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 #: 0022

Corresponding Measures:

De.2. Measure Title: Use of High-Risk Medications in Older Adults (DAE)

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 received at least two dispensing events for the same high-risk medication. A lower rate represents better performance.

1b.1. Developer Rationale: Lowering the use of high-risk medications in the older adult population should decrease morbidity and mortality associated with adverse drug reactions.

S.4. Numerator Statement: Patients who received at least two dispensing events for the same high-risk medication during the measurement year.

S.6. Denominator Statement: All patients 65 years of age and older.

S.8. Denominator Exclusions: Patients who were enrolled in hospice care at any time during the measurement year.

De.1. Measure Type: Process

S.17. Data Source: Claims, Electronic Health Data, Electronic Health Records

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **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 2016

- The evidence provided in the previous review was based on specific recommendations in the American Geriatrics Society (AGS) Beers Criteria identifying which drugs are potentially inappropriate for all older adults.

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:

- The developer provided a logic model linking older adults at risk of adverse drug events to clinicians prescribing potentially harmful medications, selecting alternative pharmacologic and non-pharmacologic treatment approaches when possible thus avoiding adverse drug events, which leads to reduction in morbidity and mortality.
- The developer also cited 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 medications would be included in the measure were also provided.
- The developer noted that 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 first rate (former Numerator 1) for members who received at least one dispensing event for a high-risk medication was retired. The remaining rate is a better assessment of the riskier, more long-term use of high-risk medications among older adults.
- 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.

Exception to evidence

N/A

Questions for the Committee:

If the developer provided updated evidence for this measure:

- The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?
- Is the evidence directly applicable to the process of care being measured?

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3)àQQC present (Box 4)à Quantity: high; Quality: high; Consistency: high (Box 5a)à 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 summarized data extracted from HEDIS data collection for Medicare Advantage Health Plans (including all HMO and PPO plans) from 2016 to 2018.
- Performance data is summarized at the health plan level and further by mean, standard deviation, and the 10th, 25th, 50th, 75th and 90th percentile.
- The average performance increased from 9.1% in 2016 to 9.6% in 2018 with an average eligible population of 25,642 and 28,463 respectively.

Disparities

- The developer cites a cross-sectional study examining the prevalence of potentially inappropriate medications in community-dwelling Medicare beneficiaries in California, which 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 was cited, showing 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).

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.

- Agree with structure and the changes with certain medications from 2016 to 2020.
- High- this maintenance measure with updated evidence applying directly to the measure.

- Measure developers used Beers criteria to support the list of high risk medications included in the report. The measure was updated based on the most recent iteration. No new data has emerged to meaningfully change these definitions in a way that would affect validity.
- High evidence
- Evidence supports that the list of drugs are potentially harmful to older patients.
- Problems with numerator and denominator vs. the measure title
- The evidence (Beer's Criteria 2019) is a systematic review directly related to the measure. The measure is trying to measure the process outcome of whether the high risk medications that the Beer's Criteria recommend against are being prescribed long term (2+ prescriptions in one year). I am not aware of additional new information. 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?

- HEDIS measures included. Pop subgroups included as well as noting disparities in women and low-income individuals.
- High level of certainty that the criterion is met.
- Yes, measure developers note a threefold relative difference in best vs worst tertiles in their testing for 2+ high-risk medications and a 1.5 fold difference in 1+, representing a clinically significant numbers of patients at risk for harms.
- Agree with high prelim rating.
- Yes, performance gap information was provided. Most of the cited studies were older. One was published in 2018 that cited 2 specific categories of drugs (pain and insomnia drugs) used by a subset of population. My overall critique of this measure is the high number of mediations and categories included - more useful measures focus on one category and link to outcomes.
- not readily available
- As of 2018, out of 502 health plans, the % of older adults with at least 2 prescriptions for high risk medication is 9.6%, was 9.9% in 2017 and 9.1 % in 2016 (n=485) at the time. The 10th percentile is 5.8% and 90th percentile is 14.9% with about 9% difference which demonstrates variation and opportunity for improvement. There was no data on subgroups or demonstrated disparities. Cited literature suggests that women, white and low-income beneficiaries may be more likely to use inappropriate medications but it's not clear if the outcome in the literature is necessarily comparable to the measure captured here. 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 care 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:

Reliability

- Empirical reliability testing was conducted at the performance measure score level utilizing the Beta-Binomial model to calculate signal-to-noise reliability.
- Using the Beta-Binomial methodology, an estimate of the reliability for each reporting entity (health plan) was calculated then averaged across all reporting entities to produce a point estimate of signal-to-noise reliability labeled "mean signal-to-noise reliability". This mean signal-to-noise reliability measures how well, on average, the measure can differentiate between reporting entity performance on the measure.
- Additionally, 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) was provided along with the distribution (minimum, 10th, 25th, 50th, 75th, 90th, maximum) of the plan-level signal-to-noise reliability estimates.
- The following tables highlight the results of the reliability testing:

Table 2a. Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the *Use of High-Risk Medications 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
Use of High-Risk Medications in Older Adults	**502**	**32 - 679844**	**0.936**	**0.006**	**(0.924, 0.947)**

Stratification	Number of Plans	Number of Eligible Members per Plan (min - max)	Mean Signal-To-Noise Reliability	SE	95% CI
Tercile 1	166	32 - 2456	0.857	0.012	(0.833, 0.881)
Tercile 2	165	2469 - 15564	0.986	0.001	(0.985, 0.988)
Tercile 3	171	15856 - 679844	0.997	0.000	(0.997, 0.997)

SE: Standard Error of the mean.

95% CI: 95% confidence interval.

Table 2b. Distribution of Plan-Level Signal-To-Noise Reliability for the *Use of High-Risk Medications 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
Use of High-Risk Medications in Older Adults	502	0.193	0.798	0.950	0.988	0.998	0.999	1.000
Tercile 1	166	0.249	0.615	0.812	0.928	0.962	0.980	1.000
Tercile 2	165	0.962	0.975	0.982	0.988	0.993	0.995	0.997
Tercile 3	171	0.987	0.994	0.996	0.998	0.999	0.999	1.000

- The reliability estimate is 0.936, and the 95% CI is (0.924, 0.947), indicating very good reliability for the measure. Stratified analyses show that reliability increases as plan size gets larger. (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.)
- The distribution of plan-level signal-to-noise reliability estimates range from 0.193 to 1.000. The 50th percentile is 0.988, which exceeds the 0.70 threshold for reliability.

Validity

- The developer tested the measure for construct validity by exploring whether the *Use of High-Risk Medications in Older Adults* measure correlated with the HEDIS *Potentially Harmful Drug-Disease Interactions in Older Adults* measure.
- The developer hypothesized that there would be a correlation. Furthermore, they hypothesized that organizations that perform well on Use of High-Risk Medications in Older Adults should perform well on the other medication safety measure, Potentially Harmful Drug-Disease Interactions in Older Adults, given that they address the same older adult population.

- The correlations were assessed using a Pearson correlation test.
- The developer explains that a Pearson correlation test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1.
    - A value of 0 indicates no linear association.
    - 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 -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.
- Additionally, the developer convened multi-stakeholder measurement advisory panels (MAPs) to assess the measure's face validity.
- The following table illustrates the results of the Pearson correlation test:

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

Measure	**Correlation Coefficient:** *Use of High-Risk Medications in Older Adults*
Use of High-Risk Medications in Older Adults	--
Drug-disease interaction: History of Falls*	0.62
Drug-disease interaction: Dementia*	0.53
Drug-disease interaction: Chronic Kidney Disease*	0.24

Note: All correlations are significant at $p<0.001$

*The *Potentially Harmful Drug-Disease Interactions in Older Adults* measure has three rates. The first rate assesses the percentage of patients 65 and older with a history of falls who received a high-risk medication. The second rate assesses the percentage of patients 65 and older with dementia who received a high-risk medication. The third rate assesses the percentage of patients 65 and older with chronic kidney disease who received a high-risk medication.

