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

Brief Measure Information

NQF #: 2456

Corresponding Measures:

De.2. Measure Title: Medication Reconciliation: Number of Unintentional Medication Discrepancies per Medication Per Patient

Co.1.1. Measure Steward: Brigham and Women´s Hospital

De.3. Brief Description of Measure: This measure assesses the actual quality of the medication reconciliation process by identifying errors in admission and discharge medication orders due to problems with the medication reconciliation process. The target population is any hospitalized adult patient. The time frame is the hospitalization period.

At the time of admission, the admission orders are compared to the preadmission medication list (PAML) compiled by trained pharmacist (i.e., the gold standard) to look for discrepancies and identify which discrepancies were unintentional using brief medical record review. This process is repeated at the time of discharge where the discharge medication list is compared to the PAML and medications ordered during the hospitalization.

1b.1. Developer Rationale: This measure will drive hospitals to implement interventions to truly improve their medication reconciliation processes. To date, Joint Commission requirements for medication reconciliation have led mostly to pro forma compliance, for example, checking a box saying that medication reconciliation has been performed, without knowing whether clinical care has been affected. By directly measuring error rates in medication orders, this new measure will enable hospitals to better understand where their errors are occurring and the types of errors that exist. This will enable them to implement targeted interventions that actually reduce error rates. The result will be true improvements in medication safety during transitions in care.

The rate of unintentional discrepancies per patient is unacceptably high in this country, and there is variation by site. In the six sites studied using the proposed methodology, the range was 2.78 to 4.57 discrepancies per patient (average of 3.44 per patient), thus making medication reconciliation errors the single biggest source of medication errors in the hospital (i.e., as opposed to errors in prescribing, transcribing, or administration).

Studies of medication reconciliation interventions demonstrate that improvements in important outcomes are indeed possible. In a recent systematic review conducted by our group (2), we identified 26 studies. Studies consistently demonstrated a reduction in medication discrepancies (17/17 studies), potential adverse drug events (5/6), and adverse drug events (2/2), and 2/8 studies showed a reduction in health care utilization. In the first Multi-center Medication Reconciliation Quality Improvement Study, involving 1648 patients across 5 hospitals, evidence-based interventions to improve medication reconciliation resulted in a reduction in medication discrepancies by 8% per month over baseline temporal trends (adjusted incident rate ratio 0.92,

95% CI 0.87-0.97, p=0.002), using the NQF proposed metric and methodology (technically, the measure was discrepancies per patient, using the number of medications as a model offset in the Poisson regression, which essentially is the same as discrepancies per medication per patient). In the recently completed MARQUIS2 study, involving 4947 patients across 17 hospitals, results were similar but even more robust (adjusted IRR 0.95 per month, 95% CI 0.93-0.97, p=<0.0001).

Citations for 1b.1:

1. Salanitro AH, Kripalani S, Resnic J, et al. Rationale and design of the Multicenter Medication Reconciliation Quality Improvement Study (MARQUIS). BMC health services research. 2013;13:230.

2. Mueller SK, Sponsler KC, Kripalani S, Schnipper JL. Hospital-Based Medication Reconciliation Practices: A Systematic ReviewHospital-Based Medication Reconciliation Practices. Arch Intern Med. Jun 25 2012:1-13.

3. Schnipper JL, Mixon AS, Stein J, Wetterneck TB, Kaboli P, Mueller S, Labonville S, Minahan JA, Burdick E, Orav EJ, Goldstein J, Nolido NV, Kripalani S. The effects of a multi-faceted medication reconciliation quality improvement intervention on patient safety: final results of the MARQUIS study. BMJ Qual Saf 2018; 27(12):954-964.

4. Schnipper JL, Reyes Nieva H, Mallouk M, et al. Effects of a refined evidence-based toolkit on medication reconciliation quality and safety at multiple hospitals: results of the MARQUIS2 study. Plenary, Society of Hospital Medicine Annual Meeting, National Harbor, MD.

S.4. Numerator Statement: For each sampled inpatient in the denominator, the total number of unintentional medication discrepancies in admission orders plus the total number of unintentional medication discrepancies in discharge orders.

S.6. Denominator Statement: The patient denominator is the sum of the number of medications in the gold standard medication lists plus the number of unintentionally ordered additional medications in a random sample of all adults admitted to the hospital. Our recommendation is that 25 patients are sampled per month, or approximately 1 patient per weekday.

So, for example, if among those 25 patients, there are 110 gold standard medications and 40 unintentionally ordered additional medications, and 75 unintentional discrepancies are identified, the measure outcome would be 75/150 = 0.5 discrepancies per medication per patient for that hospital for that month.

S.8. Denominator Exclusions: Patients that are discharged or expire before a gold standard medication list can be obtained.

De.1. Measure Type: Outcome

S.17. Data Source: Electronic Health Data, Electronic Health Records, Instrument-Based Data, Other, Paper Medical Records

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Sep 09, 2014 **Most Recent Endorsement Date:** Sep 09, 2014

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 health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Summary of prior review in 2014

- The measure was previously reviewed by the Care Coordination Committee and passed with a Moderate vote.
- The Committee agreed the evidence presented provided moderate support for the measure focus. The evidence included a systematic review consisting of 26 studies consistently demonstrating that medication reconciliation interventions result in a reduction in medication discrepancies, potential adverse drug events, adverse drug events, and reduction in health care utilization, however the studies were of fair quality, as graded by the United States Preventive Services Task Force.
- While the Committee viewed this measure as a proxy outcome for a short-term outcome of good care coordination around medication, they did not find a strong connection between the measure and long-term error reduction and overall better patient outcomes.

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:

- Since the 2012 systematic review discussed above, larger multi-site studies have been conducted which demonstrate the consistent link between medication reconciliation quality improvement interventions and reductions in medication discrepancies and further support the evidence base.
- The MARQUIS study included 1648 patients at 5 hospitals. The study showed evidence-based interventions to improve medication reconciliation resulted in a reduction in medication discrepancies by 8% per month over baseline temporal trends (adjusted incident rate ratio 0.92, 95% CI 0.87-0.97, p=0.002), using the NQF proposed metric and methodology.
- In the MARQUIS2 study, which included 4947 patients at 17 hospitals, results were similar but even more robust (adjusted IRR 0.95 per month, 95% CI 0.93-0.97, p=<0.0001).

Question for the Committee:

- *Is there at least one thing that hospitals can do to achieve a change in the measure results?*

- *Is the evidence or conceptual logic model strong enough to support the benefits of reduced medication discrepancies through medication reconciliation?*
- *Are there long-term benefits of medication reconciliation interventions?*

Guidance from the Evidence Algorithm

Outcome measure (box 1) → Evidence that the outcome (discrepancies) can be impacted by at least one healthcare action (box 2) → Yes → Pass

RATIONALE:

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

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.

- Data from 1427 sites from data collected by Leapfrog indicates a mean rate of 0.18, standard deviation of 0.17, 25th percentile 0.06, 75th percentile 0.25, minimum 0, and maximum 1.24.
- In MARQUIS2, involving 17 sites, the number of discrepancies per medication per patient ranged from 0.04 to 0.36. The developer offers interpretation of performance results. A discrepancy rate of 0.36 correlates with an 18% error rate. If the average patient is on 10 medications, there would be 3-4 medication errors per patient.

Disparities

- While several studies show that older age and number of medications impact discrepancies, at least one study shows patients over age 85 have fewer medication discrepancies.
- Other risk factors for discrepancies include: low health literacy, low education attainment, and poor patient understanding of medications. Medicaid insurance, patient sex, or race or ethnicity do not correlate with medication discrepancies. Having a recent medication list in the EMR has been shown to be protective.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are there disparities that exist that support risk adjustment or stratification to help focus performance improvement efforts or to compare hospitals to one another?*

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

RATIONALE:

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

1a. Evidence

Comments:

**Reasonable supporting evidence

**no

**For this maintenance measure, the developer provided updated evidence. Larger multi-site studies demonstrate the consistent link between medication reconciliation quality improvement interventions and reductions in medication discrepancies. Also, in the MARQUIS2 study, which included 4947 patients at 17 hospitals, results were similar but even more robust than the previous results from MARQUIS published in 2013. As the developer pointed out, the current requirements for medication reconciliation are mostly involved with checking a box saying medication reconciliation has been formed without knowing whether clinical care has been affected. By directly measuring the errors expressed as the discrepancy rates, this measure will enable hospitals to better understand where their errors are occurring and the types of errors. This will enable them to implement targeted interventions to improve medication safety. Also according to the developer, the rate of unintentional discrepancies per patient is high and there is variation by site, ranging from 2.78 to 4.57 (average of 3.44 per patient), making medication reconciliation errors the single biggest source of medication errors in the hospital (i.e., as opposed to errors in prescribing, transcribing, or administration). In MARQUIS2, involving 17 sites, the number of discrepancies per medication per patient ranged from 0.04 to 0.36. Interpreted by the developer, a discrepancy rate of 0.36 correlates with an 18% error rate; if the average patient is on 10 medications, there would be 3-4 medication errors per patient. But, I am disappointed that no evidence is shown that medication reconciliation directly results in a reduction in medication errors.

**moderate evidence with a prelim rating of pass

**there is a strong link to medication safety

**Evidence is on point and persuasive. I am not aware of any contradictory literature since endorsement.

1b. Performance Gap

Comments:

**Certainly a gap in performance.

**yes

**In MARQUIS2, involving 17 sites, the number of discrepancies per medication per patient ranged from 0.04 to 0.36. So the performance gap is quite large and demonstrated adequately.

**Data from 1427 sites from data collected by Leapfrog indicates a mean rate of 0.18

**the rate of error, compared to six sigma is significant

**There is ample evidence of care gaps in this area.

Disparities:

**Disparity data not provided. Its possible that this issue is widespread enough that it impacts all.

**yes

**Most studies cited showed age and number of medications impact discrepancies, and other risk factors involve with low healthcare literacy, low level of education, poor patient understanding of medications. So there is a room for improvement by addressing disparities among various patient subgroups.

**minimal data provided

**the measure shows equal areas for improvement

**yes, not correlation with race/ethnicity. Paradoxically, less of a problem in over 85 population

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.

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

Evaluators: Subgroup 1

Methods Panel Review (Combined)

Scientific Methods Panel Evaluation Summary:
This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Scientific Methods Panel Votes:
Reliability: H-0; M-4; L-2; I-0 (pass)
Validity: H-0; M-3; L-2; I-1 (consensus not reached)

In their preliminary analyses, subgroup members did not reach consensus on the validity of the measure. During the Panel's discussion of the measure, members suggested that there may be a need to incorporate some type of additional risk or case-mix adjustment. The Panel noted that it may be more difficult to reconcile medications for patients with more complex regimens (i.e., more medications), leading to a higher likelihood of discrepancies in those patients. While the measure does account for that issue by counting the number of discrepancies per medication per patient, Panel members suggested that the relationship between number of medications and complexity may not be entirely linear, meaning that the developer's approach may not adequately capture differences in risk across patients. In addition, the developer noted during the discussion that the measure is intended for internal quality improvement purposes, and may not be appropriate for between-hospital comparisons. This caused concern among Panel members, since NQF endorsement implies that measures are suitable for both quality improvement and accountability applications. Ultimately, the subgroup did not reach consensus on the validity of the measure.

