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 #: 0204

Measure Title: Skill mix (Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN], unlicensed assistive personnel [UAP], and contract)

Measure Steward: American Nurses Association

Brief Description of Measure: NSC-12.1 - Percentage of total productive nursing hours worked by RN (employee and contract) with direct patient care responsibilities by hospital unit.

NSC-12.2 - Percentage of total productive nursing hours worked by LPN/LVN (employee and contract) with direct patient care responsibilities by hospital unit.

NSC-12.3 - Percentage of total productive nursing hours worked by UAP (employee and contract) with direct patient care responsibilities by hospital unit.

NSC-12.4 - Percentage of total productive nursing hours worked by contract or agency staff (RN, LPN/LVN, and UAP) with direct patient care responsibilities by hospital unit.

Note that the skill mix of the nursing staff (NSC-12.1, NSC-12.2, and NSC-12.3) represent the proportions of total productive nursing hours by each type of nursing staff (RN, LPN/LVN, and UAP); NSC-12.4 is a separate rate.

Measure focus is structure of care quality in acute care hospital units.

Developer Rationale: Despite the consistent evidence that better nurse staffing contributes significantly to improved patient outcomes, there is considerable variations in skill mix across and within different unit types. Skill mix has been addressed in the research literature with respect to patient safety and quality of care. A lack of total nursing time and a lack of RN time for patients is thought to constrain the amount and quality of care that can be provided, as well as contribute to stress and fatigue among nursing staff. This creates a hazardous situation for patients and represents a major opportunity for use of the measures quality improvement at the patient care level and accountability (e.g.,public reporting, an identified driver of improved patient safety). The Skill Mix measures allows hospitals, including nurse administrators/managers, to assess and plan their nurse staffing and develop strategies to provide adequate skill mix on a unit-by-unit and hospital-level basis allowing comparisons with regional, state, and national staffing data.

Numerator Statement: Four separate numerators are as follows:

RN hours – Productive nursing care hours worked by RNs with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

LPN/LVN hours – Productive nursing care hours worked by LPNs/LVNs with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

UAP hours – Productive nursing care hours worked by UAP with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

Contract or agency hours – Productive nursing care hours worked by nursing staff (contract or agency staff) with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

Denominator Statement: Denominator is the total number of productive hours worked by employee or contract nursing staff with direct patient care responsibilities (RN, LPN/LVN, and UAP) for each hospital in-patient unit during the calendar month.

Denominator Exclusions: Same as numerator; nursing staff with no direct patient care responsibilities are excluded.

Measure Type: Structure

Data Source: Management Data, Other

Level of Analysis: Facility, Other

IF Endorsement Maintenance – Original Endorsement Date: Aug 05, 2009 **Most Recent Endorsement Date:** Dec 10, 2015

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Evidence Summary or **Summary of prior review in 2015**

- The developer provided a summary of the links between nurse work environment/RN hours per patient day/skill mix and job enjoyment/missed nursing care which can lead to outcomes of nurse turnover/patient safety outcomes/safety/quality of care.
- The developer provided the following systematic review: Shekelle, P.G. (2013). Chapter 34. Effect of nurse-to-patient ratios on patient morbidity and mortality. In Shekelle, P. G., Wachter, R. M.,

Pronovost, P. J., Schoelles, K., McDonald, K. M., Dy, S. M., … Winters, B. D. (Ed.), *Making Health Care Safer II: An updated critical analysis of the evidence for patient safety practices*.

- Nurse staffing ratios were defined broadly. Included staffing measures were FTEs per patient day, nurse-to-patient ratios, skill mix, and staffing in relation to adjusted targets (prevention of death, falls, pressure ulcers, and other nursing sensitive outcomes)
- The evidence concluded that nurse staffing ratios are consistently associated with reduced risk of death, however the evidence in this area cannot be rated as high because of a lack of experimental/intervention studies in which nurse staffing was deliberately altered to assess effect. The strength of evidence for the association of nurse staffing ratios was low for the other targets of falls, pressure ulcers, etc. because of a lack of data and a lack of studies showing a dose-response relationship.
- The developer summarized the Quality, Quantitiy, and Consistency of the body of evidence associated with the systematic review.

- Because of the lack of systematic reviews addressing skill mix, the developer provided an overview table (evidence of the association between nursing staff skill mix (% of hours supplied by RNs) an 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:

- The developer provided the following systematic review/meta-analysis: Driscoll, A., Grant, M. J., Carroll, D., Dalton, S., Deaton, C., Jones, I., … Astin, F. (2018). The effect of nurse-to-patient ratios on nurse-sensitive patient outcomes in acute specialist units: a systematic review and meta-analysis. *European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, *17*(1), 6–22. https://doi.org/10.1177/1474515117721561
  - Per developer, in the meta-analysis, a higher nurse staffing level decreased the risk of in-hospital mortality by 14% in intensive care unit and/or cardiac/cardiothoracic units.
  - Developer noted that grading not provided in this systematic review.

Questions for the Committee:

If the developer provided updated evidence for this measure:

- The evidence provided by the developer is updated and directionally the same compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?
- For structure, process, and intermediate outcome measures:
  - What is the relationship of this measure to patient outcomes?
  - How strong is the evidence for this relationship?
  - Is the evidence directly applicable to the process of care being measured?

Guidance from the Evidence Algorithm

Structure measure with systematic review (Box 3) → Summary of the QQC provided (Box 4) →Systematic review concludes low to moderate quality evidence (Box 5b) → Moderate

The highest possible rating is “High” for Evidence.

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

RATIONALE:

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided data of skill mix by unit type across all National Database of Nursing Quality Indicators (NDNQI) participating hopsitals that provided nurse staffing data for 2017.
    - By unit type, the mean RN skill mix ranged from 0.69 to 0.95.
    - By unit type, the mean LPN skill mix ranged from 0.002 to 0.053
    - By unit type, the mean UAP skill mix ranged from 0.04 to 0.34.
- The developer also provided skill mix data by hospital characteristics such as hospital type, bed size, teaching status, rural/metropolitan, and magnet status.

Disparities

- The developer noted limited data in the NDNQI for patient-level demographic characteristics. However, the developer was able to provide nurse staffing by Hospital-level percentage of Medicaid days.
    - The mean RN skill mix ranged from 71.44 to 72.57
    - The mean LPN skill mix ranged from 2.17 to 4.63
    - The mean UAP skill mix ranged from 25.37 to 26.78

Questions for the Committee:

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

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

RATIONALE:

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

1a. Evidence

Comments:

**updated evidence to support measure

**No

**Evidence is same as for HPPD:

- The developer provided the following systematic review/meta-analysis: Driscoll, A., Grant, M. J., Carroll, D., Dalton, S., Deaton, C., Jones, I., … Astin, F. (2018). The effect of nurse-to-patient ratios on nurse-sensitive patient outcomes in acute specialist units: a systematic review and meta-analysis. European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology, 17(1), 6–22. https://doi.org/10.1177/1474515117721561
- Per developer, in the meta-analysis, a higher nurse staffing level decreased the risk of in- hospital mortality by 14% in intensive care unit and/or cardiac/cardiothoracic units.
- Developer noted that grading not provided in this systematic review.

1b. Performance Gap

Comments:

**demonstrated gap

**performance information was provided

**The developer provided data of skill mix by unit type across all National Database of Nursing Quality Indicators (NDNQI) participating hopsitals that provided nurse staffing data for 2017.

- By unit type, the mean RN skill mix ranged from 0.69 to 0.95. o By unit type, the mean LPN skill mix ranged from 0.002 to 0.053
- By unit type, the mean UAP skill mix ranged from 0.04 to 0.34.

The developer also provided skill mix data by hospital characteristics such as hospital type, bed size, teaching status, rural/metropolitan, and magnet status.

Disparities

The developer noted limited data in the NDNQI for patient-level demographic characteristics. However, the developer was able to provide nurse staffing by Hospital-level percentage of Medicaid days.

- o The mean RN skill mix ranged from 71.44 to 72.57
- o The mean LPN skill mix ranged from 2.17 to 4.63
- o The mean UAP skill mix ranged from 25.37 to 26.78

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: Patient Safety project team staff

Evaluation of Reliability and Validity: Link A (Project Team staff)

Reliability:

The developer did updated reliability testing in the maintenance. Reliability testing was done at the performance score level and tested the stability of measures across time for nursing care hours data collected from the National Databse of Quality Indicators from January 1 2016-April 30, 2017. The data looked at 18,142

units from 1911 hospitals. The data are collected separately by nursing licensure (RN, Non-RN, Total Nursing (TN)).

The developer utilized descriptive statistics at the the unit level analysis and calculated the intraclass correlation coefficient (ICC). The developer utilized a weighted z-score procedure at the hospital level data analysis and calculated the ICC score.

Per developer, the ICC is the ratio of between level variance to the total model variance. A high ICC indicates that the amount of variance between units is greater than the variance within a given unit over time, and provides support for the reliability of the nursing care hours measures. Generally, an ICC above 0.8 indicates high reliability, and ICC above 0.6 indicates an acceptable level of reliability. Reliability at the Unit-Level and Hospital-Level were reported for Skill Mix and ranged from 0.86-0.92. (>0.8 is high reliability).

Validity:

The developer did convergent validity testing with correlation coefficients and compared Skill Mix (%RN) in the NDNQI® database with the staffing levels reported by RNs in each unit from the RN survey. Two items from the NDNQI® RN Survey ask nurses to provide information about staffing levels on their unit. Specifically these items ask how many total patients were assigned to a nurse on his or her last shift, and the maximum number of patients assigned to a nurse at any one time on his or her last shift.

At unit level, the correlation coefficients between the %RN Skill Mix and RN reported nurse staffing measures were -0.71 for RN reported maximum number of patients on last shift, and -0.69 for RN reported total number of patients on last shift, indicating "strong" convergent validity.

At the hospital level, the correlation coefficients were lower at –0.31 to -0.40 RN reported maximum number of patients on last shift and -0.30 to -0.40 for RN reported total number of patients on last shift. The developer attributes these lower coefficients at the hospital-level are likely because of unit-level variation in nurse staffing throughout a hospital and corresponding variation in survey-reported patient load and/or skill mix and the number of patients assigned on the last shift may be more conceptually different as staffing measures.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

Evaluation A: Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0204

Measure Title: Skill Mix Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN, Unlicensed Assistive Personnel [UAP], and Contract

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-** Payroll or staffing records submitted to the National Database of Nursing Quality Indicators (NDNQI®)

Level of Analysis:

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

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.*

Developer noted they have made modification to specification to provide clear descriptions such as:

- Providing clear description on float staff (hospital employees temporarily assigned to provide direct patient care for all or part of a shift on a unit other than their unit of employment) when reporting their nursing care hours in the NDNQI data collection guidelines

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

No concerns.

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**

   N/A

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

   Submission document: Testing attachment, section 2a2.2

   The developer did updated reliability testing in the maintenance. Reliability testing was done at the performance score level and tested the stability of measures across time for nursing care hours data collected from the National Databse of Quality Indicators from January 1 2016-April 30, 2017. The data looked at 18,142 units from 1911 hospitals. The data are collected separately by nursing licensure (RN, Non-RN, Total Nursing (TN)).

   The developer utilized descriptive statistics at the the unit level analysis and calculated the intraclass correlation coefficient (ICC). The developer utilized a weighted z-score procedure at the hospital level data analysis and calculated the ICC score.

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

   Submission document: Testing attachment, section 2a2.3

   Per developer, the ICC is the ratio of between level variance to the total model variance. A high ICC indicates that the amount of variance between units is greater than the variance within a given unit over time, and provides support for the reliability of the nursing care hours measures. Generally, an ICC above 0.8 indicates high reliability, and ICC above 0.6 indicates an acceptable level of reliability.

   Reliability at the Unit-Level and Hospital-Level were reported for Skill Mix and ranged from 0.86-0.92. (>0.8 is high reliability).

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

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

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

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

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

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

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

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

    Reliability methodology and results are appropriate and yielded high scores.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

The developer did not conduct exclusions analysis. However exclusions to this measure, as indicated in measure submission by developer, is nursing staff with no direct patient care responsibilities are excluded.