-- cell intentionally left blank

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The staff is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

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**
- **Low level of concern regarding reliability using methods appropriate for this pass/fail measure.**
- **None**
- **No concerns with reliability**
- **I cannot find information as to how "sliding scale insulin" is identified in claims/administrative data.**
- **OK**
- **Reliability data demonstrates good reliability with reliability estimate of 0.936. No concerns. Rating High.**

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

- **No**
- **No**
- **For small health plans, there seems to be an elevated risk for bias.**
- **No**
- **Only with regard to sliding scale insulin identification.**
- **Marginally reliable considering the intent vs the method of calculation.**
- **No concerns.**

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

- **No**
- **No**

- **Minor concern that validity is so much lower for chronic kidney disease patients and represented a drop from 2016.**
- **No**
- **No**
- **Yes. The measure may capture proper prescribing if there is a rationale for using the same Beer's List drug twice in one year.**
-

- **In section S.5 Numerator Details, Table DAE-B included "high risk medications with days supply criteria" – I did not see the day supply criteria from the Beers list or systematic review. I'd like clarification on why the additional criteria of days supplied was added for these medications. No other concerns. Validity rating Moderate.**

-
- 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?

- **No concerns**
- **N/A**
- **Exclusion of hospice participants seems appropriate. It is unclear if other clinical subpopulations have higher rates of multiple high-risk prescriptions that may be clinically indicated.**
- **I am satisfied with validity analyses.**
- **Not an outcome measure**
- **ok**
- **No risk adjustment which is appropriate. Exclusion includes those patients on hospice, which 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**
- **Moderate validity identified, and potential threats disclosed. May need further discussion specific to missing data.**
- **No**
- **I am satisfied with validity analyses.**
- **Comparability is problematic in this measure as the overall performance may be 10% but the user cannot identify fluctuations in differing drug category usage. This is the underlying problem with this broad measure.**
- **Approach needs revision.**
- **No concerns, the presented statistical analysis and approach to missing data both appear rigorous and appropriate.**

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:

- Is the data collection strategy ready to be put into operational use?

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?

- **No concerns in change from 2016**
- **High feasibility. No concerns.**
- **Feasible as it leverages claims data at the plan level.**
- **N/a**
- **Measure is feasible.**
- **none**
- **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 ratings.
- 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.
- NCQA PATIENT-CENTERED MEDICAL HOME (PCMH): This measure is used in the Patient Centered Medical Home Recognition program, which identifies medical practices that have invested in a model of care that puts patients at the forefront and where continuous quality improvement is a priority.
- CMS QUALITY PAYMENT PROGRAM: This measure is used in the Quality Payment Program (QPP) which is a reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs).

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
- 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 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

- In 2018, the average performance was 9.6%. There was a 9-percentage point difference between plans at the 10th and 90th percentiles.
- This large difference in performance represents a persistent gap in care and room for improvement in medication safety for older adults, particularly given the substantially large average denominator size

of all plans reporting on this measure and therefore the great number of older adults at risk for adverse drug events.

- Although overall rates aren't changing, there have been an increase in the number of plans reporting from 2016 (n=485) to 2018 (n=502).

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

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

Potential harms

- 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 and clinicians should weigh the risks and benefits of using these medications for their individual patients.

Additional Feedback:

- N/A

Questions for the Committee:

- How can the performance results be used to further the goal of high-quality, efficient healthcare?

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?

- Used in health plan report cards
- High feasibility and transparency. Pass
- It appears plans have been made aware of data and confirmed accuracy of rates. No concerns were noted that would change measure.
- Publicly reported use measure
- This is a difficult measure for a user to respond to for quality improvement due to the broad range of PIM drugs and categories. It is only useful if it is broken down into each PIM category which creates multiple additional burdensome analytics.
- marginal
- 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.

- No concerns
- The benefits of the measure outweigh risks. Low risk for unintended consequences.
- There is a potential for bias in small plans or plans that may have outsized proportions of medically complicated populations for whom high risk medications are indicated and appropriate. However, for the vast majority of older adults, this measure has value in keeping all stakeholders accountable for reducing high risk medication use to the maximally extent possible and thus benefits outweigh risks.
- High usability
- This is a difficult measure for a user to respond to for quality improvement due to the broad range of PIM drugs and categories. It is only useful if it is broken down into each PIM category which creates multiple additional burdensome analytics. It would be better to isolate PIM categories and link directly to outcomes for measures. The harm in this measure is a lot of wasted time sifting through data and presenting all of these PIM categories as if they are of equal harm - which is not true.
- Only with a way to deal with evidence-based prescribing that might be included as improper.
- Has not led to much improvement in the rates over the years, however no identified unintended consequences and overall the benefits of the measure outweigh the harms. Rating high.

Criterion 5: Related and Competing Measures

Related or competing measures

- 2993 : Potentially Harmful Drug-Disease Interactions in Older Adults (DDE)

Harmonization

- Measure 2993 and 0022 have a similar focus (measuring potentially inappropriate medication use in older adults) and reporting level (health plan), however they have different target populations.
- The DDE measure targets patients with a specific condition or disease that can experience adverse effects when combined with certain medications that are recommended to be avoided for that condition.
- 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.
- 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?

- **Measure 2993**
- **Addressed by different target populations. Sufficient.**
- **No**
- **2993: Potentially Harmful Drug-Disease Interactions in Older Adults (DDE)**
- **No comment**

- **yes**
- **There is related measure on condition-drug interaction. I think the two are complementary and not redundant, no additional steps needed to harmonize.**

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: 0022

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

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.**

   None identified

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 attachment, section 2a2.2

   - Completed measure testing using claims data. Measure Specified to Measure Performance of and tested at the Health Plan, Integrated Delivery System.
   - The data came from Medicare plans submitting data to NCQA in 2018 for this HEDIS measure.
   - A total of 502 Medicare plans that were geographically diverse and varied in size.
   - Mean Signal-to-noise reliability was calculated using the Beta-binomial model to estimate variance between plans and within plans.
   - The developer provided 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 distribution (minimum, 10th, 25th, 50th, 75th, 90th, maximum) of the plan-level signal-to-noise reliability estimates was also provided.

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

   Submission document: Testing attachment, section 2a2.3

   - Based on the mean Signal-To-Noise reliability estimate of 0.936, with a 95% CI (0.924, 0.947), the measure is considered reliable.
   - Table 2a. Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the *Use of High-Risk Medications 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
Use of High-Risk Medications in Older Adults	**502**	**32 - 679844**	**0.936**	**0.006**	**(0.924, 0.947)**
Tercile 1	166	32 - 2456	0.857	0.012	(0.833, 0.881)
Tercile 2	165	2469 - 15564	0.986	0.001	(0.985, 0.988)

Stratification	Number of Plans	Number of Eligible Members per Plan (min - max)	Mean Signal-To-Noise Reliability	SE	95% CI
Tercile 3	171	15856 - 679844	0.997	0.000	(0.997, 0.997)

SE: Standard Error of the mean.

95% CI: 95% confidence interval.

- Table 2b. Distribution of Plan-Level Signal-To-Noise Reliability for the *Use of High-Risk Medications 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
Use of High-Risk Medications in Older Adults	502	0.193	0.798	0.950	0.988	0.998	0.999	1.000
Tercile 1	166	0.249	0.615	0.812	0.928	0.962	0.980	1.000
Tercile 2	165	0.962	0.975	0.982	0.988	0.993	0.995	0.997
Tercile 3	171	0.987	0.994	0.996	0.998	0.999	0.999	1.000

- The 50th percentile of the reported distribution of plan-level signal-to-noise reliability estimate is 0.988, which exceeds the 0.70 threshold for reliability.
- The developer noted that low minimum reliability estimate (0.193) and low observed reliability in the first tercile is likely explained by a handful of very small plans (small denominators) who inflate the sigma-squared error in the signal-to-noise calculation.

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.**

    Box 1: Measure specifications precise, unambiguous, and complete → Box 2: Empirical testing conducted using statistical tests → Box 4: Reliability testing conducted with computed performance measure scores → Box 5: Method described and appropriate for assessing the proportion of variability due to real differences among measured entities → Box 6a -HIGH

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    N/A

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

    Submission document: Testing attachment, section 2b4.

    N/A

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.
    N/A

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

    Submission document: Testing attachment, section 2b6.