Reliability

- Developer tested reliability of the data elements with an inter-rater reliability assessment, wherein two study pharmacists independently collected medication histories for 19 randomly-selected patients, calculating the percentage of patients for whom there was complete agreement in medication, dose, route, and frequency across the two assessments. Among all the medications

recorded for each patient, there was complete agreement in medication, dose, route, and frequency for 147 of 192 medications (77% agreement).

- The developers evaluated inter-rater reliability of the discrepancy scoring system by analyzing the last 4 quarterly cases, consisting of a total of 44 medications and 128 ratings each for admission and discharge discrepancies. For the presence of admission discrepancies, the developer found agreement for 116/128 ratings (91% agreement); Kappa = 0.64 (substantial agreement). For the presence of discharge discrepancies, the developer found agreement for 116/128 ratings (91% agreement); Kappa = 0.64 (substantial agreement).
- In the analysis of reliability of the scoring system, kappas were statistically significant from zero (for admission discrepancies, $Z=7.29$, $p<0.0001$; for discharge discrepancies, $Z=7.34$, $p<0.0001$).

Validity

- The developer provided empirical score-level testing showing that that hospitals that had significant improvement in their medication discrepancy rates (the critical element of the proposed measure) from the beginning to the end of the study had a greater improvement in the proportion of patients who received patient-level medication reconciliation interventions (such as a "best possible medication history" in the emergency department) than those hospitals that did not see improvement in their discrepancy rates. The data provided shows that 9 of 17 study sites had significant improvement in their discrepancy rates in the last 6 months of the study compared to the first 6 months of the study. Compared with those sites that did not show improvement, those that did show improvement had a greater increase in the proportion of patients who received patient-level interventions (55% absolute improvement vs. 22% absolute improvement).
- The developer also notes that the literature shows that pharmacists take more accurate medication histories than nurses or physicians, suggesting that a preadmission medication history taken by a trained expert pharmacist is itself a reasonable proxy for a "gold standard" medication history. The developer's implication is that the measure is using the gold standard itself to identify discrepancies in medication histories. The developer provided materials to show how expert pharmacists are trained and materials showing that the process used to measure discrepancies is transparent and systematic.
- The measure is not risk adjusted. However, since last endorsement this measure has been updated. The number of discrepancies is now divided by the number of medications to more accurately account for the fact that discrepancies (errors) are dependent on the number of opportunities for error. This accounts for the fact that hospitals patient populations may vary with respect to the complexity of their medication regimens.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented?*
- *The Scientific Methods Panel 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., risk-adjustment approach, testing methods/results)?*
- *The Scientific Methods Panel did not reach consensus on the validity of the measure. The Committee must discuss and vote on validity.*

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

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient** ☒ **Consensus Not Reached**

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability – Specifications

Comments:

**Not concerns about the specifications. Also do not agree that applying risk adjustment is critical to the implementation of this measure.

**none

**Scientific Methods Panel evaluates the reliability, which was passed as moderate.

**no concerns

**This is a major area of concern, the suporting methodology is very complicated and does not highlight inter-rater reliability

**I found the discussion of the methodology hard to follow, need better examples and more precise explanations for users. There is likely to be substantial inconsistency in application due to heterogeneity in pharmacist availability, unclear "randomization" instructions. Kappa scores as "substantial agreement", but this is an arbitrary definitions and likely not clinically applicable here.

2a2. Reliability – Testing

Comments:

**The reliability appears to be appropriate but their testing process did not seem to be robust.

**no

**No.

**none

**yes, there is no granularity to what constitutes a significant discrepancy

**Yes. Very subjective with respect to definitions of intentional and unintentional. Pharmacists are employed by facilities being evaluated.

2b1. Validity –Testing

Comments:

**Appears appropriate

**Scientific panel did not reach consensus

**The Scientific Methods Panel did not reach consensus on validity because of concerns over the developer's approach that may not adequately capture differences in risk across patients. In additional, the developer noted that the measure is intended for internal quality improvement purposes and may not be appropriate for comparisons among hospitals. This lack of applications for both quality improvement and accountability is of concern for NQF endorsement.

**consensus was not reached by methods panel

**no

**Yes, concerns regarding subjectivity and conflict of interest.

2b4-7. Threats to Validity
2b4. Meaningful Differences

Comments:

**Complex patients who have multiple medications are at higher risk for these types of events. Risk adjustment should not be applied. I have more concerns about the feasibility of the measure than the validity. I would also like to see more sampled charts per month.

**Scientific panel did not reach consensus

**There does not seem to be concerns about missing data by the Scientific Methods Panel.

**in my opinion, no

**the Methofoly is far more complicated than the numerator and denominator descriptions I also worry that within-institution analysis supplies the rates

**subjective scoring by pharmacists employed by facility

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment

Comments:

**Concerns about the generation of a Gold Standard list and relatedly pro forma application of this approach leading to inflated results.

**Scientific panel did not reach consensus

**There is no risk adjustment and it is considered to be appropriate.

**The measure is not risk adjusted

**by taking out OTC and other drugs there may be a problem with risk adjustment and disparity bauses of affordability

**Risk adjustment absent. One wonders if this might be improved by adjusting for admission diagnosis? Number of meds?

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2456

Measure Title: Medication Reconciliation: Number of Unintentional Medication Discrepancies per Patient

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.**

   Panel Member #1: It is not clear how many patients should be included for this measure, should there be a minimum number of patients required for this measure? Given that this measure requires trained pharmacists to ascertain the discrepancies, it will be very resource intensive to include a large number of patients. Unless it is aimed to measure all patients, a sampling scheme should be specified.

   It seems that the same number of discrepancies will be treated the same irrespective of whether they happened to one patient or several patients.

   Lack of sufficient reliability testing information is also a major concern.

   Panel Member #2: Measure is only as good at the instructions to and instruction of the pharmacist in both taking history and comparing to hospital medication list. Substantive committee should review documentation.

   Panel Member #3: The measure developers state 'The main barrier to data collection has been the availability of a trained pharmacist at each site'. Although this is not necessarily a specification problem, it does lead to questions about scalability.

   Panel Member #4: The decisions around discrepancies that are "intentional" versus not seem somewhat to depend on the individual "trained pharmacist" interpretation. They do provide extensive training materials and guidelines however. They also conducted inter rater reliability testing as described below.

Panel Member #5: The denominator statement in S6 is unclear although the example helps to add clarity. I like the metric of per medication per patient but this was difficult for me to be sure I understood. Perhaps another example would help clarify. I also have questions on the sampling method concerning "randomized in the order in which they are approached" which does not seem like a random sampling procedure to me and may introduce bias. For data source, the measure developer indicates that the measure also uses "instrument based data" but that is not defined or described.

Panel Member #6: "Gold standard medications" not defined (I cannot find it).

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

   Panel Member #1: The developer used 19 randomly selected patient records to assess if they could be abstracted consistently by two pharmacists. The developer reported the proportion of agreement.

   In addition, the developer also evaluated the inter-rater reliability of the discrepancy scoring system using 4 cases.

   Panel Member #2: Interrater reliability of discrepancies found.

   Panel Member #3: The methods seem appropriate. My two questions are the sample size – why only 4 patients for inter-rater reliability? The measure developers show results at the medication level, but all results for the same coder will be correlated. I would have liked to see results from more cases. Also, the analysis appears to be from an original trial where data collectors are likely to get more monitoring and support – since this is a renewal, I imagine implementation has moved well beyond this controlled environment.

   Panel Member #4: To evaluate inter-rater reliability of the gold-standard medication histories, they compared 19 randomly selected medication histories that were collected independently by two study pharmacists. To evaluate inter-rater reliability of the discrepancy scoring system, they analyzed a total of 44 medications and 128 ratings each for admission and discharge discrepancies (i.e., 256 data points).

   Panel Member #5: In section 1.6 it is indicated that reliability testing of the discrepancy scoring system was performed with 4 patients only (1 from each of the MARQUIS study sites) which raises questions as to the adequacy of this approach.

   Panel Member #6: Inter-rater reliability was assessed for 19 patients at one medical center.
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: The proportion of agreement for establishing the gold-standard medication histories is not high, only 77%, this calls into question all the subsequent evaluation results. It cannot be said that the "inter-rater reliability for the gold-standard medication history was high."

   Testing of the scoring system should be based on more cases, four chosen cases do not seem to be adequate.

Panel Member #2: Kappa of 0.64, considered in the literature substantial agreement. I would assess as low but potentially acceptable. Detailed analysis of differences among coders limited, but more discrepancies in labeling discrepancy than counts.

Panel Member #3: The measure developers found kappa statistics of 0.64 for both the admission discrepancies and discharge discrepancies. I would put this more at the 'moderate' level than substantial agreement (as the authors suggest), but still pretty good. As mentioned above, the rating for individual medications are probably correlated at the pharmacist level.

Panel Member #4: The kappa for the presence of admission discrepancies was 0.64 (substantial agreement) and the kappa for the presence of discharge discrepancies was also 0.64 (substantial agreement) across all raters. In the analysis of reliability of the scoring system, kappas were statistically significant from zero (for admission discrepancies, $Z=7.29$, $p<0.0001$; for discharge discrepancies, $Z=7.34$, $p<0.0001$). Overall, reliability was good, but was somewhat lower for discrepancy type and reason. However, those are to be used for internal QI purposes only and are not part of the measure. They note the picked "challenging" cases as well, and that reliability on average is likely to be higher. It may have been good to evaluate cases at various levels of complexity.

Panel Member #5: In section 2a2.2 it is described that of the 19 medication histories collected independently by 2 pharmacists, that complete agreement was reached on only 77% which seems inadequate.

Panel Member #6: Kappa = 0.64. Concerned that the low/medium reliability results were not tested at other facilities. Assuming the testing was completed at the same site as development and therefore the pharmacists were likely more familiar with the measure than those at other facilities.

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 Panel Member #1:** Proportion of agreement is not sufficient.

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

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

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

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

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

    ☐ **Insufficient** (NOTE: Should rate <u>INSUFFICIENT</u> 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.**

    Panel Member #1: Reliability testing is inadequate. The proportion of agreement for the gold-standard medication history was not high (77%, without correcting for chance).