- Persons whose primary responsibility is administrative in nature.
- Specialty teams, patient educators, or case managers who are not assigned to a specific unit.
- Unit secretaries or clerks, monitor technicians, and other with no direct patient care responsibilities.

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

    Submission document: Testing attachment, section 2b4.

    No concerns identified. The distributions of standardized skill mix scores continues to show variation across unit type.

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

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

    Submission document: Testing attachment, section 2b6.

    The developer notes data is from NDNQI which is a voluntary reporting database which tend to be from larger hospitals and magnet/teaching hospitals. The developer noted when they looked at missing data patterns, it tended to be missing data from within a unit (i.e. a missed month of reporting) rather than across units.

    The multilevel random effects models used to estimate ICC reliability were estimated using a full information maximum likelihood (FIML) estimator. The FIML estimation will correct for any bias in the calculation of the ICC for this sporadic missing data.

16. **Risk Adjustment**

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

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

    ☒ Yes ☐ No ☐ Not applicable

    Per developer, the measure is not risk adjusted because the NDNQI does not collect patient-level data for all patients in participating hospitals. Collecting and reporting such data to the NDNQI would present a significant data collection burden for hospitals. Further, a risk adjustment model has not been established for skill mix.

    The NDNQI stratifies skill mix data by a well-established nursing unit typology. In addition, comparative reports are available based on hospital structural characteristics. The developer broke down the data/testing by this stratification.

    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:** N/A

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**

N/A

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

The developer did convergent validity testing with correlation coefficients and compared Skill Mix (%RN) in the NDNQI® database with the staffing levels reported by RNs in each unit from the RN survey. Two items from the NDNQI® RN Survey ask nurses to provide information about staffing levels on their unit. Specifically these items ask how many total patients were assigned to a nurse on his or her last shift, and the maximum number of patients assigned to a nurse at any one time on his or her last shift.

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

Submission document: Testing attachment, section 2b2.3

At unit level, the correlation coefficients between the %RN Skill Mix and RN reported nurse staffing measures were -0.71 for RN reported maximum number of patients on last shift, and -0.69 for RN reported total number of patients on last shift, indicating “strong” convergent validity.

At the hospital level, the correlation coefficients were lower at –0.31 to -0.40 RN reported maximum number of patients on last shift and -0.30 to -0.40 for RN reported total number of patients on last shift. The developer attributes these lower coefficients at the hospital-level are likely because of unit-level variation in nurse staffing throughout a hospital and corresponding variation in survey-reported patient load and/or skill mix and the number of patients assigned on the last shift may be more conceptually different as staffing measures.

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

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

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

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

NOTE that data element validation from the literature is acceptable.

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

☐ **Yes**

☐ **No**

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

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

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

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

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

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

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

At the unit level, convergent validity scores are moderate at 0.60, however, at the hospital level, the convergent validity scores were lower 0.30-0.40.

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.**

No additional concerns.

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

2a1. Reliability – Specifications

Comments:

**none

**none

**The developer did updated reliability testing in the maintenance. Reliability testing was done at the performance score level and tested the stability of measures across time for nursing care hours data collected from the National Databse of Quality Indicators from January 1 2016-April 30, 2017. The data looked at 18,142 units from 1911 hospitals. The data are collected separately by nursing licensure (RN, Non-RN, Total Nursing (TN)). The developer utilized descriptive statistics at the the unit level analysis and calculated the intraclass correlation coefficient (ICC). The developer utilized a weighted z-score procedure at the hospital level data analysis and calculated the ICC score. Per developer, the ICC is the ratio of between level variance to the total model variance. A high ICC indicates that the amount of variance between units is

greater than the variance within a given unit over time, and provides support for the reliability of the nursing care hours measures. Generally, an ICC above 0.8 indicates high reliability, and ICC above 0.6 indicates an acceptable level of reliability. Reliability at the Unit-Level and Hospital-Level were reported for Skill Mix and ranged from 0.86-0.92. (>0.8 is high reliability).

2a2. Reliability – Testing

Comments:

**acceptable

**no

**no

2b1. Validity –Testing

Comments:

**none

**no

**The developer did convergent validity testing with correlation coefficients and compared Skill Mix (%RN) in the NDNQI® database with the staffing levels reported by RNs in each unit from the RN survey. Two items from the NDNQI® RN Survey ask nurses to provide information about staffing levels on their unit. Specifically these items ask how many total patients were assigned to a nurse on his or her last shift, and the maximum number of patients assigned to a nurse at any one time on his or her last shift. At unit level, the correlation coefficients between the %RN Skill Mix and RN reported nurse staffing measures were -0.71 for RN reported maximum number of patients on last shift, and -0.69 for RN reported total number of patients on last shift, indicating "strong" convergent validity. At the hospital level, the correlation coefficients were lower at –0.31 to -0.40 RN reported maximum number of patients on last shift and -0.30 to -0.40 for RN reported total number of patients on last shift. The developer attributes these lower coefficients at the hospital-level are likely because of unit-level variation in nurse staffing throughout a hospital and corresponding variation in survey-reported patient load and/or skill mix and the number of patients assigned on the last shift may be more conceptually different as staffing measures.

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

Comments:

**none

**yes it could

** The developer did not conduct exclusions analysis. However exclusions to this measure, as indicated in measure submission by developer, is nursing staff with no direct patient care responsibilities are excluded.

- Persons whose primary responsibility is administrative in nature.
- Specialty teams, patient educators, or case managers who are not assigned to a specific unit.
- Unit secretaries or clerks, monitor technicians, and other with no direct patient care responsibilities

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

Comments:

**n/a

** acceptable results. Not sufficient risk adjustment in my view

** no concerns. Per developer, the measure is not risk adjusted because the NDNQI does not collect patient-level data for all patients in participating hospitals. Collecting and reporting such data to the NDNQI would present a significant data collection burden for hospitals. Further, a risk adjustment model has not been established for skill mix. The NDNQI stratifies skill mix data by a well-established nursing unit typology. In addition, comparative reports are available based on hospital structural characteristics. The developer broke down the data/testing by this stratification

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 Specifications and Elements

- The measure is generated from electronic payroll/accounting report or electronic staffing system
- ALL data elements are in defined fields in a combination of electronic sources
- This measure is not an eMeasure.

Data Collection Strategy

- The developer noted the mean total hours required to extract clean, and submit the staffing measures was 6.0 with a range of 1-32 hours each month.
- The developer notes some education they do to promote measure:
    - Periodically provide teleconferences for site-coordinators to educate, update some changes in data collection guidelines, and address issues about the definition of NDNQI quality indicators (e.g., nursing care hours) and data collection procedures (changes on the data entry fields and the use of data summary report and data error report to verify data before reporting to NDNQI).
    - NDNQI collects nursing care hours data through a secure NDNQI website. They provided data error messages to notify site coordinators that data on nursing care hours were not entered for all 3 months of a quarter, although it may or may not be an error.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

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

RATIONALE:

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**none

**none

** The measure is generated from electronic payroll/accounting report or electronic staffing system

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

- This measure is not an eMeasure. Questions for the Committee: Are the required data elements routinely generated and used during care delivery? Are the required data elements available in electronic form, e.g., EHR or other electronic sources? Is the data collection strategy ready to be put into operational use?

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

Public Reporting:

- Illinois: "Illinois Hospital Report Care and Consumer Guide to Health Care" through the Illinois Hospital Report Card Act; Sponsor: Illinois General Assembly & Illinois Department of Health
- Maine: "Nursing Sensitive Indicator Quality Data Set"; Sponsor: Maine Health Data Organization
- New York: "Disclosure of Quality and Surveillance Related Information"; Sponsor: New York Public Health and Health Planning Council and the Commissioner of Health
- Vermont: "Hospital Community Reports," ; Sponsor: Vermont General Assembly

Professional Certification or Recognition Program:

- The American Nurses Credentialing Center (ANCC) includes skill mix as part of their Magnet Recognition Program and Pathways to Excellence Recognition Program (ANCC, 2015).

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

- In 2018, the developer solicited feedback from all NDNQI participating hospitals (approximately 1800), and received feedback from 324. In the survey of 324 site coordinators, 69.7% reported that the staffing measures are somewhat or very important to their hospital's quality improvement program.

- In previous years, the developer have examined the feedback to consider revisions to the measures that will reduce the burden of data collection, including changes to required versus optional data elements, and improved tools for data collection. No changes have been made to the measures based on the most recent feedback.

Additional Feedback:

- Hospitals participating in the National Database of Nursing Quality Indicators receive quarterly reports, which provide unit, unit type, and hospital-level benchmarks across a number of benchmarking categories (e.g. hospital type, bed size, teaching status)

Questions for the Committee:

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

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

RATIONALE: This is a maintenance measure and is currently in a public reporting and accreditation program.

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

- Developer cited the following literature- Staggs, V. S., & He, J. (2013). Recent trends in hospital nurse staffing in the United States. The Journal of Nursing Administration, 43(7-8), 388–393.
    - A study was conducted evaluating trends in nursing care hours provided on general care (medical, surgical, and medical-surgical combined) units and critical care units among NDNQI hospitals from 2004-2011.
    - During that time period there was a statistically significant 9.7% increase in the proportion of hours provided by RNs on general care units.
    - At the same time, the proportion of nursing care hours provided by LPNs dropped by more than half on both general care units and ICUs

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

- These measures have been used in several studies to demonstrate the positive relationship between nurse staffing and improved patient outcomes. Although this is not an unexpected benefit, the developer notes it can be used to argue for improved nurse staffing in hospitals.

Potential harms There are no harms identified by the developer.

Additional Feedback: N/A

Questions for the Committee:

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

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

RATIONALE:

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

4a1. Use - Accountability and Transparency

Comments:

** already used in some reporting

**yes

** Public Reporting:

- Illinois: "Illinois Hospital Report Care and Consumer Guide to Health Care" through the Illinois Hospital Report Card Act; Sponsor: Illinois General Assembly & Illinois Department of Health
- Maine: "Nursing Sensitive Indicator Quality Data Set"; Sponsor: Maine Health Data Organization
- New York: "Disclosure of Quality and Surveillance Related Information"; Sponsor: New York Public Health and Health Planning Council and the Commissioner of Health
- Vermont: "Hospital Community Reports," ; Sponsor: Vermont General Assembly Professional Certification or Recognition Program:
- The American Nurses Credentialing Center (ANCC) includes skill mix as part of their Magnet Recognition Program and Pathways to Excellence Recognition Program (ANCC, 2015).

4b1. Usability – Improvement

Comments:

**no concerns

**measure is fine

** Program: The American Nurses Credentialing Center (ANCC) includes skill mix as part of their Magnet Recognition Program and Pathways to Excellence Recognition Program (ANCC, 2015).

Criterion 5: Related and Competing Measures

Related or competing measures
0205 : Nursing Hours per Patient Day

Harmonization
Per developer, the measure is completely harmonized with 0205, as both use the same database and definition of RN hours and are applied to the same settings. Measure 0204 is actually a ratio of the RN hours and Total Nursing Hours elements that are the numerator for the rates tested in Measure 0205.

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

5. Related and Competing

Comments:

** related measure but harmonized already

**none

** 0205 : Nursing Hours per Patient Day Harmonization Per developer, the measure is completely harmonized with 0205, as both use the same database and definition of RN hours and are applied to the

same settings. Measure 0204 is actually a ratio of the RN hours and Total Nursing Hours elements that are the numerator for the rates tested in Measure 0205

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 6/5/2019

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

Brief Measure Information

NQF #: 0204

Corresponding Measures:

De.2. Measure Title: Skill mix (Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN], unlicensed assistive personnel [UAP], and contract)

Co.1.1. Measure Steward: American Nurses Association

De.3. Brief Description of Measure: NSC-12.1 - Percentage of total productive nursing hours worked by RN (employee and contract) with direct patient care responsibilities by hospital unit.

NSC-12.2 - Percentage of total productive nursing hours worked by LPN/LVN (employee and contract) with direct patient care responsibilities by hospital unit.

NSC-12.3 - Percentage of total productive nursing hours worked by UAP (employee and contract) with direct patient care responsibilities by hospital unit.

NSC-12.4 - Percentage of total productive nursing hours worked by contract or agency staff (RN, LPN/LVN, and UAP) with direct patient care responsibilities by hospital unit.

