.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

For those who meet denominator criteria for the history of falls rate (Rate 1): Exclude patients with a diagnosis of psychosis (Psychosis Value Set), schizophrenia, schizoaffective disorder (Schizophrenia Value Set), bipolar disorder (Bipolar Disorder Value Set; Other Bipolar Disorder Value Set), major depressive disorder (Major Depression or Dysthymia Value Set) or seizure disorder (Seizure Disorders Value Set) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year.

For those who meet denominator criteria for the dementia rate (Rate 2): Exclude patients with a diagnosis of psychosis (Psychosis Value Set), schizophrenia, schizoaffective disorder (Schizophrenia Value Set) or bipolar disorder (Bipolar Disorder Value Set; Other Bipolar Disorder Value Set) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

No risk adjustment or risk stratification

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Lower score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

Step 1. Determine the eligible population: All patients 67 years of age and older as of the end (i.e., December 31) of the measurement year.

Step 2: Identify the denominators for each of the three rates:

Rate 1: Those in the eligible population with a history of falls (see S.7 for details) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year. Exclude patients with a diagnosis of psychosis, schizophrenia, bipolar disorder, major depressive disorder, or seizure disorder (see S.9 for details). Identify the index episode start date (IESD) for each patient.

Rate 2: Those in the eligible population with dementia (see S.7 for details) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year. Exclude patients with a diagnosis of psychosis, schizophrenia, or bipolar disorder (see S.9 for details). Identify the IESD for each patient.

Rate 3: Those in the eligible population with chronic kidney disease (see S.7 for details) on or between January 1 of the year prior to the measurement year and December 1 of the measurement year. Identify the IESD for each patient.

Step 3: Identify the numerators: Individuals in each of the denominators who have received at least one potentially harmful medication on or after the IESD (see definitions of potentially harmful medications for each numerator in section S.5).

Step 4: Calculate the rates:

Rate 1 – Numerator 1 divided by denominator 1.

Rate 2 – Numerator 2 divided by denominator 2.

Rate 3 – Numerator 3 divided by denominator 3.

Note: For this measure, a lower rate indicates better performance for all three rates.

Index Episode Start Date. The earliest diagnosis, procedure or prescription between January 1 of the year prior to the measurement year and December 1 of the measurement year.

For an outpatient claim/encounter, the IESD is the date of service.

For an inpatient claim/encounter, the IESD is the discharge date.

For an acute inpatient encounter identified only by a professional claim (where the discharge date cannot be determined), the IESD is the date of service.

For dispensed prescriptions, the IESD is the dispense date.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

N/A

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Health Plan

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Outpatient Services

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

N/A

2. Validity – See attached Measure Testing Submission Form

DDE_2993_Testing_Form-637396682473177598.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 2993
Measure Title: Potentially Harmful Drug-Disease Interactions in Older Adults

Date of Submission: 11/2/2020

Type of Measure:

Measure	Measure (continued)
☐ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

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:	☐ other:

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*).

2020 Submission

N/A

2016 Submission

N/A

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

2020 Submission

HEDIS Health Plan performance data from 2018

2016 Submission

HEDIS submission data from 2014

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:	☐ other:

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*)

2020 Submission

This measure assesses the percentage of Medicare members 65 years and older who have evidence of an underlying disease, condition or health concern and who were dispensed an ambulatory prescription for a potentially harmful medication, concurrent with or after the diagnosis. There are three measure rates:

- A history of falls and a prescription for anticonvulsants, antipsychotics, benzodiazepines, nonbenzodiazepine hypnotics, or antidepressants.
- Dementia and a prescription for antipsychotics, benzodiazepines, nonbenzodiazepine hypnotics, tricyclic antidepressants, or anticholinergic agents.
- Chronic kidney disease and prescription for Cox-2 selective NSAIDs or non-aspirin NSAIDs.

Testing was completed at the health plan level which is appropriate for the level of reporting for this measure.

Measure score reliability testing and construct validity testing: Data used to assess reliability and validity were calculated from all Medicare plans submitting data to NCQA for this HEDIS measure. This data came from at least 377 Medicare plans in total that were geographically diverse and varied in size.

Systematic evaluation of face validity:
The measure was assessed for face validity with three independent panels of experts.

- The Geriatric Measurement Advisory Panel includes 15 experts in geriatric health, including representation by consumers, health plans, health care providers, and policy makers.

- The Technical Measurement Advisory Panel includes 12 members, including representation by health plan methodologists, clinicians, HEDIS auditors and state/federal users of measures.
- NCQA's Committee on Performance Measurement (CPM) oversees measures used in NCQA programs and includes representation by purchasers, consumers, health plans, health care providers, and policy makers. This panel is composed of 17 independent members that reflect the diversity of constituencies that performance measurement serves. The CPM's recommendations are reviewed and approved by NCQA's Board of Directors.

2016 Submission

Empiric reliability and validity statistics were calculated from HEDIS data that included 412 Medicare health plans. This included all Medicare health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

Face Validity: This measure was tested for face validity with two panels of experts. See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliations of expert panel members.

- The Geriatric Measurement Advisory Panel (GMAP) included 11 experts in geriatrics, including representation by consumers, health plans, health care providers and policy makers.
- NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. This panel is made up of 16 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

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*)
2020 Submission

HEDIS data are summarized at the health plan level for the Medicare product line. Below is a description of the sample. It includes number of health plans submitting measure data for HEDIS and the average and median eligible population for the measure across health plans.