    Panel Member #2: Training materials need to be reviewed by substantive experts.

Kappa is low but in acceptable range. I could be convinced it is too low.

Panel Member #3: Moderate is the highest you can get without score level testing.

Panel Member #4: Score level testing was not performed. Also the number of histories evaluated for inter-rater reliability was quite small (19) and the number evaluated for reliability of the discrepancy scoring systems was only 4 which may not be enough to say the results are repeatable.

Panel Member #6: See above: single site testing with small sample size. Concerns about the generalizability of the reliability results.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member #2: None.

Panel Member #3: No concerns about exclusions; like the stratification for med, surg, IUC and other.

Panel Member #4: The developers claim that the exclusions are practical ones and "there is no alternative". However, excluding patients who are unavailable to be seen by a pharmacist or who decline to talk to the pharmacist could introduce bias into the results. Is there really no alternative to a live interview to getting medication history?

Panel Member #5: I'm not sure about the defendability of indicating that "patients who are discharged before a gold standard medication list can be obtained" are excluded because I worry that this could be used as an excuse for not getting data on difficult patients. To address this, should a time limit be set (e.g., those discharged in 6 hours or less)?

Panel Member #6: Exclusions only for circumstances that would result in no data collection. No discussion of the impact of excluding patients that decline to talk to a pharmacist. This may result in a cultural or SES bias.

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

Submission document: Testing attachment, section 2b4.

Panel Member #1: For this reason, it would be very helpful to specify the number of cases required for this measure. The results shown in 2b1.1 do seem to indicate meaningful differences across sites.

Panel Member #2: None. Error rates in med reconciliation appear high and declines observed in trials substantial.

Panel Member #3: No concerns

Panel Member #5: In section 2b4.1 the developers indicate they used a ttest. Is this the correct test for count data (which has a Poission distribution and not a normal distribution)?

Panel Member #6: None

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.

Panel Member #1: It is important to specify a sampling scheme for hospitals to follow, if different hospitals select cases differently, then it becomes difficult to compare results across hospitals.

Panel Member #2: Efforts need to be made to improve training so kappa measure of interrater agreement increases over time.

Panel Member #5: No concerns.

Panel Member #6: None

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

Submission document: Testing attachment, section 2b6.

Panel Member #2: No concerns about missing data, but concern about selection of patients to be assessed.

Panel Member #3: No concerns

Panel Member #5: I have no concerns about missing data – but rather have concerns that some of the medication reconciliation data may be incorrect as there is no way to tell if the med rec done produces correct data.

Panel Member #6: May result in bias – excluding patients that decline to speak to pharmacist. No data to show that this will occur at random.

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**

Panel Member #1: I think it is acceptable this measure is not risk adjusted.

Panel Member #2: No risk adjustment. Adjusting for mix of patients with high vs low number of meds, which could differ across facilities, might be appropriate, and revision shifts to errors per medication.

Panel Member #4: It is not clear that "in the name of simplicity, we have chosen not to recommend risk adjustment" is an adequate rationale**.** They go on to state that the number of discrepancies is highly correlated with the total number of medications and "suggest modifying the previous metric or adding a second metric to be the number of unintentional medication discrepancies per medication per patient. It is not clear they have actually done this or what impact it would have?

Panel Member #5: I found section 2b3.2 to be difficult to understand (when describing how the developers recommend modifying the previous metric). An example would increase clarity.

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**

Submission document: Testing attachment, section 2b2.2 Panel Member #4: 2b1.2

Panel Member #1: Relying on pharmacists to ascertain PAML is supported by the literature. However, given the importance of PAML, the fact that two trained pharmacists could only achieve 77% agreement is deeply concerning and calls into questions any subsequent results based on PAML.

Relating improvement in discrepancies with patient-level interventions is a worthy attempt at establishing validity.

Panel Member #2: Literature review justifies using pharmacists to establish gold standard history of drugs and pharmacists can review and compare lists to hospital medical system.
Need to further examine sources of discrepancy in interrater tests and apply lessons to documentation and training.

Panel Member #3: It is not clear to me why the authors are limiting their validity testing to data elements and face validity for a measure that has been in use for many years. Even if this measure is the gold standard, it seems like predictive or concurrent validity would be an option.

Panel Member #4: The rationale provided for using face validity was somewhat weak. "the measure looks directly at medication orders and compares them to a gold-standard medication history, then determines whether these discrepancies are intentional or not based on the medical record +/- provider interview. *No other measure in existence looks at this process more directly than the measure proposed here."*

There was no data provided to back this up, such as agreement by a TEP or group of experts that the measure provides consistently valid scores.

Panel Member #5: The face validity was accomplished through a systematic and transparent process (see the Leapfrog worksheet and workbook for details on the process). Performance scores for the MARQUIS study are also included.

Panel Member #6: Submitter states testing at the data element level, but not able to discern the method in the submission.

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

Submission document: Testing attachment, section 2b2.3 Member #4: 2b1.3

Panel Member #1: The results shown in the table in 2b1.1 are somewhat inconsistent.

Panel Member #2: Face validity seems justified by literature. But sources of discrepancy across raters needs to be better understood and lessons incorporated into training**.**

Panel Member #4: They didn't provide results, said not applicable

Panel Member #5: The applicant indicated "Not Applicable" for this item but I have concerns that some of the medication reconciliation data may be incorrect as there is no way to tell if the med rec (even if it is a gold standard) produces correct data.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes Panel Member #3: – for data elements**

☒ **No Panel Member #1:** Insufficient.

☒ **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.**

    Panel Member #1: If PAML cannot be established consistently, then all subsequent results become questionable.

    Panel Member #2: Kappa seems adequate enough. Training methods need to be reviewed by substantive experts.

    Panel Member #3: The measure does appear to have face validity from a non-clinical, common sense perspective. That said, I feel like the validity testing was limited raising doubts in my mind about why they couldn't offer a more compelling case. As stated above, I wonder about decreasing vigilance on the part of data collectors over time.

    Panel Member #4: See comments above.

    Panel Member #5: I have concerns that some of the medication reconciliation data may be incorrect as there is no way to tell if the med rec (even if it is a gold standard) produces correct data.

    Panel Member #6: Critical element validity testing could not be located.

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.**

    Panel Member #2: Substantive committee needs to review training materials and discuss with developer what has been learned from areas where two rates disagreed and how this has been used to improve materials. Also, what are plans for continued testing of interrater reliability, or including it in protocol for hospitals to implement when using measure.

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 is collected when a trained pharmacist provides care. A gold standard list must be created then the list must be compared to admission and discharge orders. The process takes about an hour per patient on average. The minimum sample size is 20 patients per month.

- Some data elements are in defined fields, but the measure score has to be calculated. The measure specifications and materials needed to calculate the measure are available to download at no cost. If sites use Leapfrog or become part of the MARQUIS Collaborative, there are fees associated. These groups facilitate measurement and benchmark results, but neither is required to use the measure.

Questions for the Committee:

- Do the benefits of a substantive medication reconciliation (e.g., assessing the quality to the medication reconciliation, rather than only if it was performed) outweigh the considerable measurement process?
- Can the data collection strategy be widely operationalized?

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

RATIONALE:

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**Concerns about the application of this at a facility level. The intent of the measure is needed but this execution may have too many holes as well as needing additional pharmacy resources to collect the data. Also have concerns about as stated above that the data collected (minimum of 20 charts a month) may be inadequate and too easily gamed.

**none

**Feasibility is rated as moderate.

**No concerns with the feasibility

**I think the measure could be strengthen by comparing to pharmacy data rather than pharmacist to pharmacist

**No concerns, manual audit.

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**

OR

Planned use in an accountability program? ☐ **Yes** ☒ **No**

Accountability program details

- The measure is currently reported to Leapfrog by 1427 hospitals. The data is not publicly reported, but sites are given their own results with national averages of similar hospital types.
- The current goal of the measure is to drive quality improvement within hospitals, but the developer reports the measure could be considered for accountability in the future (no specific plan noted). Certain updates may need to be considered if sites were to be compared against one another for accountability purposes.

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

- Leapfrog provides guidance on data collection and tracks and explains results for each site. Members of the MARQUIS Collaborative are also provided data each month and sessions are held to assist in interpreting results.
- Feedback collected by Leapfrog includes questions about the need to have a second medication history taken to obtain a gold standard, timing of admission orders, the difference between number of additionally ordered medications and discrepancies, and auto-checking in the Leapfrog Worksheet.
- A FAQ document was created, and webinars were held to offer additional education.

Additional Feedback: N/A

Questions for the Committee:

- NQF-endorsed measures are supposed to be used for accountability after 3 years of initial endorsement and publicly reported 6 years from initial endorsement. Is the increased use across hospitals since the last review supportive of the measure's use and justification for continued endorsement?
- Have performance results been used to further the goal of high-quality, efficient healthcare?

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

RATIONALE: The measure is reported by 1427 hospitals, but it is not currently used for accountability. The Committee should consider if increased use across hospitals since initial endorsement justifies the "use" criterion. Additional information from the developer regarding the plan for use of this measure for accountability and/or public reporting would help infom the Committee's discussion.

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

- Results from two MARQUIS studies demonstrate improvement. Discrepancy rates are used to refine quality improvement.
- More sites are now using the measure. The developer hopes increased use will drive national improvement efforts.

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

- No unexpected finding reported by the developer.

Potential harms

- No potential harms reported by the developer.

Additional Feedback: N/A

***Questions for the Committee**:*

- Can the performance results be used to further the goal of high-quality, efficient healthcare?
- Are there any unintended consequences of this measure and, if so, do the benefits outweigh the risks?

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

RATIONALE:

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

4a1. Use - Accountability and Transparency

Comments:

**Concerns as above about accountability. Not used for public reporting. Concerns about accuracy of Leapfrog data given the concerns with the measure listed above.

**none

**It is disconcerting that this measure is neither intended for public reporting nor for accountability programs. The measure was first endorsed in 2014, but the measure has not been used for public reporting. NQF-endorsed measures are supposed to be used for accountability programs after 3 years of initial endorsement and publicly reported 6 years from initial endorsement.

**Leapfrog reported

**this can be a very important proxy for hospital quality

**Feedback has been obtained and incorporated.

4b1. Usability – Improvement

Comments:

**Could be used for improvement. Harm would be alternate resource allocation to accomplish this measure.

**one

**Results from two MARQUIS studies demonstrate improvement. Discrepancy rates are used to refine quality improvement. The developer hopes more facilities will use the measure, which could drive national improvement efforts. However, no dada were shown to demonstrate an increased adoption of the measure by the facilities since the last endorsement in 2014. It would have been helpful if the developer could demonstrate whether or not there has been such an observation.