Note that the skill mix of the nursing staff (NSC-12.1, NSC-12.2, and NSC-12.3) represent the proportions of total productive nursing hours by each type of nursing staff (RN, LPN/LVN, and UAP); NSC-12.4 is a separate rate.

Measure focus is structure of care quality in acute care hospital units.

1b.1. Developer Rationale: Despite the consistent evidence that better nurse staffing contributes significantly to improved patient outcomes, there is considerable variations in skill mix across and within different unit types. Skill mix has been addressed in the research literature with respect to patient safety and quality of care. A lack of total nursing time and a lack of RN time for patients is thought to constrain the amount and quality of care that can be provided, as well as contribute to stress and fatigue among nursing staff. This creates a hazardous situation for patients and represents a major opportunity for use of the measures quality improvement at the patient care level and accountability (e.g.,public reporting, an identified driver of improved patient safety). The Skill Mix measures allows hospitals, including nurse administrators/managers, to assess and plan their nurse staffing and develop strategies to provide adequate skill mix on a unit-by-unit and hospital-level basis allowing comparisons with regional, state, and national staffing data.

S.4. Numerator Statement: Four separate numerators are as follows:

RN hours – Productive nursing care hours worked by RNs with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

LPN/LVN hours – Productive nursing care hours worked by LPNs/LVNs with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

UAP hours – Productive nursing care hours worked by UAP with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

Contract or agency hours – Productive nursing care hours worked by nursing staff (contract or agency staff) with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

S.6. Denominator Statement: Denominator is the total number of productive hours worked by employee or contract nursing staff with direct patient care responsibilities (RN, LPN/LVN, and UAP) for each hospital in-patient unit during the calendar month.

S.8. Denominator Exclusions: Same as numerator; nursing staff with no direct patient care responsibilities are excluded.

De.1. Measure Type: Structure

S.17. Data Source: Management Data, Other

S.20. Level of Analysis: Facility, Other

IF Endorsement Maintenance – Original Endorsement Date: Aug 05, 2009 **Most Recent Endorsement Date:** Dec 10, 2015

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

0204_nqf_evidence_attachment_ver_7.1.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0204

Measure Title: Skill Mix (Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN], unlicensed assistive personnel [UAP], and contract)

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: 4/10/2015

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

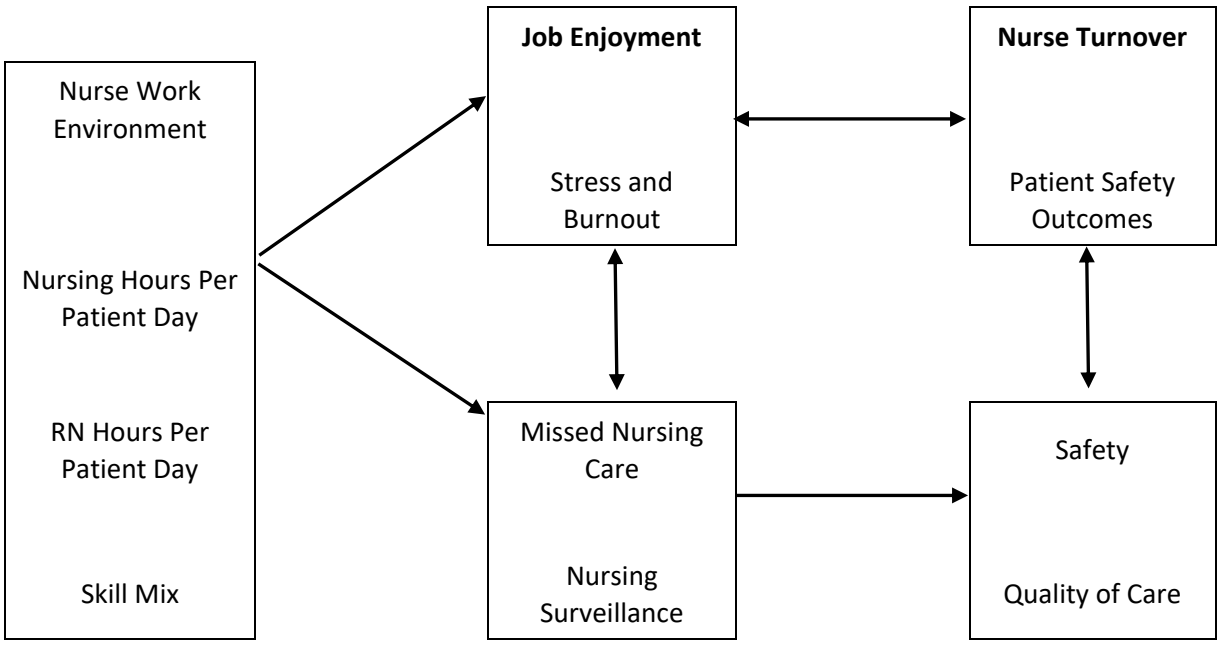
☐ Process: Click here to name what is being measured

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

☒ Structure: : Skill Mix (Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN], unlicensed assistive personnel [UAP], and contract)

☐ 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.



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.)

****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)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ 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*)

X Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Title: Effect of nurse-to-patient ratios on patient morbidity and mortality** • **Author: Shekelle, P.G.** • **Date: 2013** • **Citation, including page number: Shekelle, P.G. (2013). Chapter 34. Effect of nurse-to-patient ratios on patient morbidity and mortality. In Shekelle, P. G., Wachter, R. M., Pronovost, P. J., Schoelles, K., McDonald, K. M., Dy, S. M., … Winters, B. D. (Ed.),** ***Making Health Care Safer II: An updated critical analysis of the evidence for patient safety practices*****.** • **URL: http://www.ahrq.gov/research/findings/evidence-based-reports/ptsafetyuptp.html**
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.	Nurse staffing ratios were defined broadly. Included staffing measures were FTEs per patient day, nurse-to-patient ratios, skill mix, and staffing in relation to adjusted targets.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Nurse to patient ratios to prevent death: • Scope of the problem (frequency/severity): Common/high • Strength of the evidence for effectiveness: Moderate • Evidence of potential harm: Low • Estimate of cost: High • Implementation (how much do we know/how hard is it): A lot/Not difficult

	Nurse to patient ratios to prevent falls, pressure ulcers, and other nursing sensitive outcomes: • Scope of the problem (frequency/severity): Common/high • Strength of the evidence for effectiveness: Low • Evidence of potential harm: Low • Estimate of cost: High • Implementation (how much do we know/how hard is it): A lot/Not difficult
Provide all other grades and definitions from the evidence grading system	Defintions: **Scope of the problem:** Included the frequency of the safety problem and the severity of each average event. **Strength of evidence for effectiveness.** Included assessments evidence about context, implementation, and the use of theory or logic models, inconsistency, precision, and the possibility of reporting bias. **Evidence on potential for harmful unintended consequences.** includes actual harm and the potential for harm, ranging from high risk of harm to low (or negligible) risk of harm **Estimate of costs.** **Low cost.** Patient safety practice that does not require hiring new staff or large capital outlays **Medium cost.** Patient safety practices that might require hiring one or a few new staff members, have modest capital outlays, or incur ongoing monitoring costs. **High cost.** Patient Safety Practices that require hiring substantial numbers of new staff, have considerable capital outlays, or both. **Implementation issues.** how much we know about how to implement the PSP and how difficult it is to implement.
Grade assigned to the **recommendation** with definition of the grade	They concluded that nurse staffing ratios are consistently associated with reduced risk of death, however the evidence in this area cannot be rated as high because of a lack of experimental/intervention studies in which nurse staffing was deliberately altered to assess effect. The grade for the association of nurse staffing ratios was considered low for other outcomes because of a lack of data and a lack of studies showing a dose-response relationship.
Provide all other grades and definitions from the recommendation grading system	

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	• Quantity: 9 longitudinal studies, 1 cross-sectional • Quality: The studies as a whole lacked consistency with staffing definitions making comparisons across studies complicated.
Estimates of benefit and consistency across studies	Although the review did not examine pooled effects, the studies in general found that better nurse staffing was related to lower odds of mortality. The findings for other outcomes were more mixed.
What harms were identified?	Three studies mentioned potential harms. One study noted a decrease in the use of non-RN nursing staff. Two other studies noted increases in pressure ulcers, although this could have been due to increased detection.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	More recently, Driscoll and colleagues (2018) published a systematic literature review and meta-analysis of nurse staffing, measured as either nurse-to-patient ratios or NHPPD, and patient outcomes. While the quality of the individual studies was graded, overall grades for the evidence was not provided. In the meta-analysis, a higher nurse staffing level decreased the risk of in-hospital mortality by 14% in intensive care unit and/or cardiac/cardiothoracic units. Driscoll, A., Grant, M. J., Carroll, D., Dalton, S., Deaton, C., Jones, I., … Astin, F. (2018). The effect of nurse-to-patient ratios on nurse-sensitive patient outcomes in acute specialist units: a systematic review and meta-analysis. *European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, *17*(1), 6–22. https://doi.org/10.1177/1474515117721561

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.

Other systematic reviews on nurse staffing exist, however they do not provide graded evidence. Kane (2007) conducted a meta-analysis of studies examining the effects of nurse-to-patient ratios on patient outcomes. Increased RN staffing was associated with lower odds of mortality, hospital acquired pneumonia, unplanned extubation, respiratory failure, cardiac arrest, and failure to rescue. More recently, Driscoll and colleagues (2018) published a systematic literature review and meta-analysis of nurse staffing, measured as either nurse-

to-patient ratios or NHPPD, and patient outcomes. While the quality of the individual studies was graded, overall grades for the evidence was not provided. In the meta-analysis, a higher nurse staffing level decreased the risk of in-hospital mortality by 14% in intensive care unit and/or cardiac/cardiothoracic units.

Driscoll, A., Grant, M. J., Carroll, D., Dalton, S., Deaton, C., Jones, I., … Astin, F. (2018). The effect of nurse-to-patient ratios on nurse-sensitive patient outcomes in acute specialist units: a systematic review and meta-analysis. *European Journal of Cardiovascular Nursing: Journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, *17*(1), 6–22. https://doi.org/10.1177/1474515117721561

Kane, R. L., Shamliyan, T. A., Mueller, C., Duval, S., & Wilt, T. J. (2007). The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. *Medical Care*, *45*(12), 1195–1204. https://doi.org/10.1097/MLR.0b013e3181468ca3

Because of the lack of systematic reviews addressing skill mix, we provide an overview of the evidence below.

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

Evidence of the Association between **Nursing Staff Skill Mix (% of Hours Supplied by RNs)** and Patient Outcomes

Patient Outcome	Author (year)	Result
Falls	He et al (2016)	(+)
	Tzeng et al (2012)	(-)
	He et al (2012)	(-)
	Patrician et al (2011)	(-) Medical-surgical and intensive care units
	Dunton et al (2007)	(-)
	Dunton et al (2004)	(-)
	Cho et al (2003)	(NS)
	Blegen & Vaughn (1998)	(-)
Pressure Ulcers	He (2016)	(+)
	Choi et al (2014)	(-)
	Yang et al (2012)[a]	(NS)
	Blegen et al (2011)	(NS)
	Cho et al (2003)	(NS)
	Dunton et al (2007)	(-)
Mortality	Yang et al (2012)[a]	(NS)
	Blegen et al (2011)	(NS)
	Estabrooks et al (2005)	(-)

Failure to Rescue	Talsma et al (2014)	(NS)
Length of Stay	Pitkäaho et al (2016)	(NS)
	Esparza (2012)	(-)
	Yang et al (2012)[a]	(NS)
	Blegen et al (2011)	(NS)
	Needleman et al (2002)	(-) Medical patients only
Medication Administration Errors	Patrician et al (2011)	(-) Medical surgical and intensive care units
	Stratton (2008)	(NS)
	McGillis Hall et al (2004)	(-)
Urinary Tract Infections	Esparza (2012)	(-)
	Yang et al (2012)[a]	(-)
	Cho et al (2003)	(NS)
	Needleman et al (2002)	(-)
Pneumonia	Cho et al (2003)	(NS)
	Needleman et al (2002)	(-) Medical patients only
Bloodstream Infections	Yang et al (2012)[a]	(+)
	Stratton (2008)	(-)
Wound Infections	McGillis Hall et al (2004)	(-)
Ventilator Weaning	Yang et al (2012)[a]	(+)
Restraint Use	Staggs et al (2017)	(-)
Satisfaction	Oppel & Young (2018)	(NS)
	Martsolf (2016)	(NS)
Complaints	Stratton (2008)	(NS)
Costs	Martsolf et al (2014)	(-)

a: RN/Aide composition vs. 100% RN care

Note. (-) statistically inverse relationship between RN skill mix and patient outcomes (higher proportion of nursing hours provided by RNs is related to lower rates of the patient outcomes); (+) statistically positive relationship between RN skill mix and patient outcomes (higher proportion of nursing hours provided by RNs is related to higher rates of the patient outcomes); NS = results were not significant; RN = registered nurses.