    None identified

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**

VALIDITY: TESTING

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

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

☒ Face validity

☒ Empirical validity testing of the measure score

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

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

- The developer tested the measure for construct validity by exploring whether the *Use of High-Risk Medications in Older Adults* measure correlated with the HEDIS *Potentially Harmful Drug-Disease Interactions in Older Adults* measure.
- The developer hypothesized that there would be a correlation. Furthermore, they hypothesized that organizations that perform well on Use of High-Risk Medications in Older Adults should perform well on the other medication safety measure, Potentially Harmful Drug-Disease Interactions in Older Adults, given that they address the same older adult population.
- The correlations were assessed using a Pearson correlation test.
- The developer explains that a Pearson correlation test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1.
  - A value of 0 indicates no linear association.
  - 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 -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.
- Additionally, the developer convened multi-stakeholder measurement advisory panels (MAPs) to assess the measure's face validity.

Submission document: Testing attachment, section 2b2.2

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

- The developer provides the following table reporting the results of the Pearson correlation:

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

Measure	**Correlation Coefficient:** *Use of High-Risk Medications in Older Adults*
Use of High-Risk Medications in Older Adults	- -
Drug-disease interaction: History of Falls*	0.62

Measure	Correlation Coefficient: *Use of High-Risk Medications in Older Adults*
Drug-disease interaction: Dementia*	0.53
Drug-disease interaction: Chronic Kidney Disease*	0.24

Note: All correlations are significant at p<0.001

*The Potentially Harmful Drug-Disease Interactions in Older Adults measure has three rates. The first rate assesses the percentage of patients 65 and older with a history of falls who received a high-risk medication. The second rate assesses the percentage of patients 65 and older with dementia who received a high-risk medication. The third rate assesses the percentage of patients 65 and older with chronic kidney disease who received a high-risk medication.

- - cell intentionally left blank

- 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 developer reported that correlations between the *Use of High-Risk Medications in Older Adults* and the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure rates were moderate, indicating that plans that perform well on the *Use of High-Risk Medications in Older Adults* measure are moderately likely to perform well on the *Potentially Harmful Drug-Disease Interactions in Older Adults* measure.
- The developer notes that input from advisory panels and the public comment indicate the measure has face validity.

Submission document: Testing attachment, section 2b2.3

21. **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)

22. **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)

23. **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.)

24. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

Box 1: Potential threats to validity assessed Box 2: Empirical validity testing conducted using the measure as specified à Box 5: Testing conducted at the measure score level à Box 6: Testing method described and appropriate à Box 7b: Moderate certainty or confidence that measure scores are reliable à MODERATE

ADDITIONAL RECOMMENDATIONS

25. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

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

DAE_nqf_evidence_attachment_7.1-637393060756503336.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*)**:** 0022

Measure Title: Use of High-Risk Medications 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: 11/2/2020

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.

Older adults at risk of adverse drug events >> clinician judiciously prescribes potentially harmful medications, selecting alternative pharmacologic and non-pharmacologic treatment approaches when possible >> adverse drug events are avoided >> morbidity and mortality is reduced

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

Evidence

Source of Systematic Review:

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

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 medications would be included in the measure.

Guiding Principles

1. Include only medications listed in Table 2: 2019 AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults.
2. Include only prescription medications.
3. Include only medications where the AGS Recommendation indicates "avoid" and that can be identified reliably from prescription drug claims data.
4. Include only medications where the AGS Recommendation or Rationale includes caveats ("except in") that can be identified reliably from administrative claims data.
5. Do not include medications that are rarely prescribed and would not provide a sufficient denominator count for quality measurement.
6. If including a medication in the measure would likely result in the increased use of another potentially harmful medication that is not included in the measure, an exception to these guiding principles may be warranted to reduce this unintended consequence.

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.

Provide all other grades and definitions from the evidence grading system

N/A

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

2020 Submission

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

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

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

Provide all other grades and definitions from the recommendation grading system

N/A

Body of evidence:

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

2020 Submission

Methods used for the 2019 update were similar to those used in the 2015 update of the Beers Criteria. The American Geriatrics Society 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.

Overall, the quality of the evidence for each of the medications included in the Beers Criteria recommendations is good. See table above for the quality of evidence rating for the recommendation for each medication or medication class. The table also includes the AGS 2019 Beers Criteria Update Expert Panel rating for the strength of the evidence supporting each recommendation.

2016 Submission

The Beers Criteria was 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 AGS 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, which this measure is based on, 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 medications included in the Beers Criteria recommendations is good. See table under section 1c.16 for the quality of evidence rating for the recommendation for each medication or medication class. The table also includes the AGS 2015 Beers Criteria Update Expert Panel rating for the strength of the evidence supporting each recommendation. Definitions of these ratings are listed in section 1c.21.

Estimates of benefit and consistency across studies

2020 Submission

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

Each updated study contributes to the strength of the measure by updating the medication lists. See section 1c.16 for a table that contains the Beers Criteria recommendations for each drug and drug class that are included in the measure. Evidence tables containing summaries of each study supporting the recommendations can be found on the American Geriatrics Society's website: http://www.americangeriatrics.org/.

The studies consistently mention similar drugs. Since the bodies of evidence all relate to the original Beers list, they maintain consistency in process. See section 1c.16 for a table that contains the Beers Criteria recommendations for each drug and drug class that are included in the measure.

What harms were identified?

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. However, the criteria are unable to account for the complexity of patients and subpopulations; there may be a small portion of individuals who will 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.

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

2020 Submission

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

2020 Submission

Language in the table below is taken verbatim from Table 2 (pages 5-9) 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 DAE measure during the most recent update:

- Added *Pyrilamine* to the list of anticholinergics, first-generation antihistamines
- Added *Methscopolamine* to the list of antispasmodics