**Prelim pass rating for use

**I think there is a strong liklihood of variation due to compulsiveness of the pharmacist

**No concerns.

Criterion 5: Related and Competing Measures

Related or competing measures

Related measures:

- 0097 Medication Reconciliation Post-Discharge
- 2988 Medication Reconciliation for Patients Receiving Care at Dialysis Facilities
- 0419e Documentation of Current Medications in the Medical Record
- 0553 Care for Older Adults (COA)-Medication Review
- 3317 Medication Reconciliation on Admission, and

Harmonization

- This measure is different than the other medication reconciliation/review measures since it focuses on the results of the process and goes beyond documentation.
- NQF has been engaged in an effort to further harmonize these measures and make them complementary to one another. The developer notes their willingness to be involved in these efforts.

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

5. Related and Competing

Comments:

**Several other med rec measures. Would like to hear of the progress noted in the Measure Worksheet on the other measures.

**many

**None identified.

**A number of related measures. NQF has been engaged in harmonization

**I think the measure is great and adds to the existing measure and does not complete

**None that are crucial.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 1/21/2020

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

Brief Measure Information

NQF #: 2456

Corresponding Measures:

De.2. Measure Title: Medication Reconciliation: Number of Unintentional Medication Discrepancies per Medication Per Patient

Co.1.1. Measure Steward: Brigham and Women´s Hospital

De.3. Brief Description of Measure: This measure assesses the actual quality of the medication reconciliation process by identifying errors in admission and discharge medication orders due to problems with the medication reconciliation process. The target population is any hospitalized adult patient. The time frame is the hospitalization period.

At the time of admission, the admission orders are compared to the preadmission medication list (PAML) compiled by trained pharmacist (i.e., the gold standard) to look for discrepancies and identify which discrepancies were unintentional using brief medical record review. This process is repeated at the time of discharge where the discharge medication list is compared to the PAML and medications ordered during the hospitalization.

1b.1. Developer Rationale: This measure will drive hospitals to implement interventions to truly improve their medication reconciliation processes. To date, Joint Commission requirements for medication reconciliation have led mostly to pro forma compliance, for example, checking a box saying that medication reconciliation has been performed, without knowing whether clinical care has been affected. By directly measuring error rates in medication orders, this new measure will enable hospitals to better understand where their errors are occurring and the types of errors that exist. This will enable them to implement targeted interventions that actually reduce error rates. The result will be true improvements in medication safety during transitions in care.

The rate of unintentional discrepancies per patient is unacceptably high in this country, and there is variation by site. In the six sites studied using the proposed methodology, the range was 2.78 to 4.57 discrepancies per patient (average of 3.44 per patient), thus making medication reconciliation errors the single biggest source of medication errors in the hospital (i.e., as opposed to errors in prescribing, transcribing, or administration).

Studies of medication reconciliation interventions demonstrate that improvements in important outcomes are indeed possible. In a recent systematic review conducted by our group (2), we identified 26 studies. Studies consistently demonstrated a reduction in medication discrepancies (17/17 studies), potential adverse drug events (5/6), and adverse drug events (2/2), and 2/8 studies showed a reduction in health care utilization. In the first Multi-center Medication Reconciliation Quality Improvement Study, involving 1648 patients across 5 hospitals, evidence-based interventions to improve medication reconciliation resulted in a reduction in medication discrepancies by 8% per month over baseline temporal trends (adjusted incident rate ratio 0.92, 95% CI 0.87-0.97, p=0.002), using the NQF proposed metric and methodology (technically, the measure was discrepancies per patient, using the number of medications as a model offset in the Poisson regression, which essentially is the same as discrepancies per medication per patient). In the recently completed MARQUIS2 study, involving 4947 patients across 17 hospitals, results were similar but even more robust (adjusted IRR 0.95 per month, 95% CI 0.93-0.97, p=<0.0001).

Citations for 1b.1:

1. Salanitro AH, Kripalani S, Resnic J, et al. Rationale and design of the Multicenter Medication Reconciliation Quality Improvement Study (MARQUIS). BMC health services research. 2013;13:230.

2. Mueller SK, Sponsler KC, Kripalani S, Schnipper JL. Hospital-Based Medication Reconciliation Practices: A Systematic ReviewHospital-Based Medication Reconciliation Practices. Arch Intern Med. Jun 25 2012:1-13.

3. Schnipper JL, Mixon AS, Stein J, Wetterneck TB, Kaboli P, Mueller S, Labonville S, Minahan JA, Burdick E, Orav EJ, Goldstein J, Nolido NV, Kripalani S. The effects of a multi-faceted medication reconciliation quality improvement intervention on patient safety: final results of the MARQUIS study. BMJ Qual Saf 2018; 27(12):954-964.

4. Schnipper JL, Reyes Nieva H, Mallouk M, et al. Effects of a refined evidence-based toolkit on medication reconciliation quality and safety at multiple hospitals: results of the MARQUIS2 study. Plenary, Society of Hospital Medicine Annual Meeting, National Harbor, MD.

S.4. Numerator Statement: For each sampled inpatient in the denominator, the total number of unintentional medication discrepancies in admission orders plus the total number of unintentional medication discrepancies in discharge orders.

S.6. Denominator Statement: The patient denominator is the sum of the number of medications in the gold standard medication lists plus the number of unintentionally ordered additional medications in a random sample of all adults admitted to the hospital. Our recommendation is that 25 patients are sampled per month, or approximately 1 patient per weekday.

So, for example, if among those 25 patients, there are 110 gold standard medications and 40 unintentionally ordered additional medications, and 75 unintentional discrepancies are identified, the measure outcome would be 75/150 = 0.5 discrepancies per medication per patient for that hospital for that month.

S.8. Denominator Exclusions: Patients that are discharged or expire before a gold standard medication list can be obtained.

De.1. Measure Type: Outcome

S.17. Data Source: Electronic Health Data, Electronic Health Records, Instrument-Based Data, Other, Paper Medical Records

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Sep 09, 2014 **Most Recent Endorsement Date:** Sep 09, 2014

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

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

NQF_Evidence_Form_Attachment_092017_Updated_and_Submitted_SHM.docx

1a.1 <u>For Maintenance of Endorsement:</u> 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)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2456

Measure Title: Medication Reconciliation: Number of Unintentional Medication Discrepancies per Medication Per Patient

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: Click here to enter a date

Instructions

- ***Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.***
- ***Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.***
- ***For composite performance measures:***
  - ***A separate evidence form is required for each component measure unless several components were studied together.***
  - ***If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.***
- **All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- **Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.**
- **Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.**
- **Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.**
- **Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.**
- **Efficiency: [6] evidence not required for the resource use component.**
- **For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.**
- **Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.**

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Outcome: Click here to name the health outcome

☐ Patient-reported outcome (PRO): Click here to name the 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*): Click here to name the intermediate outcome

☒ Process: Unintentional medication discrepancies are errors in inpatient admission or discharge orders due to faults in the medication reconciliation process. These errors can lead directly to patient harm.

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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.

Errors in the medication reconciliation process → Unintentional medication discrepancies in admission or discharge orders → adverse drug events → patient harm

Faulty medication reconciliation processes lead to unintentional medication discrepancies in admission and discharge orders. Some of these discrepancies are potentially harmful (i.e., potential adverse drug events), and among these, some will lead to actual adverse drug events, which by definition cause patient injury.

The intervention data described below (see systematic review #2 in section 1.a.3) clearly demonstrates that our measure is responsive to improvements in the medication reconciliation process. Moreover, it is also clearly related to more distal and relevant patient outcomes: total number of unintentional medication discrepancies tracks closely with potentially harmful medication discrepancies (a kind of potential adverse drug event (potential ADE))(1). Multiple studies have also shown a clear relationship between potential ADEs and actual ADEs (injury due to a medication), and both are similarly responsive to interventions(2).

References:

1. Tam VC, Knowles SR, Cornish PL, Fine N, Marchesano R, Etchells EE. Frequency, type and clinical importance of medication history errors at admission to hospital: a systematic review. *Cmaj.* Aug 30 2005;173(5):510-515.
2. Bates DW, Boyle DL, Vander Vliet MB, Schneider J, Leape L. Relationship between medication errors and adverse drug events. J Gen Intern Med. 1995;10(4):199-205

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.

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)

X Clinical Practice Guideline recommendation (with evidence review)

The Joint Commission National Patient Safety Goal NPSG 03.06.01

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review (1): • **Title**	**The Joint Commission** **National Patient Safety Goals Effective January 2019**

• **Author** • **Date** • **Citation, including page number** • **URL**	**Hospital Accreditation Program** **NPSG.03.06.01** **https://www.jointcommission.org/assets/1/6/NPSG_Chapter_HAP_Jan2019.pdf**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Maintain and communicate accurate patient medication information. Elements of Performance: 1. Obtain information on the medications the patient is currently taking when he or she is admitted to the hospital or is seen in an outpatient setting. This information is documented in a list or other format that is useful to those who manage medications. Note 1: Current medications include those taken at scheduled times and those taken on an as-needed basis. See the Glossary for a definition of medications. Note 2: It is often difficult to obtain complete information on current medications from a patient. A good faith effort to obtain this information from the patient and/or other sources will be considered as meeting the intent of the EP. 2. Define the types of medication information to be collected in non-24-hour settings and different patient circumstances. Note 1: Examples of non-24-hour settings include the emergency department, primary care, outpatient radiology, ambulatory surgery, and diagnostic settings. Note 2: Examples of medication information that may be collected include name, dose, route, frequency, and purpose. 3. Compare the medication information the patient brought to the hospital with the medications ordered for the patient by the hospital in order to identify and resolve discrepancies. Note: Discrepancies include omissions, duplications, contraindications, unclear information, and changes. A qualified individual, identified by the hospital, does the comparison. (See also HR.01.06.01, EP 1) 4. Provide the patient (or family as needed) with written information on the medications the patient should be taking when he or she is discharged from the hospital or at the end of an outpatient encounter (for example, name, dose, route, frequency, purpose). Note: When the only additional medications prescribed are for a short duration, the medication information the hospital provides may include only those medications. For more information about communications to other providers of care when the patient is discharged or transferred, refer to Standard PC.04.02.01. 5. Explain the importance of managing medication information to the patient when he or she is discharged from the hospital or at the end of an outpatient encounter. Note: Examples include instructing the patient to give a list to his or her primary care physician; to update the information when medications are discontinued, doses are changed, or new medications (including over-the-counter products) are added; and to

	carry medication information at all times in the event of emergency situations. (For information on patient education on medications, refer to Standards MM.06.01.03, PC.02.03.01, and PC.04.01.05.)
Source of Systematic Review (2): • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Mueller SK, Sponsler KC, Kripalani S, Schnipper JL. Hospital-Based Medication Reconciliation Practices: A Systematic Review. *Arch Intern Med.* Jun 25 2012:1-13. http://www.ncbi.nlm.nih.gov/pubmed/22733210
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Fifteen of 26 studies reported pharmacist-related [medication reconciliation] interventions, 6 evaluated IT interventions, and 5 studied other interventions. Six studies were classified as good quality. The comparison group for all the studies was usual care; no studies compared different types of interventions. Studies consistently demonstrated a reduction in medication discrepancies (17 of 17 studies). Rigorously designed studies comparing different inpatient medication reconciliation practices and their effects on clinical outcomes are scarce. Available evidence supports medication reconciliation interventions that heavily use pharmacy staff and focus on patients at high risk for adverse events.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	6 studies were classified as good quality, 5 studies were classified as fair quality, the remainder were classified as poor quality.
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	Overall recommendation not assigned a grade. The evidence for benefit of interventions was fair. Most studies were small, single-site investigations. Only ten were randomized controlled trials. Descriptions of the interventions and usual care were suboptimal. Evidence was best for interventions that heavily utilized pharmacy staff and focused on patients at high risk for adverse events.