Table 1. Mean and median eligible population for *Potentially Harmful Drug-Disease Interactions in Older Adults,* 2018

Rate	Number of Plans	Mean number of eligible members per plan	Median number of eligible members per plan
Rate 1 (History of Falls)	402	2481	719
Rate 2 (Dementia)	407	1974	686
Rate 3 (Chronic Kidney Disease)	377	999	365

2016 Submission

Data from the HEDIS submission for 2014 are summarized at the health plan level by numerator rate. Below is a description of the sample. It includes number of health plans included that reported this measure for HEDIS and the median eligible population for the measure across health plans.

Measure	Number of Plans	Median number of eligible patients per plan
Rate 1 (History of Falls)	387	534
Rate 2 (Dementia)	385	579
Rate 3 (Chronic Kidney Disease)	356	297
Rate 4 (Total)	412	1213

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.

2020 Submission

No differences in the data used for reliability and construct validity testing. The systematic assessment of face validity was done with multiple multi-stakeholder expert panels as described in Section 1.5 above.

2016 Submission

This question was not on the 2016 form.

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.

2020 Submission

We did not assess data by social risk factors. Social risk factor data were not available in reported results. This measure is specified for Medicare older adults, 65 years and older. NCQA is actively engaged with partners including the CMS Office of Minority Health in identifying feasible methods to further integrate social risk factors into health plan quality measures, with a focus on stratification. This is aligned with recent recommendations from MedPAC and ASPE on optimal methods for addressing social risk in quality measurement and programs.[1,2] This is an NCQA wide initiative. Our intent is to implement methods to bridge data concerns in the future.

1. Medicare Payment Advisory Commission. (2020). The Medicare Advantage program: Status report. In Report to the Congress: Medicare Payment Policy (p. 397). http://medpac.gov/docs/default-source/reports/mar20_medpac_ch13_sec.pdf
2. Office of the Assistant Secretary for Planning and Evaluation, & U.S. Department of Health & Human Services. (2020). Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-basedpurchasing-programs

2016 Submission

This question was not on the 2016 form.

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*)

2020 Submission

Reliability testing of performance measure score

We utilized the methodology described by John Adams (Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009) to calculate signal-to-noise reliability. This methodology uses the Beta-binomial model to assess how well one can confidently distinguish the performance of one reporting entity from another. Conceptually, the Beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences across reporting entities (plans, physicians, etc.) in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures, such as the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance across reporting entities.

For the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure, health plans are the reporting entity. For the formulas and explanations below, we use health plans as the reporting entity.

The formula for signal-to-noise reliability is:

$$\text{Signal-to-noise reliability} = \sigma^2_{\text{plan-to-plan}} / (\sigma^2_{\text{plan-to-plan}} + \sigma^2_{\text{error}})$$

More simply, the formula is the numerator is the variation across plans, and the denominator is the sum of the variation across plans plus the variation within the plan (across members).

Therefore, we need to estimate two variances: 1) variance between plans ($\sigma^2_{\text{plan-to-plan}}$); 2) variance within plans (σ^2_{error}).

1. Variance between plans = $\sigma^2_{\text{plan-to-plan}} = (\alpha\ \beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$

   α and β are two shape parameters of the Beta-Binomial distribution, $\alpha > 0$, $\beta > 0$

2. Variance within plans: $\sigma^2_{\text{error}} = \hat{p}(1- \hat{p})/n$

   $\hat{p}$ = observed rate for the plan

   n = plan-specific denominator for the observed rate (most often the number of eligible plan members)

Using Adams' 2009 methodology, we estimated the reliability for each reporting entity, then averaged these reliability estimates across all reporting entities to produce a point estimate of signal-to-noise reliability. We

label this point estimate "mean signal-to-noise reliability". The mean signal-to-noise reliability measures how well, on average, the measure can differentiate between reporting entity performance on the measure.

Along with the point estimate of mean signal-to-noise reliability, we are also providing:

1. The standard error (SE) and 95% confidence interval (95% CI) of the mean signal-to-noise reliability for all plans and stratified by the denominator size (number of eligible members per plan). The SE and 95% CI of the mean signal-to-noise reliability provides information about the stability of reliability. The 95% CI is the mean signal-to-noise reliability ± (1.96*SE). The narrower the confidence interval, the less the mean signal-to-noise reliability estimate will change due to idiosyncratic features of specific plans. We also stratified the results by the denominator size using terciles of the distribution to provide additional information about the stability of reliability.

2. The distribution (minimum, 10th, 25th, 50th, 75th, 90th, maximum) of the plan-level signal-to-noise reliability estimates. Each plan's reliability estimate is a ratio of signal to noise, as described above [$\sigma^2_{\text{plan-to-plan}} / (\sigma^2_{\text{plan-to-plan}} + \sigma^2_{\text{error}})$]. Variability between plans ($\sigma^2_{\text{plan-to-plan}}$) is the same for each plan, while the specific plan error (σ^2_{error}) varies. Reliability for each plan is an ordinal measure of how well one can determine where a given plan lies in the distribution of reliability across all plans, with higher estimates indicating better reliability. We also stratified the results by the denominator size using terciles of the distribution to provide additional information about the distribution of plan-level signal-to-noise reliability estimates. The number of plans in each stratum and the per-plan denominators of the performance rates are displayed in the summary tables.