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

Databases, including PubMed and CINAHL were used to identify research utilizing skill mix as a staffing measure in relation to patient outcomes.

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

References

Blegen, M. A., & Vaughn, T. (1998). A multisite study of nurse staffing and patient occurrences. *Nursing Economic$*, *16*(4), 196–203.

Blegen, M., Goode, C., Spetz, J., Vaughn, T., & Park, S. (2011). Nurse staffing effects on patient outcomes: safety-net and non-safety-net hospitals. *Medical Care*, *49*(4), 406–414. http://doi.org/10.1097/MLR.0b013e318202e129

Choi, J., & Staggs, V. S. (2014). Comparability of nurse staffing measures in examining the relationship between RN staffing and unit-acquired pressure ulcers: a unit-level descriptive, correlational study. *International Journal of Nursing Studies*, *51*(10), 1344–1352. http://doi.org/10.1016/j.ijnurstu.2014.02.011

Cho, S.-H., Ketefian, S., Barkauskas, V. H., & Smith, D. G. (2003). The effects of nurse staffing on adverse events, morbidity, mortality, and medical costs. *Nursing Research*, *52*(2), 71–79.

Dunton, N., Gajewski, B., Klaus, S., & Pierson, B. (2007). The relationship of nursing workforce characteristics to patient outcomes. *OJIN: The Online Journal of Issues in Nursing*, *12*(3), Manuscript 3. http://doi.org/10.3912/OJIN.Vol12No03Man03

Dunton, N., Gajewski, B., Taunton, R. L., & Moore, J. (2004). Nurse staffing and patient falls on acute care hospital units. *Nursing Outlook*, *52*(1), 53–59. http://doi.org/10.1016/j.outlook.2003.11.006

Esparza, S. J., Zoller, J. S., White, A. W., & Highfield, M. E. F. (2012). Nurse staffing and skill mix patterns: Are there differences in outcomes? *Journal of Healthcare Risk Management: The Journal of the American Society for Healthcare Risk Management*, *31*(3), 14–23. http://doi.org/10.1002/jhrm.20092

Estabrooks, C. A., Midodzi, W. K., Cummings, G. G., Ricker, K. L., & Giovannetti, P. (2005). The impact of hospital nursing characteristics on 30-day mortality. *Nursing Research*, *54*(2), 74–84.

He, J., Dunton, N., & Staggs, V. (2012). Unit-level time trends in inpatient fall rates of US hospitals. *Medical Care*, *50*(9), 801–807. http://doi.org/10.1097/MLR.0b013e31825a8b88

He, J., Staggs, V. S., Bergquist-Beringer, S., & Dunton, N. (2016). Nurse staffing and patient outcomes: a longitudinal study on trend and seasonality. *BMC Nursing*, *15*, 60. https://doi.org/10.1186/s12912-016-0181-3

Martsolf, G. R., Auerbach, D., Benevent, R., Stocks, C., Jiang, H. J., Pearson, M. L., … Gibson, T. B. (2014). Examining the value of inpatient nurse staffing: an assessment of quality and patient care costs. *Medical Care*, *52*(11), 982–988. https://doi.org/10.1097/MLR.0000000000000248

Martsolf, G. R., Gibson, T. B., Benevent, R., Jiang, H. J., Stocks, C., Ehrlich, E. D., … Auerbach, D. I. (2016). An Examination of Hospital Nurse Staffing and Patient Experience with Care: Differences between Cross-Sectional and Longitudinal Estimates. *Health Services Research*, *51*(6), 2221–2241. https://doi.org/10.1111/1475-6773.12462

McGillis Hall, L., Doran, D., & Pink, G. H. (2004). Nurse staffing models, nursing hours, and patient safety outcomes. *The Journal of Nursing Administration*, *34*(1), 41–45.

Needleman, J., Buerhaus, P., Mattke, S., Stewart, M., & Zelevinsky, K. (2002). Nurse-staffing levels and the quality of care in hospitals. *The New England Journal of Medicine*, *346*(22), 1715–1722. http://doi.org/10.1056/NEJMsa012247

Oppel, E.-M., & Young, G. J. (2018). Nurse Staffing Patterns and Patient Experience of Care: An Empirical Analysis of U.S. Hospitals. *Health Services Research*, *53*(3), 1799–1818. https://doi.org/10.1111/1475-6773.12756

Patrician, P. A., Loan, L., McCarthy, M., Fridman, M., Donaldson, N., Bingham, M., & Brosch, L. R. (2011). The association of shift-level nurse staffing with adverse patient events. *The Journal Of Nursing Administration*, *41*(2), 64–70.

Pitkäaho, T., Partanen, P., Miettinen, M. H., & Vehviläinen-Julkunen, K. (2016). The relationship between nurse staffing and length of stay in acute-care: a one-year time-series data. *Journal of Nursing Management*, *24*(5), 571–579. https://doi.org/10.1111/jonm.12359

Staggs, V. S., Olds, D. M., Cramer, E., & Shorr, R. I. (2017). Nursing Skill Mix, Nurse Staffing Level, and Physical Restraint Use in US Hospitals: a Longitudinal Study. *Journal of General Internal Medicine*, *32*(1), 35–41. https://doi.org/10.1007/s11606-016-3830-z

Stratton, K. M. (2008). Pediatric nurse staffing and quality of care in the hospital setting. *Journal of Nursing Care Quality*, *23*(2), 105–114. http://doi.org/10.1097/01.NCQ.0000313758.33654.49

Talsma, A., Jones, K., Guo, Y., Wilson, D., & Campbell, D. A. (2014). The relationship between nurse staffing and failure to rescue: where does it matter most? *Journal of Patient Safety*, *10*(3), 133–139. https://doi.org/10.1097/PTS.0b013e31829954e2

Tzeng, H.-M., Titler, M. G., Ronis, D. L., & Yin, C.-Y. (2012). The contribution of staff call light response time to fall and injurious fall rates: an exploratory study in four US hospitals using archived hospital data. *BMC Health Services Research*, *12*, 84. http://doi.org/10.1186/1472-6963-12-84

Yang, P.-H., Hung, C.-H., Chen, Y.-M., Hu, C.-Y., & Shieh, S.-L. (2012). The impact of different nursing skill mix models on patient outcomes in a respiratory care center. *Worldviews on Evidence-Based Nursing / Sigma Theta Tau International, Honor Society of Nursing*, *9*(4), 227–233. http://doi.org/10.1111/j.1741-6787.2012.00246.x

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.

Despite the consistent evidence that better nurse staffing contributes significantly to improved patient outcomes, there is considerable variations in skill mix across and within different unit types. Skill mix has been addressed in the research literature with respect to patient safety and quality of care. A lack of total nursing time and a lack of RN time for patients is thought to constrain the amount and quality of care that can be provided, as well as contribute to stress and fatigue among nursing staff. This creates a hazardous situation for patients and represents a major opportunity for use of the measures quality improvement at the patient care level and accountability (e.g.,public reporting, an identified driver of improved patient safety). The Skill Mix measures allows hospitals, including nurse administrators/managers, to assess and plan their nurse staffing and develop strategies to provide adequate skill mix on a unit-by-unit and hospital-level basis allowing comparisons with regional, state, and national staffing data.

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

The following are descriptive statistics of skill mix by unit type across all NDNQI participating hospitals that provided nurse staffing data for 2017.

Skill Mix Measures

Descriptive by Unit Type (Unit Level Measure)

RN Skill Mix	Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl	Maximum
Adult Critical Care		0.89	0.07	0.33	0.85	0.89	0.95	1.00		
Adult Step Down		0.74	0.10	0.03	0.68	0.73	0.79	1.00		
Adult Medical	0.69	0.10	0.17	0.63	0.69	0.74	1.00			
Adult Surgical	0.70	0.09	0.30	0.64	0.69	0.75	1.00			
Adult Medical-Surgical		0.68	0.10	0.02	0.63	0.68	0.74	1.00		
Neonatal	0.95	0.08	0.17	0.93	0.99	1.00	1.00			
Pediatric Critical Care		0.93	0.08	0.27	0.88	0.94	0.99	1.00		
Pediatric Medical/Surgical			0.83	0.12	0.06	0.76	0.83	0.92	1.00	
Psychiatric	0.57	0.16	0.07	0.46	0.55	0.65	1.00			
Rehabilitation	0.61	0.13	0.08	0.52	0.60	0.69	1.00			
LPN Skill Mix	Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl	Maximum
Adult Critical Care		0.002	0.016	0.000	0.000	0.000	0.000	0.675		
Adult Step Down		0.009	0.046	0.000	0.000	0.000	0.000	0.766		
Adult Medical	0.014	0.043	0.000	0.000	0.000	0.001	0.816			
Adult Surgical	0.014	0.040	0.000	0.000	0.000	0.001	0.460			
Adult Medical-Surgical		0.018	0.064	0.000	0.000	0.000	0.001	0.830		
Neonatal	0.005	0.040	0.000	0.000	0.000	0.000	0.831			
Pediatric Critical Care		0.004	0.047	0.000	0.000	0.000	0.000	0.726		
Pediatric Medical/Surgical			0.010	0.044	0.000	0.000	0.000	0.000	0.936	
Psychiatric	0.029	0.063	0.000	0.000	0.000	0.027	0.624			
Rehabilitation	0.053	0.095	0.000	0.000	0.000	0.068	0.787			
UAP Skill Mix	Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl	Maximum
Adult Critical Care		0.10	0.07	0.00	0.04	0.10	0.15	0.55		
Adult Step Down		0.25	0.09	0.00	0.20	0.26	0.31	0.97		
Adult Medical	0.30	0.10	0.00	0.25	0.31	0.36	0.82			
Adult Surgical	0.28	0.09	0.00	0.24	0.29	0.34	0.70			
Adult Medical-Surgical		0.30	0.10	0.00	0.25	0.31	0.36	0.93		
Neonatal	0.04	0.07	0.00	0.00	0.01	0.06	0.65			
Pediatric Critical Care		0.07	0.07	0.00	0.00	0.05	0.11	0.49		
Pediatric Medical/Surgical			0.16	0.11	0.00	0.05	0.16	0.24	0.84	
Psychiatric	0.19	0.20	0.00	0.00	0.12	0.37	0.91			
Rehabilitation	0.34	0.14	0.00	0.25	0.36	0.44	0.78			

Descriptives by Hospital Characteristics

RN Skill Mix

Hospital Type	Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl	Maximum
General-0.06	0.74	-4.55	-0.48	-0.06	0.33	3.24				