	Several, although not all, used an outcome measure similar to the one presented here and using similar patient populations.
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	26 studies, including 13 RCTs, 10 non-randomized trials with concurrent controls, and 3 pre-post studies.
Estimates of benefit and consistency across studies	Reductions in medication discrepancies were consistent in every study that measured this as an outcome (17 of 17).
What harms were identified?	None, but most studies did not explicitly measure harms. In the two studies that measured adverse drug events, they were reduced as a result of the medication reconciliation interventions.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Larger multi-site studies have been conducted since then, demonstrating the consistent link between medication reconciliation quality improvement interventions and reductions in medication discrepancies as defined by this measure. These include MARQUIS (BMJ Qual Saf. 2018; 27(12): 954-964. PMID: 30126891), which included 1648 patients at 5 hospitals, and MARQUIS2 (Plenary, Society of Hospital Medicine Annual Meeting, National Harbor, MD), which included 4947 patients at 17 hospitals. These studies have substantially added to the evidence base linking medication reconciliation QI interventions and reductions in medication discrepancies.

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.

This measure will drive hospitals to implement interventions to truly improve their medication reconciliation processes. To date, Joint Commission requirements for medication reconciliation have led mostly to pro forma compliance, for example, checking a box saying that medication reconciliation has been performed, without knowing whether clinical care has been affected. By directly measuring error rates in medication orders, this new measure will enable hospitals to better understand where their errors are occurring and the types of errors that exist. This will enable them to implement targeted interventions that actually reduce error rates. The result will be true improvements in medication safety during transitions in care.

The rate of unintentional discrepancies per patient is unacceptably high in this country, and there is variation by site. In the six sites studied using the proposed methodology, the range was 2.78 to 4.57 discrepancies per patient (average of 3.44 per patient), thus making medication reconciliation errors the single biggest source of medication errors in the hospital (i.e., as opposed to errors in prescribing, transcribing, or administration).

Studies of medication reconciliation interventions demonstrate that improvements in important outcomes are indeed possible. In a recent systematic review conducted by our group (2), we identified 26 studies. Studies consistently demonstrated a reduction in medication discrepancies (17/17 studies), potential adverse drug events (5/6), and adverse drug events (2/2), and 2/8 studies showed a reduction in health care utilization. In the first Multi-center Medication Reconciliation Quality Improvement Study, involving 1648 patients across 5 hospitals, evidence-based interventions to improve medication reconciliation resulted in a reduction in medication discrepancies by 8% per month over baseline temporal trends (adjusted incident rate ratio 0.92, 95% CI 0.87-0.97, p=0.002), using the NQF proposed metric and methodology (technically, the measure was discrepancies per patient, using the number of medications as a model offset in the Poisson regression, which essentially is the same as discrepancies per medication per patient). In the recently completed MARQUIS2 study, involving 4947 patients across 17 hospitals, results were similar but even more robust (adjusted IRR 0.95 per month, 95% CI 0.93-0.97, p=<0.0001).

Citations for 1b.1:

1. Salanitro AH, Kripalani S, Resnic J, et al. Rationale and design of the Multicenter Medication Reconciliation Quality Improvement Study (MARQUIS). BMC health services research. 2013;13:230.

2. Mueller SK, Sponsler KC, Kripalani S, Schnipper JL. Hospital-Based Medication Reconciliation Practices: A Systematic ReviewHospital-Based Medication Reconciliation Practices. Arch Intern Med. Jun 25 2012:1-13.

3. Schnipper JL, Mixon AS, Stein J, Wetterneck TB, Kaboli P, Mueller S, Labonville S, Minahan JA, Burdick E, Orav EJ, Goldstein J, Nolido NV, Kripalani S. The effects of a multi-faceted medication reconciliation quality improvement intervention on patient safety: final results of the MARQUIS study. BMJ Qual Saf 2018; 27(12):954-964.

4. Schnipper JL, Reyes Nieva H, Mallouk M, et al. Effects of a refined evidence-based toolkit on medication reconciliation quality and safety at multiple hospitals: results of the MARQUIS2 study. Plenary, Society of Hospital Medicine Annual Meeting, National Harbor, MD.

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 Leapfrog group has the largest database on the performance of our measure, using it at hundreds of sites. Here are the results from the 1427 sites that chose to report data this past year:

Mean (SD): 0.18 (0.17)

Median (IQR): 0.14 (0.06, 0.25)

Mix-Max: 0-1.24

Deciles: 0.02, 0.05, 0.07, 0.10, 0.14, 0.18, 0.22, 0.28, 0.38, 1.24

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.

In MARQUIS2, involving 17 sites, the number of discrepancies per medication per patient ranged from 0.04 to 0.36. To put this into perspective, the maximum number of discrepancies per medication per patient is 2 (an error on admission and an error on discharge for each medication). A discrepancy rate of 0.36 therefore correlates with an 18% error rate (almost one fifth of all the errors that could be made due to the medication reconciliation process were made). Another way to state this is that if the average patient is on 10 medications (which is typical in these studies), there would be 3-4 errors in medication errors per patient. Previous studies, which were smaller, provide consistent results. Moreover, studies of interventions show improvements in discrepancy rates with medication reconciliation interventions, generally in the 42-59% range. These data clearly demonstrate opportunity for improvement in this measure.

Mueller SK, Sponsler KC, Kripalani S, Schnipper JL. Hospital-Based Medication Reconciliation Practices: A Systematic Review. Arch Intern Med. Jun 25 2012:1-13.

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.*

Perhaps paradoxically, patients over age 85 have fewer medication discrepancies, perhaps because providers are more careful with polypharmacy in these patients and pay more attention to their medication regimens. Not surprisingly, low health literacy in general, low educational attainment, and specifically poor patient understanding of their medications (dose, frequency, indication) is a major risk factor for discrepancies. Health literacy tracks with socioeconomic status. Medicaid insurance is a known risk factor for post-discharge medication non-adherence but is not as established risk factor for inpatient medication discrepancies. There is no known correlation with patient sex or race/ethnicity. The biggest risk factors have more to do with system factors and the complexity of the medication regiment rather than patient demographics.

See attached article for details:

Pippins JR, Gandhi TK, Hamann C, Ndumele CD, Labonville SA, Diedrichsen EK, Carty MG, Karson AS, Bhan I, Coley CM, Liang CL, Turchin A, McCarthy PC, Schnipper JL. Classifying and predicting errors of inpatient medication reconciliation. J Gen Intern Med. 2008;23(9):1414-1422. PMID: 18563493.

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

There are no known disparities by race, ethnicity or gender that have been reported in the literature in relation to medication reconciliation. Several studies have shown that the main predictors of discrepancies are older age and number of medications.(1-5) However, at least some evidence suggests that the very old (over 85) may actually have a lower risk of potentially harmful medication discrepancies.(2)Other risk factors for discrepancies may include low patient understanding of their medications, while having a recent medication list in the electronic medication record has been shown to be protective.(4)

Citations for 1b.5.

1. Climente-Marti M, Garcia-Manon ER, Artero-Mora A, Jimenez-Torres NV. Potential risk of medication discrepancies and reconciliation errors at admission and discharge from an inpatient medical service. Ann Pharmacother. 2010;44(11):1747-1754.

2. Pippins JR, Gandhi TK, Hamann C, et al. Classifying and predicting errors of inpatient medication reconciliation. J Gen Intern Med. 2008;23(9):1414-1422.

3. Gleason KM, McDaniel MR, Feinglass J, et al. Results of the Medications At Transitions and Clinical Handoffs (MATCH) Study: An Analysis of Medication Reconciliation Errors and Risk Factors at Hospital Admission. J Gen Intern Med. 2010.

4. Salanitro AH, Osborn CY, Schnipper JL, et al. Effect of patient- and medication-related factors on inpatient medication reconciliation errors. J Gen Intern Med. 2012; 27(8):924-932.

5. Unroe KT, Pfeiffenberger T, Riegelhaupt S, Jastrzembski J, Lokhnygina Y, Colon-Emeric C. Inpatient medication reconciliation at admission and discharge: A retrospective cohort study of age and other risk factors for medication discrepancies. Am J Geriatr Pharmacother. 2010;8(2):115-126.

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):*

Care Coordination, Person-and Family-Centered Care, Safety : Medication

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

Elderly

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.)*

We are in the process of developing a webpage.

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** MedRec_Workbook_Leapfrog_2017_Final_NQF.xlsx

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

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.

S.3.1. <u>For maintenance of endorsement:</u> 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. <u>For maintenance of endorsement,</u> please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

The number of discrepancies is now divided by the number of medications to more accurately account for the fact that discrepancies (errors) are dependent on the number of opportunities for error. For each gold standard medication or unintentionally ordered additional medication, it can be ordered incorrectly ordered at admission, at discharge, both, or neither. Therefore, the number of discrepancies per medication per patient can range from zero to two. This more fairly judges hospitals because patient populations may vary with respect to the complexity of their medication regimens.

We attempted to address concerns regarding how to reconcile the measure for patients with numerous medications versus patients with a lower number of medications in part by modifying the metric so that it is now discrepancies per medication per patient.

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.*

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

For each sampled inpatient in the denominator, the total number of unintentional medication discrepancies in admission orders plus the total number of unintentional medication discrepancies in discharge orders.

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)*

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

First, a "gold-standard" preadmission medication history is taken by one or more trained pharmacists at each site. Every site can have a trained pharmacist. We have stopped calling them study pharmacists, just trained pharmacists. Pharmacist training materials have been developed to support pharmacists (please see training materials in attachment), which specifically reviews how to take a gold standard medication history, including compliance with a best practices checklist (see attached materials). The pharmacist utilizes all available

sources of information to take the medication history, including subject and family/caregiver interviews, prescription pill bottles, outpatient electronic medical records, community pharmacy data, and prescription fill information (see Appendix A for complete protocol). The gold-standard medication history is taken within 24 hours of admission but after the medication history has been taken as part of usual care.