This methodology allows us to estimate the reliability for each plan and summarize the distribution of these estimates.

2016 Submission

In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009). The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures. The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

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*)

2020 Submission

Table 2a. Point Estimates of Mean Signal-to-Noise Reliability, 2018

Measure	Point estimate: Mean Signal-To-Noise Reliability
Rate 1 (History of Falls)	0.872
Rate 2 (Dementia)	0.880
Rate 3 (Chronic Kidney Disease)	0.879

Table 2b. Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the *Potentially Harmful Drug-Disease Interactions in Older Adults* Measure by Terciles of the Denominator Size and for All Submissions, 2018

Stratification	Number of Plans	Number of Eligible Members per Plan (min - max)	Mean Signal-To-Noise Reliability	SE	95% CI
Rate 1 (History of Falls)	**402**	**32 - 46847**	**0.872**	**0.007**	**(0.857, 0.886)**
Tercile 1	133	32 – 335	0.704	0.011	(0.682, 0.726)
Tercile 2	132	336 – 1546	0.933	0.002	(0.928, 0.938)
Tercile 3	137	1549 – 46847	0.982	0.001	(0.981, 0.984)
Rate 2 (Dementia)	**407**	**30 - 36535**	**0.880**	**0.007**	**(0.867, 0.893)**
Tercile 1	134	30 – 255	0.616	0.012	(0.593, 0.639)
Tercile 2	134	259 – 1266	0.926	0.003	(0.920, 0.931)
Tercile 3	139	1268 - 36535	0.989	0.001	(0.988, 0.990)
Rate 3 (Chronic Kidney Disease)	**377**	**30 - 16878**	**0.879**	**0.007**	**(0.865, 0.893)**
Tercile 1	125	30 – 153	0.683	0.015	(0.682, 0.726)
Tercile 2	124	155 – 663	0.928	0.004	(0.682, 0.726)
Tercile 3	128	664 - 16878	0.982	0.001	(0.682, 0.726)

SE: Standard Error of the mean.

95% CI: 95% confidence interval.

Table 2c. Distribution of Plan-Level Signal-To-Noise Reliability for the *Potentially Harmful Drug-Disease Interactions in Older Adults* Measure by Terciles of the Denominator Size and for All Submissions, 2018

Stratification	Number of Plans	Distribution of Plan Estimates of Signal-to-Noise Reliability: Min	Distribution of Plan Estimates of Signal-to-Noise Reliability: P10	Distribution of Plan Estimates of Signal-to-Noise Reliability: P25	Distribution of Plan Estimates of Signal-to-Noise Reliability: P50	Distribution of Plan Estimates of Signal-to-Noise Reliability: P75	Distribution of Plan Estimates of Signal-to-Noise Reliability: P90	Distribution of Plan Estimates of Signal-to-Noise Reliability: Max
Rate 1 (History of Falls)	402	0.378	0.630	0.821	0.934	0.976	0.991	0.999
Tercile 1	133	0.383	0.511	0.595	0.727	0.818	0.857	0.874
Tercile 2	132	0.876	0.892	0.910	0.938	0.958	0.967	0.972
Tercile 3	137	0.962	0.967	0.973	0.985	0.991	0.995	0.999
Rate 2 (Dementia)	407	0.430	0.664	0.815	0.946	0.981	0.992	0.999
Tercile 1	134	0.317	0.396	0.511	0.651	0.726	0.776	0.807
Tercile 2	134	0.850	0.870	0.903	0.938	0.951	0.961	0.967
Tercile 3	139	0.977	0.980	0.985	0.990	0.994	0.997	0.999
Rate 3 (Chronic Kidney Disease)	377	0.329	0.686	0.824	0.937	0.980	0.992	1.000
Tercile 1	125	0.271	0.431	0.580	0.716	0.802	0.860	1.000
Tercile 2	124	0.775	0.866	0.908	0.940	0.965	0.973	0.992
Tercile 3	128	0.908	0.963	0.977	0.986	0.992	0.996	0.999

2016 Submission

Using 2014 Health Plan performance data, reliability for the rates in this measure are as shown below. Strong reliability is demonstrated since majority of variances is due to signal and not to noise

Rate	Beta Binomial Rate
Rate 1 (History of Falls)	0.96565
Rate 2 (Dementia)	0.97552
Rate 3 (Chronic Kidney Disease)	0.95273
Rate 4 (Total)	0.98571

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?*)

2020 Submission

Table 2a shows the point estimates of mean signal-to-noise reliability using above methodology.

The value of the signal-to-noise reliability estimate is greater than 0.7 for all rates, suggesting the measure has adequate reliability.

Table 2b provides the point estimate of mean signal-to-noise reliability, its standard error, and the 95% CI for the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure stratified by the denominator size (distribution of the number of eligible members per plan). For Rate 1, the reliability estimate is 0.872, and the 95% CI is (0.857, 0.886). For Rate 2, the reliability estimate is 0.880, and the 95% CI is (0.867, 0.893). For Rate 3, the reliability estimate is 0.879, and the 95% CI is (0.865, 0.893), indicating good reliability for all rates. Stratified analyses show that reliability increases as plan size gets larger.