Pediatric	-0.25	0.67	-5.23	-0.75	-0.26	0.21	1.45				
Rehabilitation	-0.51	0.92	-2.23	-1.17	-0.59	-0.16	3.04				
Psychiatric	-0.12	1.13	-1.77	-0.95	-0.30	0.35	2.73				
Other specialty	0.59	0.73	-0.73	0.05	0.40	1.04	2.43				
LTAC	-0.03	1.09	-1.69	-1.01	-0.11	0.55	2.84				
Critical Access Hospital		0.14	0.98	-1.92	-0.66	0.04	1.01	2.70			
Oncology Specialty		0.14	0.66	-1.37	-0.31	0.08	0.58	1.50			
Orthopedic Specialty		-0.41	0.79	-1.56	-0.95	-0.53	-0.11	3.05			
Women´s Specialty		0.20	0.77	-1.14	-0.40	0.00	0.58	2.45			
Cardiac Specialty		0.04	0.58	-1.09	-0.37	-0.10	0.39	1.48			
Bed Size	Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum
<100	-0.05	0.92	-5.23	-0.63	-0.09	0.44	3.24				
100-199		-0.10	0.69	-2.73	-0.50	-0.09	0.30	2.71			
200-299		-0.08	0.67	-2.95	-0.47	-0.10	0.31	2.84			
300-399		-0.05	0.70	-2.17	-0.43	-0.04	0.33	2.82			
400-499		-0.01	0.67	-1.91	-0.38	-0.01	0.27	2.51			
>=500	-0.06	0.70	-4.55	-0.38	-0.01	0.31	2.12				
Teaching Status	Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum
Academic Medical Center		0.04	0.62	-3.06	-0.29	0.02	0.35	2.84			
Teaching Hospital		-0.12	0.75	-5.23	-0.53	-0.10	0.29	3.24			
Non-Teaching Hospital		-0.05	0.79	-3.74	-0.54	-0.09	0.37	3.24			
Location	Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum
Rural	-0.14	0.91	-2.75	-0.73	-0.26	0.30	3.24				
Metropolitan		-0.05	0.74	-5.23	-0.48	-0.06	0.34	3.24			
Micropolitan		-0.21	0.85	-3.41	-0.64	-0.16	0.27	3.24			
Magnet Status	Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum
Not Magnet Designated		-0.09	0.82	-5.23	-0.60	-0.10	0.36	3.24			
Magnet Applicant		-0.06	0.69	-4.55	-0.41	-0.04	0.31	3.24			
Magnet Designated		-0.03	0.66	-2.73	-0.41	-0.04	0.31	3.05			
LPN Skill Mix											
Hospital Type	Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum
General	0.04	0.78	-0.46	-0.27	-0.23	-0.01	14.27				
Pediatric		-0.04	0.55	-0.32	-0.20	-0.18	-0.09	4.75			
Rehabilitation		-0.01	0.94	-0.56	-0.56	-0.45	0.16	4.36			
Psychiatric		0.05	1.09	-0.46	-0.46	-0.35	-0.03	5.30			
Other specialty		0.13	1.07	-0.36	-0.36	-0.29	-0.14	3.54			
LTAC	-0.01	0.66	-0.56	-0.56	-0.56	0.60	1.52				
Critical Access Hospital		0.33	1.82	-0.46	-0.26	-0.21	0.14	10.80			
Oncology Specialty		0.00	0.79	-0.34	-0.28	-0.26	-0.21	3.22			

Orthopedic Specialty			-0.10	0.58	-0.36	-0.36	-0.35	-0.27	1.82				
Women´s Specialty			-0.12	0.12	-0.30	-0.19	-0.16	-0.12	0.21				
Cardiac Specialty			1.10	2.03	-0.30	-0.25	-0.19	1.19	5.34				
Bed Size		Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum	
<100	0.17	1.03	-0.56	-0.27	-0.22	0.17	10.80						
100-199		-0.02	0.65	-0.56	-0.28	-0.24	-0.04	7.08					
200-299		-0.01	0.58	-0.56	-0.27	-0.23	-0.03	4.72					
300-399		-0.04	0.51	-0.38	-0.27	-0.22	-0.04	6.71					
400-499		0.00	0.56	-0.46	-0.26	-0.20	0.02	4.97					
>=500	0.06	1.27	-0.37	-0.26	-0.22	-0.03	14.27						
Teaching Status		Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum	
Academic Medical Center			-0.07	0.56	-0.56	-0.27	-0.21	-0.06	5.67				
Teaching Hospital			0.01	0.78	-0.56	-0.28	-0.23	-0.02	14.27				
Non-Teaching Hospital			0.09	0.88	-0.56	-0.27	-0.23	0.04	10.80				
Location		Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum	
Rural	0.61	1.38	-0.46	-0.26	0.02	0.96	7.08						
Metropolitan		-0.01	0.73	-0.56	-0.28	-0.23	-0.03	14.27					
Micropolitan		0.42	1.17	-0.39	-0.25	-0.09	0.52	7.95					
Magnet Status		Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum	
Not Magnet Designated			0.08	0.78	-0.56	-0.27	-0.22	0.07	7.95				
Magnet Applicant			0.21	1.26	-0.56	-0.27	-0.21	0.09	14.27				
Magnet Designated			-0.12	0.49	-0.56	-0.28	-0.24	-0.12	10.80				
UAP Skill Mix													
Hospital Type		Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum	
General	0.05	0.74	-3.11	-0.33	0.08	0.49	3.29						
Pediatric		0.25	0.66	-1.38	-0.23	0.24	0.76	3.60					
Rehabilitation		0.49	1.09	-2.50	0.29	0.78	1.06	2.48					
Psychiatric		-0.62	0.53	-0.95	-0.95	-0.94	-0.41	1.14					
Other specialty		-0.62	0.68	-2.08	-1.33	-0.52	-0.14	0.89					
LTAC	0.10	0.90	-2.45	-0.59	0.23	0.86	1.57						
Critical Access Hospital			-0.37	1.16	-3.01	-1.22	-0.36	0.66	1.64				
Oncology Specialty			-0.16	0.92	-2.67	-0.46	-0.04	0.45	1.49				
Orthopedic Specialty			0.45	0.76	-2.92	0.15	0.57	0.90	1.37				
Women´s Specialty			-0.12	0.80	-2.36	-0.58	0.10	0.53	1.41				
Cardiac Specialty			-0.52	1.21	-2.59	-1.31	0.00	0.37	1.26				
Bed Size		Mean	Std Dev	Minimum		25th Pctl		50th Pctl		75th Pctl		Maximum	
<100	-0.05	0.92	-3.11	-0.56	0.02	0.58	3.60						
100-199		0.12	0.70	-2.73	-0.26	0.14	0.54	2.92					
200-299		0.08	0.68	-2.74	-0.30	0.14	0.51	2.30					

300-399	0.06	0.68	-2.69	-0.31	0.08	0.47	1.82		
400-499	0.02	0.73	-2.39	-0.34	0.05	0.47	2.03		
>=500	0.03	0.67	-2.91	-0.27	0.03	0.41	1.78		

Teaching Status	Mean	Std Dev	Minimum	25th Pctl		50th Pctl		75th Pctl	Maximum
Academic Medical Center		-0.02	0.65	-2.91	-0.33	-0.01	0.30	3.29	
Teaching Hospital		0.12	0.76	-3.11	-0.28	0.13	0.59	3.60	
Non-Teaching Hospital		0.01	0.79	-3.11	-0.41	0.10	0.52	3.01	

Location	Mean	Std Dev	Minimum	25th Pctl		50th Pctl		75th Pctl	Maximum
Rural	-0.15	0.97	-3.11	-0.62	0.03	0.45	1.98		
Metropolitan	0.05	0.76	-3.11	-0.34	0.09	0.52	3.60		
Micropolitan	0.03	0.76	-3.11	-0.42	0.08	0.56	2.70		

Magnet Status	Mean	Std Dev	Minimum	25th Pctl		50th Pctl		75th Pctl	Maximum
Not Magnet Designated		0.05	0.82	-3.11	-0.39	0.09	0.58	3.60	
Magnet Applicant		-0.06	0.71	-3.11	-0.40	0.03	0.40	1.69	
Magnet Designated		0.09	0.68	-3.01	-0.24	0.12	0.49	2.92	

National Database of Nursing Quality Indicators (NDNQI), Q1-Q3 2014 data. The NDNQI is owned by Press Ganey Associates.

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.

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.*

NDNQI has limited patient-level demographic characteristics, however, by linking our staffing data to AHA hospital survey data, we were able to examine nurse staffing by the Hospital-level percentage of Medicaid days. Because Medicaid days are a continuous variable, we categorized the hospitals by quartile for ease of presentation.

%RN Skill Mix by Hospital Percentage of Medicaid Days in Quartiles

N=880 Hospitals

	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
1st Quartile	71.44	14.87	62.91	71.74	80.41
2nd Quartile	71.65	13.81	63.99	72.19	79.77
3rd Quartile	72.57	14.97	65.02	73.66	82.34
4th Quartile	72.22	17.14	63.91	73.17	81.68

%LPN/LVN Skill Mix by Hospital Percentage of Medicaid Days in Quartiles

N=880 Hospitals

	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
1st Quartile	3.27	9.34	0.00	0.00	1.45

2nd Quartile	4.63	13.25	0.00	0.07	3.10
3rd Quartile	2.95	6.82	0.00	0.18	2.59
4th Quartile	2.17	6.51	0.00	0.02	1.77

%UAP Skill Mix by Hospital Percentage of Medicaid Days in Quartiles

N=880 Hospitals

	Mean	Standard Deviation		25th Percentile	50th Percentile	75th Percentile
1st Quartile	26.78	14.22	18.65	27.44	35.83	
2nd Quartile	25.37	12.81	16.87	27.03	32.78	
3rd Quartile	25.59	13.08	17.25	25.60	33.96	
4th Quartile	26.46	15.22	17.13	26.47	35.16	

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

N/A

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):*

Person-and Family-Centered Care : Workforce

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

Children, Populations at Risk

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.)*

None

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** Codebook_staffing-636542899570465618.pdf

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

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

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

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

No

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

Additional nursing unit types have been tested at both the unit level, and in the hospital composite measure. Additionally, a new version of the hospital-level measure that includes only medical, surgical, and medical-surgical combined units in the calculation of the composite was tested.

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

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

Four separate numerators are as follows:

RN hours – Productive nursing care hours worked by RNs with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

LPN/LVN hours – Productive nursing care hours worked by LPNs/LVNs with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

UAP hours – Productive nursing care hours worked by UAP with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

Contract or agency hours – Productive nursing care hours worked by nursing staff (contract or agency staff) with direct patient care responsibilities for each hospital in-patient unit during the calendar month.

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

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

Nursing care hours are defined as the number of productive hours worked by nursing staff (registered nurse [RN], licensed vocational/practical nurse [LVN/LPN], and unlicensed assistive personnel [UAP]) assigned to the unit who have direct patient care responsibilities for greater than 50% of their shift.

Productive hours are actual direct patient care hours worked by nursing staff including overtime, not budgeted or scheduled hours. Vacation, sick time, orientation, education leave, or committee time are considered non-productive hours. However, orientation programs vary from hospital to hospital. Once orientees reach the point where they are considered part of the staffing matrix, their work hours are charged to the unit and they would be replaced if they call in sick, then their hours are counted as productive.

Direct patient care responsibilities: Patient centered nursing activities by unit-based staff in the presence of the patient and activities that occur away from the patient that are patient related:

- Medication administration
- Nursing treatments

• Nursing rounds

• Admission, transfer, discharge activities

• Patient teaching

• Patient communication

• Coordination of patient care

• Documentation time

• Treatment planning

• Patient screening (e.g. risk) and assessment

Nursing staff included are either staff employed by the facility or temporary staff who are not employed by the facility (contracted/agency staff). Float staff—those are assigned to a unit other than their unit of employment on an as-needed basis—must be counted and reported in the unit's total nursing care hours where they provided direct patient care.

Included nursing staff:

Staff who are counted in the unit's staffing matrix, and

Are replaced if they call in sick, and

Work hours are charged to the unit's cost center

Excluded nursing staff:

1)Persons whose primary responsibility is administrative in nature

2)Specialty teams, patient educators, or case managers who are not assigned to a specific unit

3)Unit secretaries or clerks, monitor technicians, and other with no direct patient care responsibilities (Therapy assistants, student nurses who are fulfilling educational requirements, sitters who either are not employed by the facility or who are employed by the facility, but are not providing typical UAP activities)

Unlicensed Assistive Personnel (UAPs): Individuals trained to function in an assistive role to nurses in the provision of patient care, as delegated by and under the supervision of the registered nurse. Typical activities performed by UAPs may include (but are not limited to): taking vital signs, bathing, feeding, or dressing patients, assisting patients with transfers, ambulation or toileting.

Included UAPs: nursing assistants, orderlies, patient care technicians/assistants, graduate nurses (not yet licensed) who have completed unit orientation.