The resulting preadmission medication list is then compared with the medical team's documented preadmission medication list and with all admission and discharge medication orders. Any discrepancies between the gold-standard history and medication orders are identified and reasons for these changes sought from the medical record. Pharmacists may also need to communicate directly with the medical team to clarify reasons for discrepancies, as needed. Medication discrepancies that are not clearly intentional are then recorded, along with the reason for the discrepancy:

1. History discrepancies: the order is incorrect because the medical team's preadmission medication list is incorrect (e.g., the team did not know the patient was taking aspirin prior to admission, does not record it in the preadmission medication list, and therefore does not order it at admission)

2. Reconciliation discrepancies: the medical team's preadmission medication list is correct, but there is still an error in the orders. For example, the team knew the patient was taking aspirin prior to admission and documents it in the preadmission medication list. The team decides to hold the aspirin on admission for a clinical reason such as bleeding, but the team forgets to restart the aspirin at discharge. The admission discrepancy would be considered intentional (no error, not counted in the numerator), but the discharge discrepancy would be counted as a reconciliation error.

The type of error should also be recorded: omission, discrepancy in dose, route, frequency, or formulation, or an additional medication. Lastly, the time of the error should be recorded: admission vs. discharge.

See attached materials for a flow diagram explaining how history discrepancies, reconciliation discrepancies (PowerPoint slides), intentional and unintentional discrepancies are defined and operationalized.

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

The patient denominator is the sum of the number of medications in the gold standard medication lists plus the number of unintentionally ordered additional medications in a random sample of all adults admitted to the hospital. Our recommendation is that 25 patients are sampled per month, or approximately 1 patient per weekday.

So, for example, if among those 25 patients, there are 110 gold standard medications and 40 unintentionally ordered additional medications, and 75 unintentional discrepancies are identified, the measure outcome would be 75/150 = 0.5 discrepancies per medication per patient for that hospital for that month.

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).

Patients are randomly selected each day from a list of admitted patients the day before. A target number of patients are selected (e.g. one patient per weekday) and these patients are interviewed by the pharmacist.

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

Patients that are discharged or expire before a gold standard medication list can be obtained.

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.)*

Please see exclusion listed above.

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.)*

Stratification could be done by service if desired by NQF, for example: non-ICU medicine, non-ICU surgery, ICU, and other.

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:

Continuous variable, e.g. average

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.)*

See Appendix Attached (2019 Leapfrog Hospital Town Hall Call-Medication Discrepancies for NQF-Final (PowerPoint Presentation)

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.

For statistical process control charts, the recommended minimal sample size is 20 data points per time period (in this case, 20 patients per month). Beyond that, depending on several factors, additional data does not have a large impact on the SPC limits. https://www.spcforexcel.com/knowledge/control-chart-basics/how-much-data-do-i-need-calculate-control-limits

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

To minimize sampling bias, we used a method whereby the list of admitted patients from the day prior are randomized in the order in which they are approached (see attached workbook, tabs for Instructions and Sampling). Once approached, the goal is for the pharmacist to take a gold-standard medication history before the patient is discharged from the hospital unless the patient declines. This may require multiple attempts (e.g., because the patient is off the floor at a procedure or test or wants the pharmacist to return at another time). By approaching patients on the first full day after admission, this method also minimizes bias by length of stay. As noted above (2b2.2.), patients who were measured were generally older, with longer lengths of stay, and on more medications, but these differences would likely be less in a non-research setting. We know of no better way to minimize selection bias for this metric.

Once selected for measurement, there should be no missing data, as all data collection is inherent to the process of taking a gold-standard medication history and comparing it to medication orders, which should always be accessible, and reviewing the medical record to determine whether discrepancies were unintentional, which should also always be available.

While the proportion of patients excluded from the measure might vary by site (e.g., due to differences in length of stay or intensity of procedures), the populations of those included in each site should be more comparable to each other. In addition, for this measure, the more important factor is the stability of a patient population within a site over time (See notes above about tracking improvements over time, S.11).

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.

Electronic Health Data, Electronic Health Records, Instrument-Based Data, Other, Paper Medical 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.

Please see Med Rec Leapfrog Workbook Excel Attachment.

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)*

Available in attached appendix at A.1

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

Facility

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

Inpatient/Hospital

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.)*

2. Validity – See attached Measure Testing Submission Form

nqf_testing_attachment_7_27_2018_Final_Submitted_revised_08012019_Final_Resubmitted_112019.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.

No

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.

Yes - Updated information is included

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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** NQF #2456
Measure Title: Medication Reconciliation: Number of Unintentional Medication Discrepancies per Patient
Date of Submission: 8/1/2019

Type of Measure:

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

Instructions

- **Measures must be tested for all the data sources and levels of analyses that are specified. *If there is more than one set of data specifications or more than one level of analysis, contact NQF staff* about how to present all the testing information in one form.**
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures, section 2b3 also must be completed.**
- **If specified for multiple data sources/sets of specificaitons (e.g., claims and EHRs), section 2b5 also must be completed.**
- **Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). *Contact NQF staff if more pages are needed.***
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**
- **For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.**

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration**

OR

- **rationale/data support no risk adjustment/ stratification.**

No risk adjustment or risk stratification

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful [16] differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures

(e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ claims	☐ claims
☐ registry	☐ registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Medication data collected from patient/caregiver interview, ambulatory providers, community pharmacies, electronic prescription fill information and information on discrepancies in medication orders (intentional or not) from provider interviews.	☒ other: Medication data collected from patient/caregiver interview, ambulatory providers, community pharmacies, electronic prescription fill information and information on discrepancies in medication orders (intentional or not) from provider interviews.

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*).

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

December 2011 – June 2014

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.20***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

Five U.S. hospitals were selected, based on their willingness to participate in the MARQUIS study and ability to collect data and implement interventions to improve the medication reconciliation process. These included two academic medical centers, two community hospitals (one teaching, one non-teaching), and one Veterans Affairs medical center. These hospitals were geographically diverse (West coast, Northeast, Southeast, and Midwest) and varied in size from 45 to 653 beds. Two utilized an electronic health record prior to data collection, two implemented an EHR during the data collection period, and one continued to use a paper medical record supplemented by medication reconciliation software.

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

For testing the reliability of gold-standard medication histories, we included 19 randomly selected medical inpatients at one large, urban, academic medical center.

For testing the reliability of the discrepancy scoring system, we included 4 patients, one each from 4 of the MARQUIS study sites, chosen by study pharmacists because of the challenging nature of the cases.

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.

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.

Not applicable. There were no social risk factors analyzed.

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*)

To evaluate inter-rater reliability of the gold-standard medication histories, 19 randomly selected medication histories were collected independently by two study pharmacists. Among all the medications recorded for each patient, there was complete agreement in medication, dose, route, and frequency for 147 of 192 medications (77%).

To evaluate inter-rater reliability of the discrepancy scoring system, we analyzed the last 4 quarterly cases, consisting of a total of 44 medications and 128 ratings each for admission and discharge discrepancies (i.e., 256 data points). We found the following:

- For the presence of admission discrepancies, we found agreement for 116/128 ratings (91% agreement)
- For the presence of discharge discrepancies, we found agreement for 116/128 ratings (91% agreement)
- When an admission discrepancy was present (according to the gold standard), we found complete agreement for discrepancy type (i.e., omission, dose, frequency, route, formulation, additional, other) in 55/64 cases (86%), partial agreement (e.g., "dose and frequency" vs. "dose") in 4/64 cases (6%), and disagreement in 5/64 cases (8%)
- When an admission discrepancy was present, we found complete agreement for discrepancy reason (i.e., history error vs. reconciliation error) in 47/64 cases (73%).
- When a discharge discrepancy was present, we found complete agreement for discrepancy type in 46/56 cases (82%), partial agreement in 4/56 cases (7%) and disagreement in 6/56 cases (11%).
- When a discharge discrepancy was present, we found complete agreement for discrepancy reason in 45/56 cases (80%).

The kappa for the presence of admission discrepancies was 0.64 (substantial agreement) and the kappa for the presence of discharge discrepancies was also 0.64 (substantial agreement) across all raters.

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*)

In the analysis of reliability of the scoring system, kappas were statistically significant from zero (for admission discrepancies, $Z=7.29$, $p<0.0001$; for discharge discrepancies, $Z=7.34$, $p<0.0001$).

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?*)

Inter-rater reliability for the gold-standard medication history was high.

Inter-rater reliability for the presence or absence of an unintentional discrepancy (i.e., the information required to calculate a discrepancy rate per patient) was substantial. Reliability was somewhat lower for discrepancy type and reason, but these are to be used for internal QI purposes only (i.e., are not part of the measure itself). It should be noted that the cases used to derive these analyses were chosen because they were challenging for study pharmacists to evaluate – reliability for an "average" case is likely higher.

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 <u>performance measure score</u> 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.

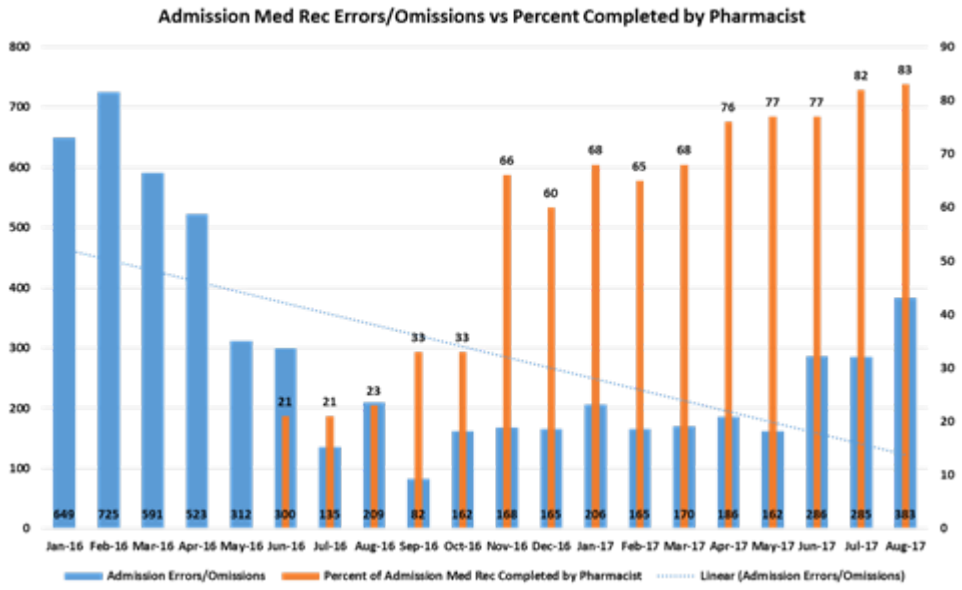
The literature supports that pharmacists take more accurate medication histories than either nurses or physicians. (1) It is therefore reasonable to assume that a specially trained pharmacist taking a preadmission medication history is a good proxy for a gold-standard medication history. Comparing errors in admission and discharge orders compared to this gold standard has high face validity as a measure of the quality of medication reconciliation. After all the goal of medication reconciliation is "to identify the most accurate list of all medications a patient is taking . . . and using this list to provide correct medications for patients anywhere within the health care system." In other words, what matters clinically is whether the orders are correct, based on knowing what medications the patient was taking previously. No other measure in existence looks at this process more directly than the measure proposed here.

The process for measuring discrepancies is systematic and transparent – see the Leapfrog worksheet and workbook for details on the process, now being used at hundreds of Leapfrog hospitals. Those who measure discrepancies are trained experts (trained pharmacists) at each site. See the attached materials for how they are trained and the above paragraph on how pharmacists can justifiably be considered medication experts. Performance scores, or at least relative improvement in scores over time, can be used to distinguish sites that truly improved from those that did not. For example, in the MARQUIS2 study, involving 17 sites, 9 sites had significant improvement in their discrepancy rates per medication per patient in the last 6 months of the study compared with the first 6 months. Compared with those sites that did not show improvement, those that did show improvement had a greater increase in the proportion of patients who received patient-level interventions (55% absolute improvement vs. 22% absolute improvement), such as a best-possible medication history taken by a dedicated trained provider while the patient was still in the Emergency Department (see below).

	First 6 months		Last 6 months		Relative Risk (95% CI)	P value	Proportion of Patients who received patient-level interventions		
	N	Discrepancies per medicati	N	Discrepancies per medicati			First 6 months	Last 6 months	Absolute improvement

		on per patient		on per patient					
Site 1	54	0.24	161	0.12	**0.50**	**<0.0001**	**24%**	**82%**	**58%**
Site 2	174	0.37	287	0.26	**0.70**	**<0.0001**	**56%**	**68%**	**13%**
Site 3	31	0.41	222	0.17	**0.41**	**<0.0001**	**17%**	**59%**	**42%**
Site 4	136	0.24	98	0.19	0.79	0.11	44%	50%	6%
Site 5	94	0.26	159	0.27	1.04	0.64	33%	45%	12%
Site 6	110	0.24	268	0.26	1.08	0.21	61%	95%	34%
Site 7	68	0.16	459	0.23	**1.44**	**0.0004**	**2%**	**61%**	**59%**
Site 8	39	0.08	149	0.04	**0.50**	**0.01**	**1%**	**91%**	**89%**
Site 9	29	0.46	319	0.32	**0.70**	**0.0001**	**1%**	**80%**	**79%**
Site 10	72	0.40	124	0.26	**0.65**	**<0.0001**	**0%**	**86%**	**86%**
Site 11	56	0.32	294	0.31	0.97	0.63	20%	35%	15%
Site 12	45	0.47	82	0.36	**0.77**	**0.003**	**37%**	**53%**	**15%**
Site 13	48	0.30	288	0.30	1.00	0.98	44%	34%	-9%
Site 14	47	0.08	100	0.11	1.38	0.17	0%	28%	28%
Site 15	84	0.34	223	0.22	**0.65**	**<0.0001**	**0%**	**100%**	**100%**
Site 16	102	0.17	225	0.14	0.82	0.053	0%	28%	28%
Site 17	40	0.15	259	0.10	**0.67**	**0.01**	**38%**	**48%**	**10%**

These results can also be seen in Leapfrog data. For example, when a particular site started having pharmacists take medication histories, their number of discrepancies went down:



References:

1. Dawson P, Gray S. Clinical significance of pharmacist-obtained drug histories. Pharm J. 1981;227:120.

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)*

Face validity: as explained above, the measure looks directly at medication orders and compares them to a gold-standard medication history, then determines whether these discrepancies are intentional or not based on the medical record +/- provider interview. No other measure in existence looks at this process more directly than the measure proposed here.

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

Not applicable.

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?*)

Not applicable.

2b2. EXCLUSIONS ANALYSIS

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

Patients may be excluded from the measure if they are discharged or expire prior to being seen by a study pharmacist, are otherwise unavailable to be seen by a pharmacist, or decline to talk to the pharmacist.

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*)

In 2006 at Brigham and Women's Hospital, we compared the 180 patients included in the analysis with the 199 excluded subjects. As expected, compared with excluded subjects, study patients were older, had longer lengths of stay, and had more medications at discharge. (1) In this study, patients were required to provide informed written consent, so the differences between included and excluded patients may have been more pronounced than if this were a routine part of hospital measurement.

References:

1. Pippins JR, Gandhi TK, Hamann C, et al. Classifying and predicting errors of inpatient medication reconciliation. J Gen Intern Med. 2008;23(9):1414-1422.

(Reprint attached – see Table 1)

This is an unusual measure, in that there really is no better measure of the quality of medication reconciliation with which is conduct a comparison. As shown in our systematic review, interventions that reduce medication discrepancies as measured using our methodology have been shown to improve more downstream patient outcomes, including potential adverse drug events, actual drug events, and (occasionally) post-discharge health care utilization.

References:

1. Mueller SK, Sponsler KC, Kripalani S, Schnipper JL. Hospital-based medication reconciliation practices: a systematic review. Arch Intern Med. 2012;172(14):1057-1069. PMID: 22733210.

To certify that pharmacists have mastery over the measurement process, "Gold standard pharmacists" receive the following in terms of training and certification:

- Review of a slide presentation and video on how to take a gold standard medication
- Completion of a pre- and post-test on the fundamentals of history-taking
- Satisfactory completion of at least one simulated case (>90% accuracy of gold standard history, >90% of best practice behaviors demonstrated)

We have created an online platform to administer these components and track completion of these requirements. The simulated cases themselves are available in the MARQUIS "train the trainer" materials: https://shm.hospitalmedicine.org/acton/media/25526/shm---bphm-train-the-trainers-material.

The training materials on being a gold standard pharmacist are here: https://shm.hospitalmedicine.org/acton/media/25526/shm-data-pharmacist-training-part-1

In addition, by definition, we require anyone taking a gold-standard history to have pharmacist-level credentials because studies have shown that pharmacists take more accurate medication histories than physicians or nurses

Lastly, as part of Leapfrog, we review the results of each site on a regular basis and talk with sites with outlier results to make sure they understand the measure and correct any misunderstanding of how to measure it.

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*)

These exclusions are simply practical ones (i.e., there is no alternative)

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

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

☐ **Other,** Click here to enter description

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.

Certain patients are at higher risk for medication discrepancies than others, but since this is an intermediate patient outcome (a measure of processes of care), risk stratification is less critical than for the evaluation of more distal outcomes such as adverse drug events. In the name of simplicity, we have chosen not to recommend risk adjustment.

However, because the number of discrepancies in medication orders is highly correlated with the total number of gold standard medications (each medication is an opportunity for error), we suggest modifying the previous metric (or adding a second metric) to be the number of unintentional medication discrepancies per medication per patient. The maximum number of this metric is two: there is a discrepancy in the admission order and a discrepancy in the discharge order for each medication.

References:

1. Pippins JR, Gandhi TK, Hamann C, et al. Classifying and predicting errors of inpatient medication reconciliation. J Gen Intern Med. 2008;23(9):1414-1422.
2. Salanitro AH, Osborn CY, Schnipper JL, Roumie CL, Labonville S, Johnson DC, Neal E, Cawthon C, Businger A, Dalal AK, Kripalani S. Effect of patient- and medication-related factors on inpatient medication reconciliation errors. J Gen Intern Med. Aug 2012;27(8):924-932.

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-Leeson 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)*

We used the two-sample T test method (1-2), using PASS power calculation software (PASS 2008, NSCC LLC, Kaysville, Utah), assuming an alpha of 0.05, beta of 0.20 (i.e., 80% power), baseline results and standard errors from the MARQUIS study, and effect sizes that are both achievable and close to the smallest sizes that would be considered clinically meaningful.

References:

1. Machin, D., Campbell, M., Fairs, P., and Pinal, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.
2. Zarf, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

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*)

Using data from MARQUIS, we can assume a baseline rate of 3.44 medication discrepancies per patient with a standard deviation of 3.62. With an alpha of 0.05, then with 6 months of pre-intervention data collection (at 25 patients per month, or 150 patients total) and 12 months of post-intervention data collection (300 patients), hospitals would have 80% power to detect a reduction from 3.44 to 2.43 discrepancies per patient.

Using data from the MARQUIS2 study using unintentional medication discrepancies per medication per patient (5022 patients at 18 hospitals), the baseline rate (i.e., among patients who did not receive interventions) was 0.62, with a standard deviation of 0.14 per site. With an alpha of 0.05, then with 6 months of pre-intervention

data collection (at 25 patients per month, or 150 patients total) and 12 months of post-intervention data collection (300 patients), hospitals would have 80% power to detect a reduction from 0.62 to 0.58 discrepancies per medication per patient, a 6.5% relative reduction, close to the smallest effect size that could be considered clinically important. This effect size was seen by all 18 of the participating MARQUIS2 sites during the study period among patients who received interventions compared with those who did not.

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?*)

With data collection on approximately 1 patient per weekday and a reasonable time-frame, hospitals will have adequate statistical power to detect effect sizes in discrepancy rates per patient that are achievable (see above for our results from the MARQUIS2 study) and close to the smallest effect size that could be considered clinically important.

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

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

Note*: This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or embrasures). 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 nonresponse's) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

To minimize sampling bias, we used a method whereby the list of admitted patients from the day prior are randomized in the order in which they are approached. Once approached, the goal is for the pharmacist to take a gold-standard medication history before the patient is discharged from the hospital unless the patient declines. This may require multiple attempts (e.g., because the patient is off the floor at a procedure or test or wants the pharmacist to return at another time). By approaching patients on the first full day after admission, this method also minimizes bias by length of stay. As noted above (**2b2.2.**), patients who were measured were generally older, with longer lengths of stay, and on more medications, but these differences would likely be less in a non-research setting. We know of no better way to minimize selection bias for this metric.

Once selected for measurement, there should be no missing data, as all data collection is inherent to the process of taking a gold-standard medication history and comparing it to medication orders, which should always be accessible, and reviewing the medical record to determine whether discrepancies were unintentional, which should also always be available.

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*)

As noted above, for the metric of number of unintentional medication discrepancies (per patient or per medication per patient), there is essentially no missing data. Reasons for discrepancies (e.g., history errors vs. reconciliation errors) was missing in X% of cases for the MARQUIS2 study.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

We believe our method minimizes selection bias to the greatest extent that is practical. As noted above, missing data is not an issue with this measure.

Addendum

Key Developments Since the Approval of the Measure

Since our metric was originally approved by NQF, there have been two major developments:

1. We completed the second Multicenter Medication Reconciliation Quality Improvement Study (MARQUIS2), which used this metric to measure the quality of medication reconciliation and response to interventions among over 5000 patients in 18 diverse hospitals, making it the largest medication reconciliation interventional study conducted to date in the U.S.

2. The Leapfrog Group adopted this metric, with the only modification, based on our advice, of dividing the total number of unintentional medication discrepancies by the total number of gold standard medications (i.e., opportunities for error). This metric has now been reported by 1,123 hospitals, including 21,347 patients, over this past year (see table below). The number of discrepancies per medication per patient was fairly stable over the last two years, with a fairly wide spread in performance among reporting hospitals. In this process of measure adoption, The Leapfrog Group has also created instructional materials and data collection tools to ease the burden of collection and reporting, held public forums, addressed questions from sites, and taken other measures to maximize consistent adoption of this metric.

Table: Preliminary Results of Leapfrog Group

Measurement Year	Number of Hospitals	Number of Patients	Discrepancies per medication per patient, median (IQR)
2017	980	12,291	0.16 (0.07-0.31)
2018	1,123	21,347	0.15 (0.06-0.28)

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**.

Some data elements are in defined fields in 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).

Data elements are in defined fields.

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. <u>Required for maintenance of endorsement.</u> 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.

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

1. When some sites started to use "intervention pharmacists" separate from "study pharmacists," sites needed to make sure that the measurement process interfered as little as possible with the intervention (e.g., by approaching patients later, not providing intervention pharmacists with information, etc.)

2. Some sites needed to work on logistics so that under most circumstances, evaluation was done after discharge orders were written but either before or not much after patient discharge. This improved access to discharge orders and the ability to contact providers in case serious errors were identified that needed to be corrected.

3. Sites have found it easier to build this evaluation into the daily work of a pharmacist. That way, if a patient is unavailable, the pharmacist can continue with their other clinical responsibilities.

4. There are efficiencies to having the same pharmacist perform admission and discharge comparisons on the same patient and do them at the same time. However, if preferable logistically, this could be a separate person from the pharmacist who collects the gold standard medication history.

5. This process takes about an hour per patient, but can take more or less depending on the patient.

6. The main barrier to data collection has been the availability of a trained pharmacist at each site. If this measure were to be endorsed by NQF, then this resource would be required for each hospital, and this problem would be solved, much in the same way that all hospitals hire study nurses to collect data for NSQIP. This has already happened at Leapfrog sites.

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)*.

There are no fees, licensing, or other requirements to use the measure as specified. If sites measure through the Leapfrog group, there are fees associated with being a member site. There are also fees associated with the MARQUIS Collaborative, which is sponsored by the Society of Hospital Medicine. Both of these can facilitate measurement and benchmark results, but neither of them are required to conduct measurement. The SHM MARQUIS web page has all the materials needed for measurement, and they can be downloaded for free.

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)
Quality Improvement (Internal to the specific organization)	

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

The Leapfrog Group currently measures discrepancy rates (as defined in this application) in 1427 sites. Currently, the scale is only 20 patients per quarter to keep the burden low on entities, but it might be increased in the future. The results are not being publicly reported, but sites are given their own results with comparison to national averages for similar hospital types (e.g., large teaching hospitals). The goal is to drive internal improvement efforts and reductions in discrepancy rates within sites. In the future, accountability might require either a certain degree of improvement in discrepancy rates over time or achievement of a certain absolute level of discrepancies per medication per patient.

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.

.

0The Leapfrog Group currently provides guidance for data collection and tracks results for each member site. While sites can be benchmarked against other sites, the main focus of interpretation is to help sites improve their own discrepancy rates over time as they engage in efforts to improve their processes, rather than compare sites against each other. As noted above, currently 1427 sites are included in this measure through Leapfrog.

MARQUIS2 involved 18 sites. We trained each site on data collection and provided data on this measure every month as part of mentored implementation. The MARQUIS Collaborative is currently enrolling sites.

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.

Leapfrog provides results quarterly along with an explanatory guide to results.

MARQUIS2 provided results monthly via phone calls with mentors, including discrepancy rates by month and the differences in discrepancy rates between those patients who did and did not receive patient-level interventions. The MARQUIS Collaborative emails results each month and holds monthly virtual "office hours" to help interpret results.

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.

Feedback from Leapfrog sites is obtained directly from member sites in 3 ways:

1. Formal 30 day public comment period each November
2. Help desk
3. Key informant interviews

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

1. Confusion about the need to have a second medication history taken in order to obtain a gold standard.
2. Questions about the timing of "admission orders" (how long after admission, etc.)
3. Questions about the difference between number of additionally ordered medications and the number of discrepancies (i.e., in orders) due to these medications.
4. Questions about auto-checking in the Leapfrog Worksheet (i.e., what it means when a number turns red, indicating a mistake in the entered data)

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

We developed an FAQ document (attached) and created a second webinar to Leapfrog member sites (attached) to help answer these questions. The measure was not changed, only the education about implementing it.

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.

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.

As noted above, improvement has clearly been demonstrated in the MARQUIS1 and MARQUIS2 studies, using monthly feedback of discrepancy rates (and lower discrepancy rates in patients who receive interventions compared to those who don't) to iteratively refine interventions and demonstrate improvement to stakeholders, leading to further spread and sustainability efforts. It is our hope that the involvement of over 1400 sites in Leapfrog's measurement program will similarly drive improvement efforts nationally.

Additionally, through the MARQUIS Collaborative, we plan to work closely with several dozen entities to drive improvement efforts, using discrepancy rates to inform refinements to improvements. This will be less intensive but more scalable than the MARQUIS studies.

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.

None

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

None

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.

No

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

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

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 other measures focus on documentation of an action related to medication reconciliation or transmission of medication data across care transitions. These are fundamentally different than measure 2456, which focuses on the results of these medication reconciliation efforts: having accurate medication orders. The fundamental problem with several of these other measures is that it is easy to "check a box" documenting that a medication reconciliation step has been completed, but it does not mean it has been completed well. In fact, there are times where these documentation efforts can be counter-productive. For example, documenting that a complete medication history has been taken, when in fact it could not be done well, could actually impede transparency among providers and efforts to fix that history the next day. Having said that, there is clearly a role for these types of measures. Further efforts are needed to harmonize these measures with each other to produce a set of complementary measures that together provide a picture of the quality of medication reconciliation. Dr. Schnipper would be happy to be involved in these efforts.

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.

Attachment **Attachment:** Measure_Maintenance_Attachments_082019_Resubmitted_112019.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Brigham and Women´s Hospital

Co.2 Point of Contact: Jeffrey, Schnipper, jschnipper@partners.org, 617-732-6201-

Co.3 Measure Developer if different from Measure Steward: Veterans Rural Health Resource Center-Central Region, VA Office of Rural Health

Co.4 Point of Contact: Peter, Kaboli, peter.kaboli@va.gov, 319-338-0581-7716

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.

• Jeffrey L. Schnipper, MD, MPH, FHM, Director of Clinical Research, BWH Hospitalist Service, Associate Physician, Division of General Medicine, Brigham and Women´s Hospital, Associate Professor of Medicine, Harvard Medical School, Boston, MA. MARQUIS Principal Investigator

Primary developer of the measure, used in several studies he has conducted

• Peter Kaboli, MD, FHM, Director, Midwest Rural Health Resource Center, VA Office of Rural Health, Iowa City VA Medical Center, Iowa City, IA. MARQUIS Co-Investigator

Expert on inpatient medication safety and roles of pharmacists. Developer of the measure with Dr. Schnipper

• Stephanie Mueller, MD, General Medicine Fellow, Division of General Medicine, Brigham and Women's Hospital, Boston, MA. MARQUIS Co-Investigator

Led systematic review of medication reconciliation interventions

• Stephanie Labonville, Pharm D. Pharmacy Services, Brigham and Women's Hospital, Boston, MA. MARQUIS Study Pharmacist

Has led implementation of measurement protocol and on-site training and evaluation of study pharmacists for MARQUIS and previous studies using this protocol led by Dr. Schnipper.

• JoAnne Resnic, MBA, BSN, RN, Former Senior Manager, Center for Hospital Innovation and Improvement, Society of Hospital Medicine, Philadelphia, PA. MARQUIS Project Manager

Project Manager for initial NQF measure application

Jenna Goldstein, MA, Directors, Center for Quality Improvement, Society of Hospital Medicine, Philadelphia, PA. MARQUIS Project Manager

Project Manager for NQF care coordination re-submission

• Peter B. Angood, MD, FRCS(C), FACS, FCCM, Senior Advisor on Patient Safety, National Quality Forum, Washington, DC. MARQUIS Steering Committee Member

Advisor to MARQUIS, especially regarding measure development

Other Advisors:

• Daniel Cobaugh, PharmD, FAACT, DABAT, Vice President, ASHP Research and Education Foundation, Bethesda, MD. MARQUIS Steering Committee Member

• Jeff Greenwald, MD, SFHM, Inpatient Clinician Educator Service, Department of Medicine, Massachusetts General Hospital and Associate Professor of Medicine, Harvard Medical School, Co-Investigator Project RED and Project BOOST, Boston, MA. Chair, MARQUIS Steering Committee

•Sunil Kripalani, MD, MSc, SFHM, Associate Professor, Chief, Section of Hospital Medicine, Associate Director, Effective Health Communication Program, Emphasis Program Area Director, Healthcare and Public Health Research and Management, Vanderbilt University Medical Center, Nashville, TN. MARQUIS Co-Investigator

• Nyryan V. Nolido, MA, Research Project Manager, Brigham and Women´s Hospital, Boston, MA. MARQUIS Data Project Manager

• Amanda Salanitro, MD, MPH, Instructor, Geriatric Research, Education and Clinical Center, Tennessee Valley VA Healthcare System and Section of Medicine at Vanderbilt University, Nashville, TN. MARQUIS Co-Investigator

• Mark Williams, MD, FACP, Professor of Medicine & Chief, Division of Hospital Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL. MARQUIS Steering Committee Member

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2012

Ad.3 Month and Year of most recent revision: 08, 2019

Ad.4 What is your frequency for review/update of this measure? The frequency of review is once every one to two years.

Ad.5 When is the next scheduled review/update for this measure? 11, 2019

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: Additional References

1. Greenwald JL, Halasyamani L, Greene J, et al. Making inpatient medication reconciliation patient centered, clinically relevant and implementable: a consensus statement on key principles and necessary first steps. J Hosp Med. 2010;5(8):477-485.