Table 2c summarizes the distribution of plan-level signal-to-noise reliability estimates for the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure. For Rate 1, the estimates range from 0.378 to 0.999. The 50th percentile is 0.934, which exceeds the 0.70 threshold for reliability. For Rate 2, the estimates range from 0.430 to 0.999. The 50th percentile is 0.946, which exceeds the 0.70 threshold for reliability. For Rate 3, the estimates range from 0.329 to 1.000. The 50th percentile is 0.937, which exceeds the 0.70 threshold for reliability. This table also includes the distribution of plan-level signal-to-noise reliability estimates stratified by the tercile of the denominator size. In general, reliability estimates are higher for plans with a larger denominator.

2016 Submission

Reliability scores can vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (signal). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. The testing suggests that all indicators within this measure have great reliability that is above 0.95.

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)*

2020 Submission

Method of testing construct validity

We tested for construct validity by exploring the following:

- Are the individual rates within the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure correlated with one another?
- Is the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure correlated with the HEDIS *Use of High-Risk Medications in Older Adults* measure, which assesses the percentage of Medicare

members ages 65 years and older who had at least 2 dispensing events for the same high-risk medication?

We hypothesized that rates within the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure would be highly positively correlated. We also hypothesized that the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure would be positively correlated with the *Use of High-Risk Medications in Older Adults* measure. In addition, organizations that perform well on *Potentially Harmful Drug-Disease Interactions in Older Adults* should perform well on the other medication safety measure, *Use of High-Risk Medications in Older Adults*, given that they address the same older adult population.

NCQA performs Pearson correlation for construct validity using HEDIS health plan data. The test estimates the strength of linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a strong positive linear association: an increase in values of one variable is associated with increase in value of another variable. A value of 0 indicates no linear association. A value of -1 indicates a strong negative relationship in which an increase in values of the first variable is associated with a decrease in values of the second variable. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The sample size for the correlation analysis is the number of plans that reported both measures. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We adjusted our p-values to account for testing multiple correlations and used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

Systematic Assessment of Face Validity of Performance Measure Score

NCQA develops measures using a standardized process. For new measures, face validity is assessed at various steps as described below.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical measurement advisory panels (MAPs), whose members are authorities on clinical priorities for measurement, participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness, and feasibility. Measures are aligned with clinical guidelines whenever possible; the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure is aligned with the American Geriatrics Society (AGS) Beers Criteria, which recommends drugs to be avoided in older adults, particularly those with certain diseases or conditions. This information is gathered into a work-up format, which is vetted by the MAPs, including the Geriatric Measurement Advisory Panel (GMAP), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM), as well as other panels as necessary.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. At this step, face validity is systematically determined by the CPM, which uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA about proposed new measures. Public comment offers an opportunity to assess the validity, feasibility,

importance and other attributes of a measure from a wider audience. For this measure, a majority of public comment respondents supported the measure. NCQA MAPs and the technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. Face validity is then again systematically assessed by the CPM. The CPM reviews all comments before making a final decision and votes to recommend approval of new measures for HEDIS. NCQA's Board of Directors then approves new measures.

2016 Submission

Method of Assessing Face Validity: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identify areas of interest or gaps in care with clinical expert panel input. Once topics are identified, a literature review and stakeholder interviews are conducted to evaluate the importance, scientific soundness and feasibility of potential measure concepts. This information is reviewed by NCQA's Measurement Advisory Panels (MAPs), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

STEP 2: MAPs participate in the development and testing of measures by advising on measure specification, testing plans and testing results demonstrating reliability, validity and feasibility of potential measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures. MAPs consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQAs Board of Directors will be included in the next HEDIS year and reported as first-year measures.

STEP 4: All new measures are collected, but results are not publicly reported in the first year. This period guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. NCQA conducts a detailed evaluation of all first-year data. The CPM uses these evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

STEP 5: If the measure is approved by the CPM, it will be publicly reported and may be used for scoring in accreditation.

Step 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Each year, NCQA prioritizes measures for re-evaluation based on changes in clinical guidelines and feedback from measure users, auditors or other stakeholders. If necessary, the measure is re-evaluated for importance, scientific soundness, and feasibility with input from the MAPs. Specifications may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume.

Method of Testing Construct Validity: We empirically tested for construct validity by exploring whether the indicators within this measure were correlated with each other and with another measure of medication safety. We hypothesized that organizations that perform well on one of the three indicators should perform well on the other indicators as well as the other medication safety measure. To test these correlations we used a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear

dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

Method for ICD-10 Conversion: Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

Steps in ICD-9 to ICD-10 Conversion Process

1. NCQA staff identify ICD-10 codes to be considered based on ICD-9 codes currently in measure. Use GEM to identify ICD-10 codes that map to ICD-9 codes. Review GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA staff identify additional codes (not identified by GEM mapping step) that should be considered. Using ICD-10 tabular list and ICD-10 Index, search by diagnosis or procedure name for appropriate codes.
3. NCQA HEDIS Expert Coding Panel review NCQA staff recommendations and provide feedback.
4. As needed, NCQA Measurement Advisory Panels perform clinical review. Due to increased specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is intended to be included in the scope of the measure. Not all ICD-10 recommendations are reviewed by NCQA MAP; MAP review items are identified during staff conversion or by HEDIS Expert Coding Panel.
5. Post ICD-10 code recommendations for public review and comment.
6. Reconcile public comments. Obtain additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
7. NCQA staff finalize ICD-10 code recommendations.