Mental Health Technicians (MHT): For Psychiatric In-Patient Units ONLY

Individuals functioning in an assistive role, for which your facility requires course work or training that is different from UAP. They may be licensed or unlicensed. MHT hours are included in UAP hours when reporting, but their hours are collected separately from UAP hours if persons in this job position also meet the following criteria:

• They are engaged in direct care activities greater than 50% time, and

• Their position is staffed 24/7 and replaced when they call in sick, and

• Their hours are included in the nursing staff budget

Data Elements:

RN hours (Employee)

RN hours (Contract/Agency)

LPN/LVN hours (Employee)

LPN/LVN hours (Contract/Agency)

UAP hours (Employee)

UAP hours (Contract/Agency)

MHT hours (Employee)

MHT hours (Contract/Agency)

Year

Month

Type of Unit

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

Denominator is the total number of productive hours worked by employee or contract nursing staff with direct patient care responsibilities (RN, LPN/LVN, and UAP) for each hospital in-patient unit during the calendar 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).

Same as numerator; Total number of productive hours worked by nursing staff with direct patient care responsibilities for each in-patient unit is obtained by summing all number of productive hours worked by specific nursing staff with direct patient care responsibilities (RN, LPN/LVN, or UAP) for each hospital in-patient unit during the calendar month.

Nursing staff included are either staff employed by the facility or temporary staff who are not employed by the facility (contracted/agency staff). Float staff—those are assigned to a unit other than their unit of employment on an as-needed basis—must be counted and reported in the unit's total nursing care hours where they provided direct patient care.

Included nursing staff:

Staff who are counted in the unit's staffing matrix, and

Are replaced if they call in sick, and

Work hours are charged to the unit's cost center.

Excluded nursing staff:

1)Persons whose primary responsibility is administrative in nature

2)Specialty teams, patient educators, or case managers who are not assigned to a specific unit

3)Unit secretaries or clerks, monitor technicians, and other with no direct patient care responsibilities

Data Elements:

RN hours (Employee)

RN hours (Contract/Agency)

LPN/LVN hours (Employee)

LPN/LVN hours (Contract/Agency)

UAP hours (Employee)

UAP hours (Contract/Agency)

MHT hours (Employee)

MHT hours (Contract/Agency)

Month

Year

Type of Unit

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

Same as numerator; nursing staff with no direct patient care responsibilities are excluded.

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.)*

Excluded nursing staff:

Persons whose primary responsibility is administrative in nature.

Specialty teams, patient educators, or case managers who are not assigned to a specific unit.

Unit secretaries or clerks, monitor technicians, and other with no direct patient care responsibilities.

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 variables are patient population and unit type. Units are stratified by patient population first and then unit type based on acuity level, age, or type of service provided.

1. Patient population

1) Adult population: limited to units generally caring for patients over 16 years old.

2) Pediatric population: limited to units generally caring for patients under 18 years old.

3) Neonate population: limited to units caring for newborn infants.

4) Psychiatric population: units caring for patients with psychiatric disorders.

5) Rehabilitation population: limited to distinct acute rehabilitation units providing intensive therapy 5 days/week.

2. Unit types by population

1) Adult population

Critical Care

Highest level of care, includes all types of intensive care units. Optional specialty designations include: Burn, Cardiothoracic, Coronary Care, Medical, Neurology, Pulmonary, Surgical and Trauma.

Step-Down

Limited to units that provide care for patients requiring a lower level of care than critical care units and higher level of care than provided on medical/surgical units. Examples include progressive care or intermediate care units. Telemetry alone is not an indicator of acuity level.

Medical

Units that care for patients admitted to medical services, such as internal medicine, family practice, or cardiology. Optional specialty designations include: BMT (Bone Marrow Transplant), Cardiac, GI, Infectious Disease, Neurology, Oncology, Renal or Respiratory.

Surgical

Units that care for patients admitted to surgical services, such as general surgery, neurosurgery, or orthopedics. Optional specialty designations include: Bariatric, Cardiothoracic, Gynecology, Neurosurgery, Orthopedic, Plastic Surgery, Transplant or Trauma.

Medical-Surgical Combined

Units that care for patients admitted to either medical or surgical services. Optional specialty designations include: Cardiac, Neuro/Neurosurgery or Oncology.

Critical Access

A unit located in a Critical Access Hospital that cares for a combination of patients that may include critical care, medical-surgical, skilled nursing (swing bed) and/or obstetrics.

2) Pediatric population

Refer to Adult unit type descriptions for corresponding unit types.

Critical care

Step-Down

Medical

Surgical

Medical-Surgical Combined

3) Neonate population

The three unit types below (Level I, II, and III/IV) are based on the Guidelines for Perinatal Care, 5th Ed., which are used by state certification programs. Level I, II, and III/IV neonatal units are the highest level of infant care provided, and are specified by sequential level of acuity.

Well-baby Nursery

Level I Continuing Care

Level II Intermediate Care

Level III/IV Critical Care

4) Psychiatric population

Adult

Units caring for adult patients with acute psychiatric disorders.

Child/Adolescent

Units caring for children and/or adolescents, predominantly ages 2-18 years old, with acute psychiatric disorders.

Geripsych

Units caring for elderly patients with acute psychiatric disorders.

Other (Behavioral Health, Specialty, Multiple Psychiatric Unit Types)

Behavioral Health

Units caring for individuals of any age with eating disorders or substance abuse (alcohol and drugs) diagnoses.

Specialty

Units caring for patients of any age with dual diagnoses (e.g., mental illness and mental retardation, or substance abuse and an additional mental illness diagnosis).

Multiple Psychiatric Unit Types

Units caring for patients that encompass 3 or more of the above unit types, but for which no one unit type comprises greater than 50% of the entire unit.

5) Rehabilitation population

Adult

Limited to units generally caring for rehab patients over 16 years old. Optional specialty designations include: Brain Injury/SCI, Cardiopulmonary, Neuro/Stroke and Orthopedic/Amputee Rehab units.

Pediatric

Limited to units generally caring for rehab patients under 18 years old.

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

Other

If other: Each unit is stratified by unit type (e.g., critical care, step down, medical), which is not identical to risk, but may be related.

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher 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.)*

Eligible unit identified and selected; input nursing care hours for each eligible staff category by month; then perform calculations to produce the quarterly nursing care hours for each eligible staff category by summing monthly values of the 3 months; then calculate the total nursing care hours by summing quarterly nursing care hours for each eligible staff category; then divide the quarterly nursing care hours for each eligible staff category by the total quarterly nursing care hours.

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Management Data, Other

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.

Database: National Database of Nursing Quality Indicators(R) [NDNQI(R)]; Hospitals have NDNQI guidelines and Excel spreadsheets to guide data collection; data are provided to NDNQI via web based data entry or XML upload.

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, Other

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

0204_Measure_Testing_ver_7.1_v3.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

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

Measure Number (*if previously endorsed*)**:** 0204

Measure Title: Skill Mix Registered Nurse [RN], Licensed Vocational/Practical Nurse [LVN/LPN, Unlicensed Assistive Personnel [UAP], and Contract

Date of Submission: 1/7/2019

Type of Measure:

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

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 records	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒other: Payroll or staffing records	☒other: Payroll or staffing records submitted to the National Database of Nursing Quality Indicators (NDNQI®)

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).

National Database of Nursing Quality Indicators® (NDNQI®)

1.3. What are the dates of the data used in testing? January 1, 2016 – April 30, 2017

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: Clinical Nursing Unit	☒ other: Clinical Nursing Unit

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*)

18,142 units from 1,911 hospitals

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*)

None/Not applicable

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; not patient data

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*)

Design

To test reliability, we conducted a secondary analysis of longitudinal data. Two NDNQI® data sets were used, including the quarterly reported NDNQI® Nursing Care Hour data. The nursing care hours data were collected from NDNQI® member hospitals at the unit level. These data were reported quarterly but collected on a monthly-basis. NDNQI® defines nursing care hours as the number of productive hours worked by nursing staff assigned to the unit who have direct patient care responsibilities for greater than 50% of their shift. The NDNQI® nursing care hours data are collected separately by nursing licensure (Registered Nurse (RN), Non-RN, Total Nursing (TN)).

For Skill Mix, the following definitions are used:
RN Hours: Productive nursing care hours worked by RNs with direct patient care responsibilities for each hospital inpatient unit during the calendar month.

All measures were calculated on a monthly-basis for each unit. To address the fluctuation of nursing care hours resulting from changes in patient days each month, we adjusted for patient days for each staffing measure. For example, RN Skill Mix was analyzed as RNHPPD/Total NHPPD. The patient days data also were

collected from units in member hospitals on a monthly-basis and reported to NDNQI® quarterly. We included units from NDNQI® member hospitals that submitted nursing care hours data on each month from January 2016 to April 2017. We limited units to the following types: critical care, step-down, medical, surgical, medical-surgical combined, and rehabilitation. These types of units are common in most general hospitals. The application of these inclusion/exclusion criteria resulted in 18,142 units from 1,911 hospitals.

Unit-level Data Analysis

Characteristics of the participating hospitals and units were assessed using descriptive statistics. Trends in nursing care hours (with and without adjusting for patient days) by nursing licensure category were examined over the study period using descriptive statistics. The reliability of the skill mix measure (adjusted for patient days) was evaluated by assessing the consistency of reporting over time, using monthly data from 2016 January-2017 April. Critical care, step-down, medical, surgical, medical-surgical combined and rehab units were included in the analysis. Only units reporting in all 24 months were included in the analysis. The staffing measures for each unit can be adjusted for unit type by subtracting the average nursing care hours for units of that type and then dividing by the standard deviation of the nursing care hours for units of that type. The resulting z-score is the difference, in standard deviations, of the unit's nursing care hours from the average nursing care hours for units of that type.

Random intercept linear mixed models, with repeated measures nested within units were estimated for each staffing measure by unit type. The intraclass correlation coefficient (ICC) was calculated as:

$$\sigma_{ij}^2 / \sigma_{ij}^2 + \sigma_i^2,$$

where, σ_(ij)^2 is the between level variance, and σ_(i)^2, is the within-level variance. The ICC is the ratio of between level variance to the total model variance. A high ICC indicates that the amount of variance between units is greater than the variance within a given unit over time, and provides support for the reliability of the nursing care hours measures. Generally, an ICC above 0.8 indicates high reliability, and ICC above 0.6 indicates an acceptable level of reliability.

Hospital-level Data Analysis

Aggregated hospital-level scores are calculated from the unit-level staffing measures using a weighted z-score procedure. These z-scores, which are all on the same metric, can be averaged for each hospital to yield an average unit z-score:

$$(z_1 + z_2 + \ldots + z_M)/M,$$

where z_j is the z-score for the *j*th unit. The average unit z-score does not account for differences in unit size. Therefore, a weighted average of unit z-scores can be computed by weighting each unit's z-score by a measure of patient volume (patient days), summing these weighted scores, and dividing by the total number of patient days for the hospital:

$$(z_1n_1 + z_2n_2 + \ldots + z_Mn_M)/(n_1 + n_2 + \ldots + n_M).$$

Two hospital level scores were tested: one including all the unit types represented in the unit level analysis, and a second using only adult medical, surgical, and medical-surgical combined unit types.

Random intercept linear mixed models, with repeated measures nested within units were estimated for each staffing measure by unit type. The intraclass correlation coefficient (ICC) was calculated as:

$$\sigma_{ij}^2 / \sigma_{ij}^2 + \sigma_i^2,$$

where, σ_(ij)^2 is the between level variance, and σ_(i)^2, is the within-level variance. The ICC is the ratio of between level variance to the total model variance. A high ICC indicates that the amount of variance between units is greater than the variance within a given unit over time, and provides support for the reliability of the nursing care hours measures. Generally, an ICC above 0.8 indicates high reliability, and ICC above 0.6 indicates an acceptable level of reliability.

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*)

ICC Reliability Results at the Unit-level and Hospital-level

	ICC							
	Critical Care	Step-down	Medical	Surgical	Med-surg	Rehab	Hospital (MS)	Hospital (All)
RN Skill Mix (% RN hours)	0.88	0.86	0.88	0.86	0.86	0.88	0.92	0.87

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?*)

Unit-level Reliability

The results of the mixed model analyses are presented above. The ICCs for patient-day-adjusted skill mix measures ranged from 0.86 to 0.88 across unit types. The ICCs across unit type were remarkable similar. Recommendations for acceptable levels of reliability differ across sources. Center for Medicare and Medicaid Services (CMS) has suggested value above 0.4 indicate an acceptable level of reliability (CMS, 2012), while others suggest thresholds from 0.7 to 0.9 in more conservative estimates (Adams et al, 2010).