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Anticholinergics: First-generation antihistamines (p. 5) Brompheniramine Carbinoxamine Chlorpheniramine Clemastine Cyproheptadine Dexbrompheniramine Dexchlorpheniramine Dimenhydrinate Diphenhydramine (oral) Doxylamine Hydroxyzine Meclizine Promethazine Pyrilamine Triprolidine	Highly anticholinergic; clearance reduced with advanced age, and tolerance develops when used as hypnotic; risk of confusion, dry mouth, constipation, and other anticholinergic effects or toxicity Use of diphenhydramine in situations such as acute treatment of severe allergic reaction may be appropriate	Avoid	Moderate	Strong
Antiparkinsonian agents (p. 5) Benztropine (oral) Trihexyphenidyl	Not recommended for prevention or treatment of extrapyramidal symptoms with antipsychotics; more-effective agents available for treatment of Parkinson disease	Avoid	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Antispasmodics (p. 5) Atropine (excludes ophthalmic) Belladonna alkaloids Clidinium-Chlordiazepoxide Dicyclomine Homatropine (excludes ophthalmic) Hyoscyamine Methscopolamine Propantheline Scopolamine	Highly anticholinergic, uncertain effectiveness	Avoid	Moderate	Strong
Antithrombotics (p. 5) Dipyridamole, oral short-acting (does not apply to the extended release combination with aspirin)	May cause orthostatic hypotension; more effective alternatives available; intravenous form acceptable for use in cardiac stress testing	Avoid	Moderate	Strong
Anti-infective (p. 5) Nitrofurantoin	Potential for pulmonary toxicity, hepatoxicity, and peripheral neuropathy, especially with long-term use; safer alternatives available	Avoid in individuals with creatinine clearance <30 mL/min or for long-term suppression of bacteria	Low	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Cardiovascular (p.5) Peripheral alpha-1 blockers for treatment of hypertension Doxazosin Prazosin Terazosin	High risk of orthostatic hypotension and associated harms, especially in older adults; not recommended as routine treatment for hypertension; alternative agents have superior risk/benefit profile	Avoid use as an antihypertensive	Moderate	Strong
Cardiovascular (p. 6) Central alpha-agonists Clonidine for first-line treatment of hypertension Other CNS alpha-agonists Guanabenz Guanfacine Methyldopa Reserpine (>0.1 mg/d)	High risk of adverse CNS effects; may cause bradycardia and orthostatic hypotension; not recommended as routine treatment for hypertension	Avoid as first-line antihypertensive Avoid other CNS alpha-agonists as listed	Low	Strong
Cardiovascular (p. 6) Disopyramide	May induce heart failure in older adults because of potent negative inotropic action; strongly anticholinergic; other antiarrhythmic drugs preferred	Avoid	Low	Strong
Cardiovascular (p. 6) Dronodarone	Worse outcomes have been reported in patients taking dronedarone who have permanent atrial fibrillation or severe or recently decompensated heart failure.	Avoid in individuals with permanent atrial fibrillation or severe or recently decompensated heart failure	High	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Cardiovascular (p. 6) Digoxin for first-line treatment of atrial fibrillation or of heart failure	Use in atrial fibrillation: should not be used as a first-line agent in atrial fibrillation, because there are safer and more effective alternatives for rate control supported by high-quality evidence. Use in heart failure: evidence for benefits and harms of digoxin is conflicting and of lower quality; most but not all of the evidence concerns use in HFrEF. There is strong evidence for other agents as first-line therapy to reduce hospitalizations and mortality in adults with HFrEF. in heart failure, higher dosages not associated with additional benefit and may increase risk of toxicity Decreased renal clearance of digoxin may lead to increased risk of toxic effects; further dose reduction may be necessary in patients with Stage 4 or 5 chronic kidney disease	Avoid this rate control agent as first-line therapy for atrial fibrillation Avoid as first-line therapy for heart failure If used for atrial fibrillation or heart failure, avoid dosages >0.125 mg/day	Atrial fibrillation: Low Heart failure: Low Dosage >0.125 mg/d: Moderate	Atrial fibrillation: Strong Heart failure: Strong Dosage >0.125 mg/d: Strong
Cardiovascular (p. 6) Nifedipine, immediate release	Potential for hypotension; risk of precipitating myocardial ischemia	Avoid	High	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Cardiovascular (p. 6) Amiodarone	Effective for maintaining sinus rhythm but has greater toxicities than other antiarrhythmics used in atrial fibrillation; may be reasonable first-line therapy in patients with concomitant heart failure or substantial left ventricular hypertrophy if rhythm control is preferred overrate control	Avoid as first-line therapy for atrial fibrillation unless patient has heart failure or substantial left ventricular hypertrophy	High	Strong
Central Nervous System (p. 6) Antidepressants, alone or in combination Amitriptyline Amoxapine Clomipramine Desipramine Doxepin >6 mg/d Imipramine Nortriptyline Paroxetine Protriptyline Trimipramine	Highly anticholinergic, sedating, and cause orthostatic hypotension; safety profile of low dose doxepin (≤6 mg/d) comparable with that of placebo	Avoid	High	Strong
Central Nervous System (p. 7) Antipsychotics, first (conventional) and second (atypical) generation	Increased risk of cerebrovascular accident (stroke) and greater rate of cognitive decline and mortality in persons with dementia Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological	Avoid, except in schizophrenia or bipolar disorder, or for short-term use as antiemetic during chemotherapy	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
	options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others			
Central Nervous System (p. 7) Barbiturates Amobarbital Butobarbital Butalbital Mephobarbital Pentobarbital Phenobarbital Secobarbital	High rate of physical dependence, tolerance to sleep benefits, greater risk of overdose at low dosages	Avoid	High	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Central Nervous System (p. 7) Benzodiazepines Short and immediate acting: Alprazolam Estazolam Lorazepam Oxazepam Temazepam Triazolam Long acting: Chlordiazepoxide (alone or in combination with amitriptyline or clidinium) Clonazepam Clorazepate Diazepam Flurazepam Quazepam	Older adults have increased sensitivity to benzodiazepines and decreased metabolism of long-acting agents; in general, all benzodiazepines increase risk of cognitive impairment, delirium, falls, fractures, and motor vehicle crashes in older adults May be appropriate for seizure disorders, rapid eye movement sleep behavior disorder, benzodiazepine withdrawal, ethanol withdrawal, severe generalized anxiety disorder, and periprocedural anesthesia	Avoid	Moderate	Strong
Central Nervous System (p. 7) Meprobamate	High rate of physical dependence; very sedating	Avoid	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Central Nervous System (p. 7) Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics (i.e., "Z-drugs") Eszopiclone Zolpidem Zaleplon	Nonbenzodiazepine benzodiazepine receptor agonist hypnotics (i.e., Z drugs) have adverse events similar to those of benzodiazepines in older adults (e.g., delirium, falls, fractures); increased emergency department visits and hospitalizations; motor vehicle crashes; minimal improvement in sleep latency and duration	Avoid	Moderate	Strong
Central Nervous System (p. 7) Ergoloid mesylates (dehydrogenated ergot alkaloids) Isoxsuprine	Lack of efficacy	Avoid	High	Strong
Endocrine (p. 8) Androgens Methyltestosterone Testosterone	Potential for cardiac problems; contraindicated in men with prostate cancer	Avoid unless indicated for confirmed hypogonadism with clinical symptoms	Moderate	Weak
Endocrine (p. 8) Desiccated thyroid	Concerns about cardiac effects; safer alternatives available	Avoid	Low	Strong
Endocrine (p. 8) Estrogens with or without progestins	Evidence of carcinogenic potential (breast and endometrium); lack of cardioprotective effect and cognitive protection in older women. Evidence indicates that vaginal estrogens for the treatment of vaginal	Avoid systemic estrogen (e.g., oral and topical patch) Vaginal cream or tablets: acceptable to use low-dose intravaginal estrogen for	Oral and patch: High Vaginal cream or tablets: Moderate	Oral and patch: Strong Topical vaginal cream or tablets: Weak

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
	dryness are safe and effective; women with a history of breast cancer who do not respond to nonhormonal therapies are advised to discuss the risk and benefits of low-dose vaginal estrogen (dosages of estradiol <25 lg twice weekly) with their healthcare provider	management of dyspareunia, recurrent lower urinary tract infections, and other vaginal symptoms		
Endocrine (p. 8) Growth hormone	Impact on body composition is small and associated with edema, arthralgia, carpal tunnel syndrome, gynecomastia, impaired fasting glucose	Avoid, except for patients rigorously diagnosed by evidence-based criteria with growth hormone deficiency due to an established etiology	High	Strong
Endocrine (p. 8) Insulin, sliding scale (insulin regimens containing only short- or rapid-acting insulin dosed according to current blood glucose levels without concurrent use of basal or long-acting insulin)	Higher risk of hypoglycemia without improvement in hyperglycemia management regardless of care setting. Avoid insulin regimens that include only short- or rapid-acting insulin dosed according to current blood glucose levels without concurrent use of basal or long-acting insulin. This recommendation does not apply to regimens that contain basal insulin or long-acting insulin.	Avoid	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Endocrine (p. 8) Megestrol	Minimal effect on weight; increases risk of thrombotic events and possibly death in older adults	Avoid	Moderate	Strong
Endocrine (p. 8) Sulfonylureas, long-duration Chlorpropamide Glimepiride Glyburide (also known as glibenclamide)	Chlorpropamide: prolonged half-life in older adults; can cause prolonged hypoglycemia; causes syndrome of inappropriate antidiuretic hormone secretion Glyburide: higher risk of severe prolonged hypoglycemia in older adults	Avoid	High	Strong
Gastrointestinal (p. 8) Metoclopramide	Can cause extrapyramidal effects, including tardive dyskinesia; risk may be greater in frail older adults and with prolonged exposure	Avoid, unless for gastroparesis with duration of use not to exceed 12 weeks except in rare cases	Moderate	Strong
Gastrointestinal (p. 8) Mineral oil, given orally	Potential for aspiration and adverse effects; safer alternatives available	Avoid	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Gastrointestinal (p. 8) Proton-pump inhibitors	Risk of *Clostridium difficile* infection and bone loss and fractures	Avoid scheduled use for >8 weeks unless for high-risk patients (e.g., oral corticosteroids or chronic NSAID use), erosive esophagitis, Barrett esophagitis, pathological hypersecretory condition, or demonstrated need for maintenance treatment (e.g., because of failure of drug discontinuation trial or H2-receptorantagonists)	High	Strong
Pain medications (p. 9) Meperidine	Not effective oral analgesic in dosages commonly used; may have higher risk of neurotoxicity, including delirium, than other opioids; safer alternatives available	Avoid	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Pain medications (p. 9) Non-cyclooxygenase-selective NSAIDs, oral: Aspirin >325 mg/day Diclofenac Diflunisal Etodolac Feneprofen Ibuprofen Ketoprofen Meclofenamate Mefenamic acid Meloxicam Nabumetone Naproxen Oxaprozin Piroxicam Sulindac Tolmetin	Increased risk of gastrointestinal bleeding or peptic ulcer disease in high-risk groups, including those >75 years or taking oral or parenteral corticosteroids, anticoagulants, or antiplatelet agents; use of proton-pump inhibitor or misoprostol reduces but does not eliminate risk. Upper gastrointestinal ulcers, gross bleeding, or perforation caused by NSAIDs occur in ~1% of patients treated for 3-6 months and in ~2%-4% of patients treated for 1 year; these trends continue with longer duration of use. Also can increase blood pressure and induce kidney injury. Risks are dose related.	Avoid chronic use, unless other alternatives are not effective and patient can take gastroprotective agent (proton-pump inhibitor or misoprostol)	Moderate	Strong
Pain medications (p. 9) Indomethacin Ketorolac, includes parenteral	Indomethacin is more likely than other NSAIDs to have adverse CNS effects. Of all the NSAIDs, indomethacin has the most adverse effects. Increased risk of gastrointestinal bleeding, peptic ulcer disease, and acute kidney injury in older adults	Avoid	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Pain medications (p. 9) Skeletal muscle relaxants Carisoprodol Chlorzoxazone Cyclobenzaprine Metaxalone Methocarbamol Orphenadrine	Most muscle relaxants poorly tolerated by older adults because some have anticholinergic adverse effects, sedation, increased risk of fractures; effectiveness at dosages tolerated by older adults questionable	Avoid	Moderate	Strong
Genitourinary (p. 9) Desmopressin	High risk of hyponatremia; safer alternative treatments	Avoid for treatment of nocturia or nocturnal polyuria	Moderate	Strong

- Removed *Ticlopidine* from the list of antithrombotics
- Added *Glimepiride* to the list of endocrine system, sulfonylureas, long-duration
- Removed *Pentazocine* from the list of pain medications, other

2016 Submission

Language in the table below is taken verbatim from Table 2 (pages 5-10) of the *American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults*. Evidence tables containing summaries of each study supporting the recommendations can be found on the American Geriatrics Society's website: http://www.americangeriatrics.org/.

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Anticholinergics: First-generation	Highly anticholinergic; clearance reduced with	Avoid	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
antihistamines (p. 5) Brompheniramine Carbinoxamine Chlorpheniramine Clemastine Cyproheptadine Dexbrompheniramine Dexchlorpheniramine Dimenhydrinate Diphenhydramine (oral) Doxylamine Hydroxyzine Meclizine Promethazine Triprolidine	advanced age, and tolerance develops when used as hypnotic; risk of confusion, dry mouth, constipation, and other anticholinergic effects or toxicity Use of diphenhydramine in situations such as acute treatment of severe allergic reaction may be appropriate			
Antiparkinsonian agents (p. 5) Benztropine (oral) Trihexyphenidyl	Not recommended for prevention of extrapyramidal symptoms with antipsychotics; more-effective agents available for treatment of Parkinson disease	Avoid	Moderate	Strong
Antispasmodics (p. 5) Atropine (excludes ophthalmic) Belladonna alkaloids Clidinium-Chlordiazepoxide Dicyclomine Hyoscyamine	Highly anticholinergic, uncertain effectiveness	Avoid	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Propantheline Scopolamine				
Antithrombotics (p. 5) Dipyridamole, oral short-acting (does not apply to the extended release combination with aspirin)	May cause orthostatic hypotension; more effective alternatives available; intravenous form acceptable for use in cardiac stress testing	Avoid	Moderate	Strong
Antithrombotics (p. 5) Ticlopidine	Safer, effective alternatives available	Avoid	Moderate	Strong
Anti-infective (p. 5) Nitrofurantoin	Potential for pulmonary toxicity, hepatoxicity, and peripheral neuropathy, especially with long-term use; safer alternatives available	Avoid in individuals with creatinine clearance <30 mL/min or for long-term suppression of bacteria	Low	Strong
Central alpha blockers (p. 6) Guanabenz Guanfacine Methyldopa Reserpine (>0.1 mg/d)	High risk of adverse CNS effects; may cause bradycardia and orthostatic hypotension; not recommended as routine treatment for hypertension	Avoid	Low	Strong
Central alpha blockers (p. 6) Disopyramide	Disopyramide is a potent negative inotrope and therefore may induce heart failure in older adults; strongly anticholinergic; other antiarrhythmic drugs preferred	Avoid	Low	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Central alpha blockers (p. 6) Digoxin	Use in atrial fibrillation: should not be used as a first-line agent in atrial fibrillation, because more effective alternatives exist and it may be associated with increased mortality Use in heart failure: questionable effects on risk of hospitalization and may be associated with increased mortality in older adults with heart failure; in heart failure, higher dosages not associated with additional benefit and may increase risk of toxicity Decreased renal clearance of digoxin may lead to increased risk of toxic effects; further dose reduction may be necessary in patients with Stage 4 or 5 chronic kidney disease	If used for atrial fibrillation or heart failure, avoid dosages >0.125 mg/d	Dosage >0.125 mg/d: Moderate	Dosage >0.125 mg/d: Strong
Central alpha blockers (p. 6) Nifedipine, immediate release	Potential for hypotension; risk of precipitating myocardial ischemia	Avoid	High	Strong
Central Nervous System (p. 7) Antidepressants, alone or in combination Amitriptyline	Highly anticholinergic, sedating, and cause orthostatic hypotension; safety profile of low dose doxepin (≤6 mg/d)	Avoid	High	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Amoxapine Clomipramine Desipramine Doxepin >6 mg/d Imipramine Nortriptyline Paroxetine Protriptyline Trimipramine	comparable with that of placebo			
Central Nervous System (p. 7) Barbiturates Amobarbital Butabarbital Butalbital Mephobarbital Pentobarbital Phenobarbital Secobarbital	High rate of physical dependence, tolerance to sleep benefits, greater risk of overdose at low dosages	Avoid	High	Strong
Central Nervous System (p. 8) Meprobamate	High rate of physical dependence; very sedating	Avoid	Moderate	Strong
Central Nervous System (p. 8) Nonbenzodiazepine, benzodiazepine receptor agonist hypnotics Eszopiclone Zolpidem Zaleplon	Benzodiazepine-receptor agonists have adverse events similar to those of benzodiazepines in older adults (e.g., delirium, falls, fractures); increased emergency department visits and hospitalizations; motor vehicle crashes; minimal improvement in sleep latency and duration	Avoid	Moderate	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Central Nervous System (p. 8) Ergoloid mesylates (dehydrogenated ergot alkaloids) Isoxsuprine	Lack of efficacy	Avoid	High	Strong
Endocrine (p. 8) Desiccated thyroid	Concerns about cardiac effects; safer alternatives available	Avoid	Low	Strong
Endocrine (p. 8) Estrogens with or without progestins	Evidence of carcinogenic potential (breast and endometrium); lack of cardioprotective effect and cognitive protection in older women. Evidence indicates that vaginal estrogens for the treatment of vaginal dryness are safe and effective; women with a history of breast cancer who do not respond to nonhormonal therapies are advised to discuss the risk and benefits of low-dose vaginal estrogen (dosages of estradiol <25 lg twice weekly) with their healthcare provider	Avoid oral and topical patch Vaginal cream or tablets: acceptable to use low-dose intravaginal estrogen for management of dyspareunia, lower urinary tract infections, and other vaginal symptoms	Oral and patch: High Vaginal cream or tablets: Moderate	Oral and patch: Strong Topical vaginal cream or tablets: Weak
Endocrine (p. 9) Megestrol	Minimal effect on weight; increases risk of thrombotic events and possibly death in older adults	Avoid	Moderate	Strong
Endocrine (p. 9)	Chlorpropamide: prolonged half-life in	Avoid	High	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Sulfonylureas, long-duration Chlorpropamide Glyburide	older adults; can cause prolonged hypoglycemia; causes syndrome of inappropriate antidiuretic hormone secretion Glyburide: higher risk of severe prolonged hypoglycemia in older adults			
Pain medications (p. 9) Meperidine	Not effective oral analgesic in dosages commonly used; may have higher risk of neurotoxicity, including delirium, than other opioids; safer alternatives available	Avoid, especially in individuals with chronic kidney disease	Moderate	Strong
Pain medications (p. 10) Non-cyclooxygenase-selective NSAIDs, oral: Indomethacin Ketorolac, includes parenteral	Indomethacin is more likely than other NSAIDs to have adverse CNS effects. Of all the NSAIDs, indomethacin has the most adverse effects. Increased risk of gastrointestinal bleeding, peptic ulcer disease, and acute kidney injury in older adults	Avoid	Moderate	Strong
Pain medications (p. 10) Pentazocine	Opioid analgesic that causes CNS adverse effects, including confusion and hallucinations, more commonly than other opioid analgesic drugs; is also a mixed agonist and antagonist; safer alternatives available	Avoid	Low	Strong

Organ System, Therapeutic Category, Drugs	Rationale	Recommendation	Quality of Evidence	Strength of Recommendation
Pain medications (p. 10) Skeletal muscle relaxants Carisoprodol Chlorzoxazone Cyclobenzaprine Metaxalone Methocarbamol Orphenadrine	Most muscle relaxants poorly tolerated by older adults because some have anticholinergic adverse effects, sedation, increased risk of fractures; effectiveness at dosages tolerated by older adults questionable	Avoid	Moderate	Strong

2020 Submission

See the table above for the grade assigned to the evidence for each medication class.

The chart below is excerpted from the 2019 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 (Measure)	ACP-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."
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

ACP-based approach (Measure)	ACP-based approach
	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."

GRADE-based approach

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

↓↓↓↓↓

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 American Geriatrics Society 2015 Beers Criteria Update Expert Panel used the Grades of Recommendation Assessment, Development, and Evaluation (GRADE) rating process to rate the quality of evidence. Each panelist independently rated the quality of evidence and strength of recommendation for each criterion using the American College of Physicians' Guideline Grading System (Qaseem et al., 2010), which is based on the GRADE scheme (The GRADE Working Group). 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.

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
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
Strength of Recommendation	
Strong	Benefits clearly outweigh harms, adverse events, and risks, or harms, adverse events, and risks clearly outweigh benefits
Weak	Benefits may not outweigh harms, adverse events, and risks
Insufficient	Evidence inadequate to determine net harms, adverse events, and risks

References:

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.

The GRADE working group. GRADE guidelines—best practices using the GRADE framework. Journal of Clinical Epidemiology [on-line]. Available at http://www.gradeworkinggroup.org/publications/jce_series.htm

2020 Submission

See the table above for the grade assigned to the evidence for each medication class.

The chart below is excerpted from the 2019 Beers Criteria article and contains the definitions for the quality of evidence ratings and the strength of recommendations.

N/A

2020 Submission

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

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

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

N/A

2020 Submission

Methods used for the 2019 update were similar to those used in the 2015 update of the Beers Criteria. The American Geriatrics Society 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.

Overall, the quality of the evidence for each of the medications included in the Beers Criteria recommendations is good. See table above for the quality of evidence rating for the recommendation for each medication or medication class. The table also includes the AGS 2019 Beers Criteria Update Expert Panel rating for the strength of the evidence supporting each recommendation.

2016 Submission

The Beers Criteria was 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 AGS 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, which this measure is based on, 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 medications included in the Beers Criteria recommendations is good. See table under section 1c.16 for the quality of evidence rating for the recommendation for each medication or medication class. The table also includes the AGS 2015 Beers Criteria

Update Expert Panel rating for the strength of the evidence supporting each recommendation. Definitions of these ratings are listed in section 1c.21.

2020 Submission

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

Each updated study contributes to the strength of the measure by updating the medication lists. See section 1c.16 for a table that contains the Beers Criteria recommendations for each drug and drug class that are included in the measure. Evidence tables containing summaries of each study supporting the recommendations can be found on the American Geriatrics Society's website: http://www.americangeriatrics.org/.

The studies consistently mention similar drugs. Since the bodies of evidence all relate to the original Beers list, they maintain consistency in process. See section 1c.16 for a table that contains the Beers Criteria recommendations for each drug and drug class that are included in the measure.

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. However, the criteria are unable to account for the complexity of patients and subpopulations; there may be a small portion of individuals who will 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.

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 use of high-risk medications 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 all 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.

At least 2 different high-risk medications

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

2016^ | 485 | 9.1% | 3.4% | 0.0% | 5.9% | 7.0% | 8.5% | 10.7% | 13.7% | 30.1%

2017 | 482 | 9.9% | 4.1% | 0.0% | 6.1% | 7.3% | 9.0% | 11.9% | 15.4% | 35.5%

2018* | 502 | 9.6% | 3.9% | 0.0% | 5.8% | 7.1% | 8.6% | 11.4% | 14.9% | 27.4%

*For 2018 the average eligible population was 28,463, with a standard deviation of 70,665

^Note: These results are based on a previous specification of the HEDIS measure in which the numerator was based on multiple prescribing events of different high-risk medications instead of the current specification which looks at multiple prescribing events for the same high-risk medication.

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 included in HEDIS data collection and the median and mean eligible population-which is the same as the denominator-for the measure across health plans.

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

2016 | 485 | 6,212 | 25,642

2017 | 482 | 6,476 | 27,903

2018 | 502 | 5,893 | 28,463

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.

N/A

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)*

No data dictionary **Attachment:**

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 first rate (former

Numerator 1) for members who received at least one dispensing event for a high-risk medication was retired. The remaining rate is a better assessment of the riskier, more long-term use of high-risk medications among older adults. It also allows organizations to address potentially inappropriate medication use after one dispensing event and work to prevent further dispensing, to improve on the remaining rate. 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.

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).

Patients who received at least two dispensing events for the same high-risk medication during the measurement year.

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).

Patients who had at least two dispensing events for the same high-risk medication during the measurement year.

Follow the steps below to identify numerator compliance. Include patients who meet criteria in more than one step only once in the numerator. Do not include denied claims.

Step 1: Identify patients with two or more dispensing events (any days supply) on different dates of service during the measurement year for a medication in Table DAE-A. The dispensing events must be for the same drug as identified by the Drug ID in the NDC list. These patients are numerator compliant.

Step 2: For each patient, identify all dispensing events during the measurement year for medications in Table DAE-B. Identify patients with two or more dispensing events on different dates of service for medications in the same medication class (as defined by the AGS Beers Criteria Table 2 and class title below). For example, a prescription for zolpidem and a prescription for zaleplon are considered two dispensing events for medications in the same medication class (these drugs share the same class title or description: Nonbenzodiazepine hypnotics). Sum the days' supply for prescriptions in the same medication class. Identify patients with two or more dispensing events for medications of the same medication class where the summed days' supply exceeds the days' supply criteria listed for the medication. These patients are numerator compliant. For medications dispensed during the measurement year sum the days' supply and include any days supply that extends beyond December 31 of the measurement year. For example, a prescription of a 90-days supply dispensed on December 1 of the measurement year counts as a 90-days supply.

- Note: The intent is to identify all patients who had multiple dispensing events where the summed days' supply exceeds the days' supply criteria; there is no requirement that each dispensing event exceed the days' supply criteria.

Step 3: For each patient, identify all dispensing events during the measurement year for medications in Table DAE-C where average daily dose exceeds the average daily dose criteria listed for the medication. Identify patients with two or more dispensing events on different dates of service that exceed the average daily dose criteria for the same drug as identified by the Drug ID in the NDC list. These patients are numerator compliant. To calculate average daily dose for each dispensing event, multiply the quantity of pills dispensed by the dose of each pill and divide by the days' supply. For example, a prescription for a 30-days supply of digoxin containing 15 pills, .250 mg each pill, has an average daily dose of 0.125 mg. To calculate average daily dose

for elixirs and concentrates, multiply the volume dispensed by daily dose and divide by the days' supply. Do not round when calculating average daily dose.

HIGH-RISK MEDICATIONS (Table DAE-A)

Anticholinergics, First-generation antihistamines---

Brompheniramine, Carbinoxamine, Chlorpheniramine, Clemastine, Cyproheptadine, Dexbrompheniramine, Dexchlorpheniramine, Diphenhydramine (oral), Dimenhydrinate, Doxylamine, Hydroxyzine, Meclizine, Promethazine, Pyrilamine, Triprolidine

Anticholinergics, anti-Parkinson agents---

Benztropine (oral), Trihexyphenidyl

Antispasmodics---

Atropine (exclude ophthalmic), Bellandonna alkaloids, Clidinium-Chlordiazepoxide, Dicyclomine, Hyoscyamine, Methscopolamine, Propantheline, Scopolamine

Antithrombotics---

Dipyridamole, oral short-acting (does not apply to the extended-release combination with aspirin)

Cardiovascular, alpha agonists, central---

Guanabenz, Guanfacine, Methyldopa

Cardiovascular, other---

Disopyramide, Nifedipine (immediate release)

Central nervous system, antidepressants---

Amitriptyline, Clomipramine, Imipramine, Trimipramine, Amoxapine, Desipramine, Nortiptyline, Paroxetine, Protriptyline

Central nervous system, barbiturates---

Amobarbital, Butabarbital, Butalbital, Mephobarbital, Pentobarbital, Phenobarbital, Secobarbital

Central nervous system, vasodilators---

Ergot mesylates, Isoxsuprine

Central nervous system, other---

Meprobamate

Endocrine system, estrogens with or without progestins; include only oral and topical patch products---

Conjugated estrogen, Esterified estrogen, Estradiol, Estropipate

Endocrine system, sulfonylureas, long-duration---

Chlorpropamide, Glimepiride, Glyburide

Endocrine system, other---

Desiccated thyroid, Megestrol

Pain medications, skeletal muscle relaxants---

Carisoprodol, Chlorzoxazone, Cyclobenzaprine, Metaxalone, Methocarbamol, Orphenadrine

Pain medications, other---

Indomethacin, Ketorolac (includes parenteral), Meperidine

HIGH-RISK MEDICATIONS WITH DAYS SUPPLY CRITERIA (Table DAE-B)

Anti-infectives, other (greater than 90 days supply, days supply criteria)---

Nitrofurantoin, Nitrofurantoin macrocrystals, Nitrofurantoin macrocrystals-monohydrate

Nonbenzodiazepine hypnotics (greater than 90 days supply, days supply criteria)---

Eszopiclone, Zolpidem, Zaleplon

HIGH-RISK MEDICATIONS WITH AVERAGE DAILY DOSE CRITERIA (Table DAE-C)

Alpha agonists, central (greater than 0.1 mg/day, average daily dose criteria)---

Reserpine

Cardiovascular, other (greater than 0.125 mg/day, average daily dose criteria)---

Digoxin

Tertiary TCAs (as single agent or as part of combination products), (greater than 6 mg/day, average daily dose criteria)---

Doxepin

Note: NCQA will post a comprehensive list of medications and NDC codes to www.ncqa.org by November 2020. For medications in Table DAE-A and DAE-C, identify different drugs using the Drug ID field located in the NDC list on NCQA's Web site (www.ncqa.org), posted by November 2020.

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

All patients 65 years of age and older.

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 that are 66 years of age and older as of December 31 of the measurement year.

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

Patients who were enrolled in hospice care at any time during the measurement year.

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.)*

N/A

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.)*

N/A

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 denominator: All patients 66 years of age and older as of the end (i.e., December 31) of the measurement year.

Step 2: Identify the numerator: Individuals in the denominator who have dispensed at least two prescriptions for the same high-risk medication (see definition of high-risk medication in section S.6) during the measurement year.

Step 3: Divide Step 2 (numerator) by Step 1 (denominator) to calculate the rate.

Note: For this measure, a lower rate indicates better performance.

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, Electronic Health Data, Electronic Health Records

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

DAE_0022_Testing_Form-637396680504932134.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*)**:** 0022
Measure Title: Use of High-Risk Medications 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 Health Plan performance data from 2012-2014

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	Measure Tested at Level of:
☒ 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 whether Medicare members ages 65 years and older had at least 2 dispensing events for the same high-risk medication. 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 in 2018 for this HEDIS measure. This data came from 502 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

Validity statistics were calculated from 2014 HEDIS Health Plan performance data that included 488 Medicare health plans. This included all Medicare health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size. The average (mean) eligible population for this measure across health plans was 22,043.

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, as well as the average and median eligible population for the measure across health plans.

Table 1. Mean and median eligible population for *Use of High-Risk Medications in Older Adults,* 2018

Year	Number of Plans	Mean number of eligible members per plan	Median number of eligible members per plan
2018	502	28,463	5,893

2016 Submission
This question was not on the 2016 form.

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 only 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 *Use of High-Risk Medications 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 *Use of High-Risk Medications 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

Reliability was estimated by using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped.

Reliability used here 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. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

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. Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the *Use of High-Risk Medications 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
Use of High-Risk Medications in Older Adults	**502**	**32 - 679844**	**0.936**	**0.006**	**(0.924, 0.947)**
Tercile 1	166	32 - 2456	0.857	0.012	(0.833, 0.881)

Stratification	Number of Plans	Number of Eligible Members per Plan (min - max)	Mean Signal-To-Noise Reliability	SE	95% CI
Tercile 2	165	2469 - 15564	0.986	0.001	(0.985, 0.988)
Tercile 3	171	15856 - 679844	0.997	0.000	(0.997, 0.997)

SE: Standard Error of the mean.

95% CI: 95% confidence interval.

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
Use of High-Risk Medications in Older Adults	502	0.193	0.798	0.950	0.988	0.998	0.999	1.000
Tercile 1	166	0.249	0.615	0.812	0.928	0.962	0.980	1.000
Tercile 2	165	0.962	0.975	0.982	0.988	0.993	0.995	0.997
Tercile 3	171	0.987	0.994	0.996	0.998	0.999	0.999	1.000

Table 2b. Distribution of Plan-Level Signal-To-Noise Reliability for the *Use of High-Risk Medications in Older Adults* Measure by Terciles of the Denominator Size and for All Submissions, 2018

2016 Submission

Using 2014 HEDIS Health Plan performance data, reliability for this measure was calculated as 0.99814 for receipt of one or more high-risk prescriptions and 0.99594 for receipt of two or more high-risk prescriptions.

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 provides the point estimate of mean signal-to-noise reliability, its standard error, and the 95% CI for the *Use of High-Risk Medications in Older Adults* measure stratified by the denominator size (distribution of the number of eligible members per plan). The reliability estimate is 0.936, and the 95% CI is (0.924, 0.947), indicating very good reliability for the measure. Stratified analyses show that reliability increases as plan size gets larger.

Table 2b summarizes the distribution of plan-level signal-to-noise reliability estimates for the *Use of High-Risk Medications in Older Adults* measure. The estimates range from 0.193 to 1.000. The 50th percentile is 0.988, 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. Note that the low minimum reliability estimate (0.193) and low observed reliability in the first tercile is likely explained by a handful of very

small plans (small denominators) who inflate the sigma-squared error in the signal-to-noise calculation (see 2a2.2, above). Very high reliability is observed in a majority of plans and 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 both indicators in this measure are highly reliable.

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:

- Is the *Use of High-Risk Medications in Older Adults* measure correlated with the HEDIS *Potentially Harmful Drug-Disease Interactions in Older Adults* measure, which 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?

We hypothesized that the *Use of High-Risk Medications in Older Adults* measure would be correlated with the rates of *Potentially Harmful Drug-Disease Interactions in Older Adults* measure, particularly Rate 2 (Dementia) and Rate 3 (Chronic kidney disease). In addition, organizations that perform well on *Use of High-Risk Medications in Older Adults* should perform well on the other medication safety measure, *Potentially Harmful Drug-Disease Interactions 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 *Use of High-Risk Medications in Older Adults* measure is aligned with the American Geriatrics Society (AGS) Beers Criteria, which recommends drugs to be avoided in older adults. 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: 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.

Method of Testing Construct Validity: We tested for construct validity by exploring whether the two rates 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 indicators should perform well on the other indicator 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.

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 *Use of High-Risk Medications in Older Adults* and *Potentially Harmful Drug-Disease Interactions in Older Adults* Performance Scores, 2018

Measure	**Correlation Coefficient:** *Use of High-Risk Medications in Older Adults*
Use of High-Risk Medications in Older Adults	- -
Drug-disease interaction: History of Falls*	0.62
Drug-disease interaction: Dementia*	0.53
Drug-disease interaction: Chronic Kidney Disease*	0.24

Note: All correlations are significant at p<0.001

*The *Potentially Harmful Drug-Disease Interactions in Older Adults* measure has three rates. The first rate assesses the percentage of patients 65 and older with a history of falls who received a high-risk medication. The second rate assesses the percentage of patients 65 and older with dementia who received a high-risk medication. The third rate assesses the percentage of patients 65 and older with chronic kidney disease who received a high-risk medication.