Tools Used to Identify/Map to ICD-10
All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).
GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation

NCQA's Geriatric Measurement Advisory Panel and Committee for Performance Measurement reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
2020 Submission

Table 3. Health-Plan Level Pearson Correlation Coefficients Among *Potentially Harmful Drug-Disease Interactions in Older Adults* and *Use of High-Risk Medications in Older Adults* Performance Scores, 2018

Measure	Correlation Coefficient: DDE Rate 1 (History of Falls)	Correlation Coefficient: DDE Rate 2 (Dementia)	Correlation Coefficient: DDE Rate 3 (Chronic Kidney Disease)
DDE Rate 1 (History of Falls)	--	--	--
DDE Rate 2 (Dementia)	0.63	--	--

Measure	Correlation Coefficient: DDE Rate 1 (History of Falls)	Correlation Coefficient: DDE Rate 2 (Dementia)	Correlation Coefficient: DDE Rate 3 (Chronic Kidney Disease)
DDE Rate 3 (Chronic Kidney Disease)	0.24	0.59	--
*Use of High-Risk Medications in Older Adults**	0.61	0.53	0.24

Note: All correlations are significant at p<0.001

*The *Use of High-Risk Medications in Older Adults* measure assesses the percentage of patients 65 and older who received at least two of the same high-risk medications.

-- cell intentionally left blank

Results of face validity assessment

NCQA worked closely with our multi-stakeholder MAPs to re-evaluate the measure based on the latest recommendations in the American Geriatric Society's 2019 Beers Criteria. The last Beers Criteria update prior to this publication was in 2015. Based on the 2019 Beers Criteria, the primary changes to the measure were updates to medications. The measure changes were evaluated in 2019. After reviewing, the CPM recommended to send the updated measure to public comment with a majority vote in 2019. The measure was released for Public Comment in 2019 prior to publication in HEDIS. Input from advisory panels and the public comment indicate the measure has face validity.

2016 Submission

Results of Face Validity Assessment:

Step 1: This measure was developed to address potentially harmful drug-disease interactions in the elderly. NCQA and the GMAP worked together to assess conditions and medications based on the AGS Beers Criteria.

Step 2: The measure was field-tested from 2004-2005. After reviewing field test results the CPM recommended to send the measure to public comment with a majority vote in 2006.

Step 3: The measure was released for Public Comment in 2006 prior to publication in HEDIS. The CPM recommended moving this measure to first year data collection by a majority vote.

Step 4: The measure was introduced in HEDIS 2007. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure public reporting with a majority vote.

Step 5: The measure is currently undergoing re-evaluation.

Conclusion: The measure was deemed to have the desirable attributes of a HEDIS measure in 2006 (relevance, scientific soundness, and feasibility).

Results of Construct Validity Testing: The results in Table 1a indicate that there was a moderate or high correlation between all rates with the exception of Rate 1 (History of Falls) and Rate 3 (Chronic Kidney Disease).

Table 1a. Correlations among all rates in DDE measure and the DAE measure[1]

Measure	**Pearson Correlation Coefficients:** DDE: Rate 1 (History of Falls)	**Pearson Correlation Coefficients:** DDE: Rate 2 (Dementia)	**Pearson Correlation Coefficients:** DDE: Rate 3 (Chronic Kidney Disease)	**Pearson Correlation Coefficients:** DDE: Rate 4 (Total)	DAE: High-Risk Med, 65+
DDE Rate 1 (History of Falls)	--	--	--	--	--
DDE: Rate 2 (Dementia)	0.694	--	--	--	--
DDE: Rate 3 (Chronic Kidney Disease)	0.155	0.585	--	--	--
DDE: Rate 4 (Total)	0.842	0.921	0.480	--	--
DAE: High-Risk Medication in those 65 and older	0.307	0.454	0.367	0.386	--

Note: All correlations are significant at $p<.05$

[1]The DAE measure assesses the percentage of patients 65 and older who receive at least one high-risk medication (as defined by Table 2 of the American Geriatrics Society Beers Criteria. There is no disease/condition requirement for this measure.

-- cell intentionally left blank

ICD-10 Conversion:

Summary of Stakeholder Comments Received

NCQA posted ICD-10 codes for public review and comment in March 2011 and March 2012. NCQA received comments from four organizations:

- Support recommendations.
- Questions about select codes.
- Recommended additional codes for consideration.

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?*)

2020 Submission

For the purposes of this analysis and the intended use of this measure to evaluate the quality of care for members across health plans, correlation was considered high (strong) if the correlation coefficient is 0.75 to 1, moderate if 0.25 to 0.75, and low (weak) if 0 to 0.25.

The individual rates within the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure are positively correlated with each other. Correlations between the *Potentially Harmful Drug-Disease Interactions in Older Adults* and the *Use of High-Risk Medications in Older Adults* measure were moderate. This suggests

that plans that perform well on the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure are moderately likely to perform well on the *Use of High-Risk Medications in Older Adults* measure. The results indicate that the measure has good validity.

2016 Submission

Interpretation of systematic assessment of face validity: These results indicate the MAPs and CPM showed agreement that the measures as specified will accurately differentiate quality across health plans. Our interpretation of these results is that this measure has sufficient face validity.

Interpretation of construct validity testing: Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. The results confirmed the hypothesis that rates in the measure are correlated with each other as well as with another measure of medication safety, suggesting they represent the same underlying quality construct of prescribing inappropriate medications for patients with the corresponding illnesses. These results indicate the measure is a valid measure of a plan's quality at managing potentially harmful drug-disease interactions.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b3*

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*)

2020 Submission

The exclusions for this measure are based on the conditions specified in the individual rates that could merit the use of a medication that is listed in the Beers Criteria. We did not perform empirical testing of the diagnosis codes used to identify the exclusions. NCQA engaged expert panels to inform the face validity of this measure's exclusions, which aligns with the recommendations in the AGS Beers Criteria. This measure has been reviewed by NCQA's Geriatric Measurement Advisory Panel and the Committee on Performance Measurement. The measure also received public comment feedback upon initial development and during recent updates to the measure. HEDIS data can be used to identify the overall rate and distribution of exclusions across health plans that report on HEDIS measures.

2016 Submission

The exclusions for this measure are based on the conditions specified in the individual rates that could merit the use of a medication that is listed in the Beers Criteria. While the diagnosis codes used to identify the exclusions have not been tested in the context of this measure for validity, they are widely used across practitioners and considered to be valid. HEDIS data can be used to identify the overall rate and distribution of exclusions across health plans that report HEDIS.

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*)

2020 Submission

Table 4. Variation in Exclusions, 2018

Measure	N	Mean rate of exclusions	SD	Min	P10	P25	P50	P75	P90	Max
Rate 1 (History of Falls) [1]	402	20.6	10.4	0.0	9.7	13.2	18.4	25.1	35.2	67.0
Rate 2 (Dementia) [2]	407	21.0	10.3	0.0	10.3	14.1	18.5	25.5	36.0	67.0

N = Number of plans reporting

[1]For the History of Falls rate, those with a diagnosis of psychosis, schizophrenia, schizoaffective disorder, bipolar disorder, major depressive disorder or seizure disorder are excluded.

[2]For the Dementia rate, those with a diagnosis of psychosis, schizophrenia, schizoaffective disorder or bipolar disorder are excluded.

2016 Submission

The following table includes the rate and distribution of exclusions across health plans reporting 2014 HEDIS data.

Measure	Number of plans	Average rate of exclusions	Standard deviation	Min	25th	50th	75th	Max
Rate 1: History of Falls[1]	384	15.3	11.5	0.0	0.0	16.8	20.9	56.0
Rate 2: Dementia[2]	382	18.9	12.8	0.0	9.1	20.6	28.4	50.9
Rate 4: Total	409	14.8	10.7	0.0	6.9	16.0	20.67	61.0

[1]For the History of Falls rate, those with a diagnosis of psychosis, schizophrenia, bipolar disorder or seizure disorder are excluded.

[2]For the Dementia rate, those with a diagnosis of psychosis, schizophrenia or bipolar disorder are excluded.

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*)

2020 Submission

The exclusions in this measure are identified using administrative claims codes and therefore do not add much burden to collection. The results indicate that on average 20.6 percent of patients meet the exclusion criteria for the History of Falls rate and 21.0 percent of patients meet the exclusion criteria for the Dementia rate.

2016 Submission

The exclusions in this measure are identified using administrative claims codes and therefore do not add much burden to collection. The results indicate that on average 15.3 percent of patients meet the exclusion criteria for the History of Falls rate and 18.9 percent meet the exclusion criteria for the Dementia 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.

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

☒ **No risk adjustment or stratification**

☒ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other**

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.

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.

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?

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)**

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

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.**

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*)

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

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

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

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

2b3.9. Results of Risk Stratification Analysis:

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*)

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*)

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)*

2020 Submission

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure.

To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the below 25th and above 75th percentile groups. The t-test method calculates a testing statistic based on the sample size, performance rate, and standard error of each plan. The test statistic is then compared against a t distribution, which is similar to a normal distribution. If the p-value of the test statistic is less than .05, then the two plans' performance is significantly different from each other.

2016 Submission

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared

against a normal distribution. If the p value of the test statistic is less than .05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the 75th percentile of performance. We used data from the most recent HEDIS submissions in 2014.

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*)

2020 Submission

Table 5. Variation in Performance, 2018

Measure	N	Mean eligible population	Mean rate (%)	SD	Min	P10	P25	P50	P75	P90	Max	IQR	p-value
Rate 1 (History of Falls)	402	2481	48.7	7.5	25.8	39.6	43.8	48.3	53.6	58.7	76.6	9.8	p <0.001
Rate 2 (Dementia)	407	1974	45.5	8.3	23.6	36.1	40.0	44.3	50.5	56.7	75.0	10.5	p <0.001
Rate 3 (Chronic Kidney Disease)	377	999	10.2	6.6	0.0	3.8	5.9	8.4	12.7	18.8	39.4	6.8	p <0.001

N = Number of plans reporting

IQR = Interquartile range

p-value = p-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile

2016 Submission

Variation in Performance across Plans (HEDIS Results using 2014 Data)

Rate	Number of Plans	Mean	SD	10th (Better)	25th	50th	75th	90th (Worse)	IQR	p-Value
Rate 1 (History of Falls)	387	48.0%	8.3	38.8%	43.1%	47.7%	51.9%	58.5%	8.8	0.0026
Rate 2 (Dementia)	385	48.5%	9.1	39.2%	42.8%	46.8%	52.8%	61.0%	10	0
Rate 3 (Chronic Kidney Disease)	356	9.6%	6.1	3.9%	5.8%	8.1%	12.0%	17.1%	6.2	0
Rate 4 (Total)	412	41.5%	7.9	33.5%	36.7%	40.3%	44.9%	51.3%	8.2	<0.001

IQR: Interquartile range

p-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

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?*)

2020 Submission

The results indicate there is a 6.8-10.5 percentage point gap in performance between the 25th and 75th performing plans.

For Rate 1, there is a 9.8 percentage point gap in performance between Medicare plans at the 25th and 75th percentiles. This gap represents on average 248 more older adults with a history of falls receiving potentially harmful drugs in low-performing plans versus high-performing plans. For Rate 2, there is a 10.5 percentage point gap in performance between Medicare plans at the 25th and 75th percentiles. This gap represents on average 197 more older adults with a diagnosis of dementia receiving potentially harmful drugs in low-performing plans versus high-performing plans. For Rate 3, there is a 6.8 percentage point gap in performance between Medicare plans at the 25th and 75th percentiles. This gap represents on average 70 more older adults with a diagnosis of chronic kidney disease receiving potentially harmful drugs in low-performing plans versus high-performing plans compared to high performing plans. The difference in performance between plans in the 25th percentile and 75th percentile is statistically significant for all three rates.

2016 Submission

The results above indicate there is a 6-10% gap in performance between the 25th and 75th performing plans. Plans at the 25th and 75th percentile have a statistically significant difference in performance. The largest gap in performance is for the dementia rate which had a 10% gap in performance between plans at the 25th and 75th percentiles.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

2020 Submission

This measure has only one set of specifications.

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*)

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*)

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*)

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 non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

2020 Submission

HEDIS measures apply to enrolled members in a health plan, and NCQA has a rigorous audit process to ensure the eligible population and numerator events for each measure are correctly identified and reported. The audit process is designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.

The HEDIS Compliance Audit addresses the following functions:

- Information practices and control procedures
- Sampling methods and procedures
- Data integrity
- Compliance with HEDIS specifications
- Analytic file production
- Reporting and documentation

2016 Submission

Plans collect this measure using all administrative data sources. NCQA's audit process checks that plans' measure calculations are not biased due to missing data.

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*)

2020 Submission

HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used. Once measures are added to HEDIS, NCQA conducts a first-year analysis to assess the feasibility of the measure when widely implemented in the field. This analysis includes an assessment of how many plans report valid rates vs. rates that are materially biased (or have other issues, such as small

denominators). These considerations are weighed in the deliberation process before measures are approved for public reporting.

2016 Submission

Plans collect this measure using all administrative data sources. NCQA's audit process checks that plans' measure calculations are not biased due to missing data.

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 non-responders) 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*)

2020 Submission

All health plans that reported 2018 HEDIS data for this measure reported valid rates as determined by NCQA-certified auditors. This means that auditors did not find any missing data sources for any of the health plan data submissions and determined that none of the rates were materially biased.

2016 Submission

Plans collect this measure using all administrative data sources. NCQA's audit process checks that plans' measure calculations are not biased

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in a combination of electronic sources

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than

electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

N/A

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

NCQA recognizes that, despite the clear specifications defined for HEDIS measures, data collection and calculation methods may vary, and other errors may taint the results, diminishing the usefulness of HEDIS data for managed care organization (MCO) comparison. In order for HEDIS to reach its full potential, NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

1) information practices and control procedures
2) sampling methods and procedures
3) data integrity
4) compliance with HEDIS specifications
5) analytic file production
6) reporting and documentation

In addition to the HEDIS Audit, NCQA provides a system to allow “real-time” feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system is vital to the regular re-evaluation of NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
*	Public Reporting Health Plan Rating http://www.ncqa.org/ReportCards/HealthPlans/HealthInsurancePlanRankings/HealthPlanRatingsPreview.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx Health Plan Rating http://www.ncqa.org/ReportCards/HealthPlans/HealthInsurancePlanRankings/HealthPlanRatingsPreview.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx Regulatory and Accreditation Programs HEDIS®-Health Plan http://www.ncqa.org/Programs/Accreditation/HealthPlanHP.aspx HEDIS®-ACO http://www.ncqa.org/Programs/Accreditation/AccountableCareOrganizationACO.aspx HEDIS®-Health Plan http://www.ncqa.org/Programs/Accreditation/HealthPlanHP.aspx HEDIS®-ACO http://www.ncqa.org/Programs/Accreditation/AccountableCareOrganizationACO.aspx Quality Improvement (external benchmarking to organizations) NCQA Quality Compass https://www.ncqa.org/programs/data-and-information-technology/data-purchase-and-licensing/quality-compass/

*cell intentionally left blank

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

HEALTH PLAN RATINGS/REPORT CARDS: This measure is used to calculate health plan ratings which are reported on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2019, a total of 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.

STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2019, the report included results from calendar year 2018 for health plans covering a record 136 million people, or 43 percent of the U.S. population.

QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.

HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2019, a total of 247 Medicare Advantage health plans were accredited using this measure among others. Health plans are scored based on performance compared to benchmarks.
HEDIS ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care. ACO standards and guidelines incorporate whole-person care coordination throughout the health care system.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

N/A

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

Health plans that report HEDIS measures calculate their rates and know their performance when submitting to NCQA. NCQA publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Update and Best Practices Conference, NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c1.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

NCQA measures are evaluated regularly. During this "reevaluation" process, we seek broad input on the measure, including input on performance and implementation experience. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

4a2.2.2. Summarize the feedback obtained from those being measured.

In general, health plans have not reported significant barriers to implementing this measure, as it uses the administrative data collection method. Questions received through the Policy Clarification Support system have generally centered around minor clarification of the specifications, such as confirmation that information in claims meets the measure intent and satisfies the measure numerator as well as questions about the index episode start date (IESD). NCQA responded to all questions to ensure consistent implementation of the

specifications. During a recent public comment period, a majority of comments from measured entities supported updates to the measure to align with the latest clinical recommendations.

4a2.2.3. Summarize the feedback obtained from other users

This measure has been deemed a priority measure by NCQA, as illustrated by its use in programs such as Health Plan Rating, NCQA Accreditation and Quality Compass. States, employers, and regional health quality organizations value this measure (and other HEDIS measures) for shining a light on quality.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

During the measure's last major update, feedback obtained through the mechanisms described in 4a2.2.1 informed how we revised the measure specification to include clarifying text to further support determining numerator compliance.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

The data demonstrate variation in all three rates of the measure (see section 1b.2 for summary of data from health plans). For 2018, 48.7 percent of individuals with a history of falls received at least one high-risk medication. Among individuals with dementia, 45.5 percent received at least one high-risk medication and among those with chronic kidney disease, 10.2 percent received at least one high-risk medication. Overall, rates from 2016 to 2018 showed relatively stable performance, yet the 2018 rates still suggest significant room for improvement in medication safety for older adults, particularly for the history of falls and dementia rates. For all rates there is a sizeable gap between the plans at the 10th percentile and 90th percentile, demonstrating a persistent gap in care between the best and worst performing health plans.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

There were no identified unintended consequences for this measure during testing or since implementation. If this measure were to be implemented poorly, there is concern that it could lead to reduced access to medications. There will always be individual cases that will warrant the use of a potentially harmful medication. For example, antidepressants are listed as potentially harmful to patients at risk for falls, however, clinicians should weigh the relative risk of increased falls against the potential benefit of the use of antidepressants for those with severe depression.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

There were no identified unexpected findings during testing or since implementation of this measure.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

0022: Use of High-Risk Medications in Older Adults (DAE)

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

No

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

The Use of High-Risk Medications in Older Adults (DAE) measure and NQF 2993 have a similar focus (measuring potentially inappropriate medication use in older adults) and reporting level (health plan), however they have different target populations. The DAE measure targets a larger population of all older adults and assesses use of high-risk medications that have been recommended to be avoided in all older adults. This measure (NQF 2993) targets patients with a specific condition or disease who can experience adverse effects when combined with certain medications that are recommended to be avoided for that condition. The DAE measure (NQF 0022) is being submitted for NQF re-endorsement during this current Patient Safety project as well. Together these measures cover a significant portion of the AGS Beers Criteria recommendations for population-level medication safety assessment.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): National Committee for Quality Assurance

Co.2 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Co.3 Measure Developer if different from Measure Steward: National Committee for Quality Assurance

Co.4 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

Geriatric Measurement Advisory Panel (GMAP):

Wade Aubry, University of California, San Francisco

Arlene Bierman, Agency for Healthcare Research and Quality (AHRQ)

Patricia Bomba, Excellus BlueCross BlueShield

Nicole Brandt, University of Maryland, School of Pharmacy

Jennie Chin Hansen, Geriatric Expert

Joyce Dubow, Consumer Representative

Pete Hollmann, Brown Medicine

Jeffrey Kelman, Department of Health and Human Services

Karen Nichols, Trinity-Health PACE

Steven Phillips, Geriatric Specialty Care

Erwin Tan, American Association of Retired Persons (AARP)

Eric G Tangalos, Mayo Clinic

Dirk Wales, Axial Healthcare

Joan Weiss, Health Resources and Services Administration

Neil Wenger, University of California, Los Angeles

Committee on Performance Measurement (CPM):

Andrew Baskin, MD, CVS Health/Aetna

Elizabeth Drye, MD, SM, Yale School of Medicine

Andrea Gelzer, MD, MS, FACP, AmeriHealth Caritas

Kate Goodrich, MD, MHS, Centers for Medicare & Medicaid Services

David Grossman, MD, MPH, Washington Permanente Medical Group

Christine Hunter, (Co-Chair), MD, Independent Board Director

David Kelley, MD, MPA, Pennsylvania Department of Human Services

Jeff Kelman, MMSc, MD, Department of Health and Human Services

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, Independent Consultant

Adrienne Mims, MD, MPH, AGSF, FAAFP, Alliant Health Solutions

Amanda Parsons, MD, MBA, MetroPlus

Wayne Rawlins, MD, MBA, ConnectiCare

Misty Roberts, MSN, RN, CPHQ, PMP, Humana

Rodolfo Saenz, MD, MMM, FACOG, Riverside Medical Clinic

Marcus Thygeson, (Co-Chair), MD, MPH, Bind Benefits

JoAnn Volk, MA, Georgetown University

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2007

Ad.3 Month and Year of most recent revision: 07, 2019

Ad.4 What is your frequency for review/update of this measure? Approximately every 3 years, sooner if the clinical guidelines have changed significantly.

Ad.5 When is the next scheduled review/update for this measure? 12, 2021

Ad.6 Copyright statement: ©2020 by the National Committee for Quality Assurance

1100 13th Street, NW, Third Floor

Washington, DC 20005

Ad.7 Disclaimers: These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA. © 2020 by the National Committee for Quality Assurance