Hospital-level Reliability

The measures also demonstrated high reliability at both the medical-surgical hospital composite (ICC = 0.74) and the all units hospital-level, with an ICC of 0.87. This is well within the recommended thresholds stabled by CMS and others.

Adams, J. L., Mehrotra, A., Thomas, J. W., & McGlynn, E. A. (2010). Physician cost profiling – reliability and risk of misclassification. *New England Journal of Medicine*, *362*(11), 1014-1021.

Centers for Medicare & Medicaid Services. (2012). Memorandum: Results of reliability analysis from Mathematica Policy Research. Baltimore, MD: U.S. Department of Health & Human Services. Available at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf

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

X Empirical validity testing

☒**Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The NDNQI® RN survey was used in assessing convergent validity. We compared the Skill Mix (%RN) in the NDNQI® database with the staffing levels reported by RNs in each unit from the RN survey. Two items from the NDNQI® RN Survey ask nurses to provide information about staffing levels on their unit. Specifically these items ask how many total patients were assigned to a nurse on his or her last shift, and the maximum number of patients assigned to a nurse at any one time on his or her last shift. In previous studies, these measures have been shown to have very high correlations with NDNQI®'s quarterly reported NCH measures (Choi & Staggs, 2014) and have demonstrated reliability in predicting patient outcomes (Ma, McHugh, & Aiken, 2015; McHugh & Ma, 2013). Individual RN responses in a unit were averaged to reflect the unit RN staffing levels. The quarterly reported skill mix measure was annualized by averaging the monthly skill mix (% RN), and matched to the annual survey data for each unit. Approximately half of all NDNQI® hospitals participate in the RN survey, and the total number of units that submitted both staffing and RN survey data was 4,372 for the six unit types included in this study.

Choi, J., & Staggs, V. S. (2014). Comparability of nurse staffing measures in examining the relationship between RN staffing and unit-acquired pressure ulcers: a unit-level descriptive, correlational study. *International Journal of Nursing Studies*, *51*(10), 1344–1352. doi:10.1016/j.ijnurstu.2014.02.011

Ma, C., McHugh, M. D., & Aiken, L. H. (2015). Organization of Hospital Nursing and 30-Day Readmissions in Medicare Patients Undergoing Surgery. *Medical Care*, *53*(1), 65–70. doi:10.1097/MLR.0000000000000258

McHugh, M. D., & Ma, C. (2013). Hospital nursing and 30-day readmissions among medicare patients with heart failure, acute myocardial infarction, and pneumonia. *Med Care, 51*(1), 52-59. doi: 10.1097/MLR.0b013e3182763284

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

Unit-level Correlations among Clinical Staffing (Administrative) Measures and RN Reports of Staffing

	1	2	3	4	5	6
1. RN Hours	1.00					
2. Non-RN Hours	-0.29	1.00				

3. Total Hours	0.94	0.04	1.00			
4. RN Skill Mix	0.71	-0.78	0.47	1.00		
5. RN Report Max Patients	-0.81	0.29	-0.76	-0.71	1.00	
6. RN Report Total Patients	-0.79	0.27	-0.74	-0.69	0.98	1.00

Medical-Surgical Unit Hospital-Level Composite Correlations among Clinical Staffing (Administrative) Measures and RN Reports of Staffing

	1	2	3	4	5
1. RN Hours	1.00				
2. Non-RN Hours	0.82	1.00			
3. RN Skill Mix	0.46	-0.11	1.00		
4. RN Report Max Patients	-0.63	-0.44	-0.40	1.00	
5. RN Report Total Patients	-0.58	-0.39	-0.40	0.94	1.00

Hospital-level Correlations among Clinical Staffing (Administrative) Measure and RN Report of Staffing

	1	2	3	4	5
1. RN Hours	1.00				
2. Total Hours	0.86	1.00			
3. Skill Mix	0.45	-0.04[n.s.]	1.00		
4. RN Report Max Patients	-0.49	-0.38	-0.31	1.00	
5. RN Report Total Patients	-0.50	-0.39	-0.30	0.97	1.00

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?*)

Unit-level Interpretation of Validity

The correlation coefficients between the %RN Skill Mix and RN reported nurse staffing measures were -0.71 for RN reported maximum number of patients on last shift, and -0.69 for RN reported total number of patients on last shift, indicating strong convergent validity.

Hospital-level Interpretation of Validity

For the medical-surgical only hospital composite the correlation coefficients between the %RN Skill Mix measure and RN reported nurse staffing measures were -0.40 for RN reported maximum number of patients on last shift, and -0.40 for RN reported total number of patients on last shift. For the hospital composite including all six unit types, the correlation coefficients between the %RN Skill Mix measure and RN reported nurse staffing measures were -0.31 for RN reported maximum number of patients on last shift, and -0.30 for RN reported total number of patients on last shift. Although the correlation coefficients are lower than at the unit-level, the hospital-level coefficients are statistically significant. The lower coefficients at the hospital-level are likely because of unit-level variation in nurse staffing throughout a hospital and corresponding variation in

survey-reported patient load. Further, skill mix and the number of patients assigned on the last shift may be more conceptually different as staffing measures. However, we feel that the correlation coefficients at the hospital-level indicate acceptable validity.

2b2. EXCLUSIONS ANALYSIS

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

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section .

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 6 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.

The measure is not risk adjusted because the NDNQI does not collect patient-level data for all patients in participating hospitals. Collecting and reporting such data to the NDNQI would present a significant data collection burden for hospitals. Further, a risk adjustment model has not been established for skill mix. A variety of risk adjustment strategies have been used in the literature for adjusting nurse staffing measures, including patient age, diagnosis, treatment stage, co-morbidities, risk for outcomes, acuity, and other health status factors (Brennan, Daly, & Jones, 2013).

The NDNQI stratifies skill mix data by a well-established nursing unit typology. In addition, comparative reports are available based on hospital structural characteristics.

Brennan, C. W., Daly, B. J., & Jones, K. R. (2013). State of the science: the relationship between nurse staffing and patient outcomes. *Western Journal of Nursing Research*, *35*(6), 760–794. http://doi.org/10.1177/0193945913476577

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

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**: 2b3.7.**

Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*): **2b3.8.**

Statistical Risk Model Calibration – Risk decile plots or calibration curves: **2b3.9. Results**

of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

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

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified

(describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat

the information provided related to performance gap in 1b)

Clinically meaningful targets are currently set using multiple sources of data by hospitals including Nursing Hours Per Patient Day and RN Hours Per Patient Day and other data (e.g., generated by a well calibrated and audited commercial patient-classification system) (Needleman, et al., 2011). These data will provide national benchmarks to inform multiple stakeholders and foster a learning health system. National comparable hospital skill mix levels (e.g., current use of quartiles data in NDNQI or future potential quintiles display (five-star rating) in transparent hospital reporting) informs hospital providers and other stakeholders given the robust evidence that skill mix is associated with better patient safety outcomes.

Needleman, J., Buerhaus, P., Pankratz, V. S., Leibson, C. L., Stevens, S. R., & Harris, M. (2011). Nurse staffing and inpatient hospital mortality. *The New England Journal of Medicine*, *364*(11), 1037–1045. http://doi.org/10.1056/NEJMsa1001025

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*)

Table 2 – Distribution of Standardized Skill Mix by Unit-Type

Unit Type	Minimum	25th percentile	50th percentile	75th percentile	Maximum
Adult Critical Care	-3.87	-0.61	-0.01	0.72	1.56
Adult Step Down	-2.73	-0.55	-0.03	0.60	2.88
Adult Medical	-3.07	-0.40	0.13	0.64	3.34
Adult Surgical	-2.75	-0.51	0.02	0.66	3.29
Adult Medical-Surgical	-2.67	-0.51	0.01	0.56	3.37
Ped. Critical Care	-3.37	-0.32	0.29	0.84	1.07
Ped. General	-2.66	-0.60	-0.11	0.65	1.87
Rehab	-2.11	-0.53	0.08	0.68	3.02

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?*)

The distributions of standardized skill mix scores show variation across unit type, particularly in the tails of the distribution. Based on the number of units in each unit-type with standardized skill mix scores at least 2 standard deviations away from the mean for that unit-type, we believe that meaningful differences can be shown.

	N units	< 2SD Number of units (%)	< 1 SD Number of units (%)	> 1 SD Number of units (%)	> 2 SD Number of units (%)
Adult Critical Care	1,045	15 (1.4%)	115 (11.0%)	272 (26.0%)	104 (10.0%)
Adult Step Down	923	14 (1.5%)	94 (10.2%)	194 (21.0%)	113 (12.2%)
Adult Medical	917	11 (1.2%)	61 (6.7%)	244 (26.6%)	161 (17.6%)
Adult Surgical	672	5 (0.7%)	56 (8.3%)	196 (29.2%)	117 (17.4%)

Adult Medical-Surgical	1,120	8 (0.7%)	106 (9.5%)	274 (24.5%)	172 (15.4%)
Ped. Critical Care	116	6 (5.2%)	14 (12.1%)	31 26.7%)	9 (7.8%)
Ped. General	416	6 (1.4%)	47 (11.2%)	117 (28.1%)	48 (11.5%)
Rehab	340	1 (0.3%)	33 (9.7%)	128 (37.6%)	93 (27.4%)

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

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

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

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

Data for the measure testing is from the National Database of Nursing Quality Indicators (NDNQI), which is a voluntary reporting database. Hospitals participating in NDNQI tend to be larger hospitals with a disproportionately large number of teaching hospitals and Magnet® recognized facilities compared to the population of all US hospitals. However, staffing rates from NDNQI data are similar to rates from other data sources in published literature. (See: Dabney & Kalisch, 2015; Kalisch & Lee, 2014; Bolton et al., 2003; Y.-F. Li et al., 2011; S. Li, Pittman, Han, & Lowe, 2017; Bae, Kelly, Brewer, & Spencer, 2014)

To examine the participation rates within the NDNQI sample, we first assessed the percent of eligible units that were participating across unit types. Among adult inpatient units the participation rates averaged 76%: critical care (75%), step down (77%), medical (76%), surgical (76%), and med-surg combined (75%). Pediatric unit types had a slightly lower participation rate, averaging 66%: pediatric critical care (68%), pediatric med-surg (64%).

Finally, we explored missing data patterns within the monthly data from units that reported on staffing from January 2016 through December 2018. The average overall missing data was 23% for RN hours per patient day and 32% for total nursing hours per patient day. However, most of the missing data was within units (i.e. a missed month of reporting) rather than across units.

Bae, S.-H., Kelly, M., Brewer, C. S., & Spencer, A. (2014). Analysis of nurse staffing and patient outcomes using comprehensive nurse staffing characteristics in acute care nursing units. *Journal of Nursing Care Quality*, *29*(4), 318–326. https://doi.org/10.1097/NCQ.0000000000000057

Bolton, L. B., Aydin, C. E., Donaldson, N., Brown, D. S., Nelson, M. S., & Harms, D. (2003). Nurse staffing and patient perceptions of nursing care. *The Journal of Nursing Administration*, *33*(11), 607–614.

Dabney, B. W., & Kalisch, B. J. (2015). Nurse Staffing Levels and Patient-Reported Missed Nursing Care. *Journal of Nursing Care Quality*, *30*(4), 306–312. https://doi.org/10.1097/NCQ.0000000000000123

Kalisch, B., & Lee, K. H. (2014). Staffing and job satisfaction: nurses and nursing assistants. *Journal of Nursing Management*, *22*(4), 465–471. https://doi.org/10.1111/jonm.12012

Li, S., Pittman, P., Han, X., & Lowe, T. J. (2017). Nurse-Related Clinical Nonlicensed Personnel in U.S. Hospitals and Their Relationship with Nurse Staffing Levels. *Health Services Research*, *52 Suppl 1*, 422–436. https://doi.org/10.1111/1475-6773.12655

Li, Y.-F., Wong, E. S., Sales, A. E., Sharp, N. D., Needleman, J., Maciejewski, M. L., … Liu, C.-F. (2011). Nurse staffing and patient care costs in acute inpatient nursing units. *Medical Care*, *49*(8), 708–715. https://doi.org/10.1097/MLR.0b013e318223a9f1

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*)

The multilevel random effects models used to estimate ICC reliability were estimated using a full information maximum likelihood (FIML) estimator. The FIML estimator uses all of the observed data scores to estimate parameters, and is less biased than traditional list-wise deletion approaches to missing data and outperforms single imputation methods. Under missing at random (MAR) conditions, the FIML estimator yields unbiased parameter estimates, and in in sufficiently large samples performs as well or better than imputation methods (Schafer and Graham, 2002).