- - 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 and retirement of rate 1, which focused on one dispensing event for a high-risk medication. 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: This measure was developed to address high-risk medication use in the elderly. NCQA and the GMAP worked together to assess which medications to include based on recommendations in the AGS Beers Criteria. 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. 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. The measure was then 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. In summary, the measure was deemed to have the desirable attributes of a HEDIS measure in 2006 (relevance, scientific soundness, and feasibility). 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.

Results of Construct Validity Testing: The results in Table 1a indicate that there was a high correlation between the first and second rate in the measure. There were moderate correlations between both rates and the four rates in the other medication safety measure.

Table 1a. Correlations among both rates in the measure and a drug-disease interaction measure[1]

Measure	(Pearson Correlation Coefficients) Rate 1: One high-risk medication	(Pearson Correlation Coefficients) Rate 2: Two high-risk medications
Rate 1: One high-risk medication	*	*
Rate 2: Two high-risk medications	.8745	*
Drug-disease interaction: History of Falls	0.307	.2735
Drug-disease interaction: Dementia	0.454	.4390
Drug-disease interaction: Chronic Kidney Disease	0.367	.3552
Drug-disease interaction: Total	0.386	.3913

Note: All correlations are significant at p<.05

[1]The *Potentially Harmful Drug-Disease Interactions in the Elderly* measure has four rates. The first rate assesses the percentage of patients 65 and older with a history of falls who received a high-risk medication. The second rate assesses the percentage of patients 65 and older with dementia who received a high-risk medication. The third rate assesses the percentage of patients 65 and older with chronic kidney disease who received a high-risk medication. The fourth rate is the sum of the three numerators divided by the sum of the three denominators for the three previous rates. Note: "high-risk" medications for each condition are based on recommendations in Table 3 of the American Geriatrics Society Beers Criteria.

*cell intentionally left blank

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.

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

2016 Submission

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 use of high-risk medications in the elderly.

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*)

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*)

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*)

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.

2020 Submission

N/A. Not an intermediate or health outcome, PRO-PM, or resource use measure.

2016 Submission

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

Comparison of means and percentiles; analysis of variance against established benchmarks: if sample size is >400, we would use an analysis of variance.

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 4. Variation in Performance, 2018

Year	Measure	N	Mean eligible population	Mean rate (%)	SD	Min	P10	P25	P50	P75	P90	Max	IQR	p-value
2018	*Use of High-Risk Medications in Older Adults*	502	28,463	9.6	3.9	0.0	5.8	7.1	8.6	11.4	14.9	27.3	4.3	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

2012 to 2014 HEDIS Health Plan Performance Data

At least one high-risk prescription

Year	Number of Plans	Mean	Standard Deviation	Min	Max	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
2012	498	21.0	6.4	5.5	54.6	14.0	16.5	19.9	24.5	30.0
2013	494	18.0	6.1	1.0	50.5	11.5	13.8	16.7	21.1	25.8
2014	488	13.2	6.0	2.6	46.8	7.6	9.2	11.6	16.1	21.7

At least two high-risk prescriptions

Year	Number of Plans	Mean	Standard Deviation	Min	Max	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
2012	498	6.5	2.9	1.2	25.2	3.5	4.7	6.0	7.8	10.1
2013	494	3.1	2.3	0.0	20.6	1.1	1.7	2.4	4.0	6.0
2014	488	2.1	2.0	0.0	20.8	0.6	0.9	1.4	2.5	4.6

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

There is a 4.3 percentage point gap in performance between Medicare plans at the 25th and 75th percentiles. This gap represents on average 1,138 more older adults with at least two high-risk medications in low performing Medicare plans compared to high performing plans. The difference in performance between plans in the 25th percentile and 75th percentile is statistically significant.

2016 Submission

This question was not on the 2016 form.

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
This question was not on the 2016 form.

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
This question was not on the 2016 form.

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
This question was not on the 2016 form.

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)

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 Ratings 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 Ratings http://www.ncqa.org/ReportCards/HealthPlans/HealthInsurancePlanRankings/HealthPlanRatingsPreview.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx Payment Program CMS Quality Payment Program https://qpp.cms.gov/ CMS Quality Payment Program https://qpp.cms.gov/ Regulatory and Accreditation Programs HEDIS®-Health Plan http://www.ncqa.org/Programs/Accreditation/HealthPlanHP.aspx Patient Centered Medical Homes (PCMH) https://www.ncqa.org/programs/health-care-providers-practices/patient-centered-medical-home-pcmh/ HEDIS®-Health Plan http://www.ncqa.org/Programs/Accreditation/HealthPlanHP.aspx Patient Centered Medical Homes (PCMH) https://www.ncqa.org/programs/health-care-providers-practices/patient-centered-medical-home-pcmh/ Quality Improvement (external benchmarking to organizations) 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 ratings.
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.
NCQA PATIENT-CENTERED MEDICAL HOME (PCMH): This measure is used in the Patient Centered Medical Home Recognition program, which identifies medical practices that have invested in a model of care that puts patients at the forefront and where continuous quality improvement is a priority.
CMS QUALITY PAYMENT PROGRAM: This measure is used in the Quality Payment Program (QPP) which is a reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs).

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 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 and questions about the supporting guidelines for the measure. 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 and other entities such as the Centers for Medicare and Medicaid Services as illustrated by its use in the Quality Payment Program.

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 and additional examples 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 2016 to 2018 data shows relatively stable performance and room for improvement across Medicare Plans (see section 1b.2 for summary of data from health plans). In 2018, the average performance was 9.6%. There was a 9 percentage point difference between plans at the 10th and 90th percentiles. This large difference in performance represents a persistent gap in care and room for improvement in medication safety for older adults, particularly given the substantially large average denominator size of all plans reporting on this measure and therefore the great number of older adults at risk for adverse drug events. Although overall rates aren't changing, there have been an increase in the number of plans reporting from 2016 (n=485) to 2018 (n=502). Many of the new plans reporting have larger denominator sizes, as demonstrated by the increasing mean denominator size over the three years of data.

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 and clinicians should weigh the risks and benefits of using these medications for their individual patients.

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)

2993 : Potentially Harmful Drug-Disease Interactions in Older Adults (DDE)

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

Pharmacy Quality Alliance: Use of High-Risk Medications in Older Adults

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 Potentially Harmful Drug-Disease Interactions in Older Adults (DDE) measure and NQF 0022 have a similar focus (measuring potentially inappropriate medication use in older adults) and reporting level (health plan), however they have different target populations. The DDE measure targets patients with a specific condition or disease that can experience adverse effects when combined with certain medications that are recommended to be avoided for that condition. This measure (NQF 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. The DDE measure (NQF 2993) 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. This measure (NQF 0022) is harmonized with PQA's Use of High-Risk Medications in the Elderly (HRM) measure. The HRM measure is also based on the AGS Beers Criteria Table 2 and targets the same population of older adults. However, CMS will retire this display measure for 2021

and no longer reports this measure in the Patient Safety reports for the 2019 measurement year. Commenters supported retiring this measure.

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: Brittany, Wade, wade@ncqa.org, 202-530-0463-

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 BlueSheild

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

Technical Measurement Advisory Panel (TMAP):

Andy Amster, MSPH, Kaiser Permanente

Sarah Bezeredi, MBA, MSHL, UnitedHealth Group

Jennifer Brudnicki, MBA, Inovalon Inc.

Lindsay Cogan, PhD, MS, New York State Department of Health

Mike Farina, RPh, MBA, Capital District Physicians' Health Plan

Matt Flores, MS, RRT, CHCA, Advent Advisory Group

Scott Fox, MS, MEd, FAMIA, The MITRE Corporation

Carlos Hernandez, CenCal Health

Harmon Jordan, ScD, Westat

Virginia Raney, LCSW, Center for Medicare and Medicaid Services

Lynne Rothney-Kozlak, MPH, Rothney-Kozlak Consulting, LLC

Laurie Spoll, Aetna

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2006

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:

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. © 2012 by the National Committee for Quality Assurance