This method is also particularly well suited for missing data within a group (i.e. missing timepoints).

Schafer, J.L. & Graham J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7(2),* 147-177.

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*)

The majority of the missing data was within units (i.e. missed month(s) of reporting) rather than across units, suggesting there is not a systematic difference between reporting patterns by individual units, but rather many units failed to report data in all twelve months. NDNQI provides hospitals with quarterly reports, and calculates quarterly rates even if one month of data is missing in the quarter, so it is not surprising to find that hospitals may have missed one or months of reporting throughout the two-year period. The FIML estimation will correct for any bias in the calculation of the ICC for this sporadic missing data.

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.

Other

If other: generated from electronic payroll/accounting report or electronic staffing system

3b. Electronic Sources

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

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

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

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

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

Attachment:

3c. Data Collection Strategy

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

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

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

SKILL MIX DATA COLLECTION PROCESS, SOURCES, AND BURDEN

According to NDNQI guidelines, all separate nursing care hours data by licensure levels (RNs, LPNs, and UAPs), as well as employment status (hospital employees and agency/contracts), must reported by the calendar month. Payroll or staffing records should be audited to remove non-direct care hours (education, sick leave, vacation leave, etc.) and to ensure that ineligible staff are not included (i.e., unit secretary, monitor techs).

A survey was conducted in October 2018 of site coordinators (N=324) who have submitted skill mix data, which are used in calculating skill mix. Respondents indicated that skill mix data are obtained from electronic payroll/accounting systems most frequently (56.7%), followed by electronic staffing system reports (38.7%). A large majority of respondents (81%) indicated that generating and submitting skill mix data takes one day or less each month. The mean total hours required to extract clean, and submit the staffing measures was 6.0 with a range of 1-32 hours each month.

REPORTING ACCURACY

Overall, 79% of 467 respondents had high or very high confidence in the accuracy of the NDNQI quarterly staffing reports. For more accurate data collection, the NDNQI implemented several strategies, including periodic site coordinator surveys, data cleaning tools, and training for site coordinators. Almost 60% of respondents use some type of verification process prior to submitting data and 81% had used the NDNQI error reports to identify errors in submitted data. Almost 75% of respondents stated that they had to conduct manual adjustments to their data prior to submission in fewer than 10% of cases and 87% had to make manual adjustments to their data after submission in fewer than 10% of cases.

Since the measure was first endorsed, NDNQI has learned/modified the skill mix measure in a variety of ways. First, the definition of nursing care hours has been clarified by providing clear description on float staff (hospital employees temporarily assigned to provide direct patient care for all or part of a shift on a unit other than their unit of employment) when reporting their nursing care hours in the NDNQI data collection guidelines. Second, we periodically provided teleconferences for site-coordinators to educate, update some changes in data collection guidelines, and address issues about the definition of NDNQI quality indicators (e.g., nursing care hours) and data collection procedures (changes on the data entry fields and the use of data summary report and data error report to verify data before reporting to NDNQI). Lastly, NDNQI collects nursing care hours data through a secure NDNQI website. We provided data error messages to notify site coordinators that data on nursing care hours were not entered for all 3 months of a quarter, although it may or may not be an error. Most recently, we have tested the reliability of aggregating at the hospital-level, and have found the measure to be reliable. In addition, we have tested the reliability and validity of aggregating to the hospital-level medical, surgical, and medical-surgical units. These three combined units at the hospital-level have demonstrated reliability and validity.

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)*.

None

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)

	Public Reporting State of Illinois http://www.healthcarereportcard.illinois.gov/ State of Maine https://mhdo.maine.gov/_pdf/NSI%20Microspec%20Manual%20Nov%202013%20edition%20rev%20B.1.pdf State of New York https://www.health.ny.gov/regulations/recently_adopted/docs/2015-01-07_disclosure_quality_surveillance.pdf State of Vermont http://legislature.vermont.gov/statutes/section/18/221/09405b State of Illinois http://www.healthcarereportcard.illinois.gov/ State of Maine https://mhdo.maine.gov/_pdf/NSI%20Microspec%20Manual%20Nov%202013%20edition%20rev%20B.1.pdf State of New York https://www.health.ny.gov/regulations/recently_adopted/docs/2015-01-07_disclosure_quality_surveillance.pdf State of Vermont http://legislature.vermont.gov/statutes/section/18/221/09405b Regulatory and Accreditation Programs Professional Certification or Recognition Program The American Nurses Credentialing Center Magnet Recognition Progam http://www.nursecredentialing.org/Magnet The American Nurses Credentialing Center Pathways to Excellence Program http://www.nursecredentialing.org/Pathway Quality Improvement (external benchmarking to organizations) National Database of Nursing Quality Indicators pressganey.com Quality Improvement (Internal to the specific organization) National Database of Nursing Quality Indicators participating hospitals pressganey.com

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

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

Public Reporting
Illinois: "Illinois Hospital Report Care and Consumer Guide to Health Care" through the Illinois Hospital Report Card Act
Sponsor: Illinois General Assembly & Illinois Department of Health
Purpose: Provide consumers with access to information about the quality of health care provided in the state
Geographic Area: All hospitals in Illinois
Information: Total Nursing Hours Per Patient Day, RN Hours Per Patient Day.
Website: http://www.healthcarereportcard.illinois.gov/
Maine: "Nursing Sensitive Indicator Quality Data Set"

Sponsor: Maine Health Data Organization
Purpose: To create and maintain a useful, objective, reliable and comprehensive health information database that is used to improve the health of Maine citizens
Geographic Area: All acute care hospitals in Maine
Information: Total Nursing Care Hours Per Patient Day, RN Hours Per Patient Day
Website:
https://mhdo.maine.gov/_pdf/NSI%20Microspec%20Manual%20Nov%202013%20edition%20rev%20B.1.pdf
New York: "Disclosure of Quality and Surveillance Related Information"
Sponsor: New York Public Health and Health Planning Council and the Commissioner of Health
Purpose: Requires hospitals to disclose nursing quality indicator information to any member of the public.
Geographic Area: Hospitals and nursing homes in the state of New York
Information: Total number of nursing hours per patient day, RN hours Per Patient Day, LPN Hours Per Patient Day, UAP Hours Per Patient Day
Website: https://www.health.ny.gov/regulations/recently_adopted/docs/2015-01-07_disclosure_quality_surveillance.pdf
Vermont: "Hospital Community Reports,"
Sponsor: Vermont General Assembly
Purpose: Statute establishing standard formats for hospital community reports
Geographic area: All hospitals in Vermont
Information: Nursing hours per patients day
Website: http://legislature.vermont.gov/statutes/section/18/221/09405b
Professional Certification or Recognition Program
The American Nurses Credentialing Center (ANCC) includes skill mix as part of their Magnet Recognition Program and Pathways to Excellence Recognition Program (ANCC, 2015).
http://www.nursecredentialing.org/Magnet
http://www.nursecredentialing.org/Pathway

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?*)

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.*)

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.

4a2.1.1

In 2018, we solicited feedback from all NDNQI participating hospitals (approximately 1800), and received feedback from 324.

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.

Hospitals participating in the National Database of Nursing Quality Indicators receive quarterly reports, which provide unit, unit type, and hospital-level benchmarks across a number of benchmarking categories (e.g. hospital type, bed size, teaching status).

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.

In 2018, we survey NDNQI site coordinators using a REDCap survey to obtain feedback about the usability of the measures and reports for quality improvement and feasibility of data collection at their facilities.

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

From a survey of 324 site coordinators, 8.2% report skill mix to a state database or regulatory program, 5.2% report to a national regulatory group, 4.1% report to a state quality registry (other than NDNQI), 4.6% report to a national registry (other than NDNQI), and 21.1% report staffing data to a state or national credentialing program.

In a survey of 324 site coordinators, 69.7% reported that the staffing measures are somewhat or very important to their hospital's quality improvement program.

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

N/A

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.

In previous years, we have examined the feedback to consider revisions to the measures that will reduce the burden of data collection, including changes to required versus optional data elements, and improved tools for data collection. No changes have been made to the measures based on the most recent feedback.

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.

A study was conducted evaluating trends in nursing care hours provided on general care (medical, surgical, and medical-surgical combined) units and critical care units among NDNQI hospitals from 2004-2011. The sample included 2,634 medical units, 1,895 surgical units, 3,561 medical-surgical units, and 2,822 critical care units from 1,499 hospitals. During that time period there was a statistically significant 9.7% increase in the proportion of hours provided by RNs on general care units. At the same time, the proportion of nursing care hours provided by LPNs dropped by more than half on both general care units and ICUs (Staggs & He, 2013).

Staggs, V. S., & He, J. (2013). Recent trends in hospital nurse staffing in the United States. The Journal of Nursing Administration, 43(7-8), 388–393.

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.

These measures have been used in several studies to demonstrate the positive relationship between nurse staffing and improved patient outcomes. Although this is not an unexpected benefit, it can be used to argue for improved nurse staffing in hospitals.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0190 : Nurse staffing hours - 4 parts

0205 : Nursing Hours per Patient Day

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?

Yes

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

The measure is completely harmonized with 0205, as both use the same database and definition of RN hours and are applied to the same settings. Measure 0204 is actually a ratio of the RN hours and Total Nursing Hours elements that are the numerator for the rates tested in Measure 0205. Measure 0190 has had its endorsement removed, but the key differences in the measure are in the setting (inpatient acute care versus nursing home/SNF) and the denominator (per 1,000 patient days for 0204 versus total number of residents for 0190).

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.)

Nursing hours per patient day and nurse staffing hours – 4 parts are related, not competing measures. Nursing hours per patient day is also a measure for which the American Nurses Association is the measure steward, and measures a different aspect of nurse staffing. There is no additional data collection burden. Therefore,

Nurse staffing skill mix is not considered to be a competing measure with the other two measures (nursing hours per patient day and nurse staffing hours – 4 parts).

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: 0204_MeasureLogic_ScientificSupplement.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American Nurses Association

Co.2 Point of Contact: Gregory, Craig, gregory.craig@ana.org, 301-628-5395-

Co.3 Measure Developer if different from Measure Steward: University of Kansas Medical Center

Co.4 Point of Contact: Emily, Cramer, ecramer2@kumc.edu, 913-588-1657-

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.

The American Nurses Association sponsored the development of the nursing hours per patient day and nursing staff skill mix measures. The Lewin Group was hired by ANA to identify measures that likely were nurse-sensitive. An interview guide was developed and various institutions were selected based on their geographical location and organizational characteristics to provide a nation-wide sample that would include an academic medical center, private hospital, public hospital, urban hospitals, rural hospitals and hospital system. JCAHO, Catholic Health Association, AHA and AHCPR were also contacted to provide broader context. The interviews were conducted with nursing executives, quality specialists and other experts identified by each organization between August 1995 and October 1995. ANA's advisory committee was Rhonda Anderson RN, FAAN, Joanne Disch, PhD, RN FAAN, Gwendolyn Johnson, MA, RN, Clair B. Jordan, MSN, RN, Norma Lang, PhD, RN, FAAN, Pamela Mitchell, PhD, CNRN, FAAN, Margaret Sovie PhD, RN, FAAN, and Mary K. Walker, PhD, RN, FAAN.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 1998

Ad.3 Month and Year of most recent revision: 01, 2014

Ad.4 What is your frequency for review/update of this measure? annual updates, with every 3 year reendorsement

Ad.5 When is the next scheduled review/update for this measure? 12, 2013

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: