BACKGROUND

Access to care is essential, particularly for our currently fragmented healthcare system, which generally delivers episodes of face-to-face treatment with minimal communication between encounters. While people agree access to healthcare is necessary, there are several definitions and interpretations of access to care, creating confusion and frustration for all. Moreover, measuring access is further confounded by interpreting what is meaningful access, what care actually was delivered, and the impact of access on intermediate outcomes or outcomes. Measuring the quality of services differs from measuring the access to services of different quality levels.

Access often is associated with the availability of resources and is frequently dependent on financing. Penchansky and Thomas describe access as a "set of dimensions that characterize the fit between the patient and the healthcare system," including geographical, temporal, financial, cultural, and digital access.[1]

Traditional access concepts, and hence measurement points, focus on in-person experiences between the patient and provider; however, there are opportunities beyond this paradigm. One potential option is to improve digital access between the patient and provider.[2] A shift in culture will be required to utilize this method that could help diminish geographical, temporal, and cultural access problems faced by patients. Access to healthcare also can be improved beyond the doctor's office or hospital by providing wellness and health promotion at work sites, which is where many individuals spend the majority of their time or through health system changes and a focus on population health as the measurement leverage point.[3]

NQF works to help improve access to care by focusing on performance measures that can both help identify key areas to measure and to identify gaps in access to care measures. During the Health and Well-Being Phase 2 project, the Standing Committee noted the measurement focus and specifications of measures #1516, #1392, #2689, and #2695,[4] which do not assess whether specific care processes are occurring during a patient encounter, rather only confirm the visit[5] even though the developer(s) explicitly stated that these measures are intended to assess access

[1] Khan AA, Bhardwaj SM. Access to health care: a conceptual framework and its relevance to health care planning. Eval Health Prof. 1994;17(1):60-76.
[2] Fortney JC, Burgess JF Jr, Bosworth HB, et al. Re-conceptualization of access for 21st century healthcare. J Gen Intern Med. 2011;26 (Suppl 2):S639-S647.
[3] Stoto M. Population Health Measurement: Applying Performance Measurement Concepts in Population Health Settings. eGEMs (Generating Evidence & Methods to improve patient outcomes): Vol. 2: Iss. 4, Article 6.
[4] NQF 1516 Well-Child Visits in the Third, Fourth, Fifth, and Sixth Years of Life; NQF 1392 Diabetes Screening for People With Schizophrenia or Bipolar Disorder Who Are Using Antipsychotic Medications (SSD)
[5] National Quality Forum (NQF). Health and Well-Being Phase 2. Final Report. Washington, DC: NQF; 2015. Available at http://www.qualityforum.org/Publications/2015/11/Health_and_Well-Being_Phase_2_Final_Report.aspx. Last accessed May 2016.

to care. As an example, the two well-child visit measures assess only that visits occurred and not whether the child received the age-appropriate vaccinations, hearing, or vision tests. Other measures were focused more globally, e.g., hospitalization for dehydration, were asserted as reflecting access to and coordination of a community's ambulatory services.

The purpose of this document is to provide guidance to developers and NQF Committees on access to care measure development and the NQF evaluation of such measures. The document also includes examples of measures and concepts that are used to assess access to care.

TABLE 1. EXAMPLES OF MEASURES & CONCEPTS

Subject/Concept	Measure Title	Steward
Dental Care Visits	1) Ambulatory Care Sensitive Emergency Department Visits in Dental Caries in Children[6] 2) Follow-Up after Emergency Department Visit by Children for Dental Caries[7]	American Dental Association/Dental Quality Alliance
Well-Child Visits	1) Well-Child Visits in the Third, Fourth, Fifth, and Sixth Years of Life[8] 2) Well-Child Visits in the First 15 Months of Life[9]	National Committee for Quality Assurance
Dehydration Admissions	Dehydration Admission Rate (PQI 10)[10]	Agency for Healthcare Research and Quality
HIV/AIDs	HIV Late Diagnoses[11]	Centers for Disease and Control
Health Insurance Coverage	Percent of persons with health insurance	NHIS (national database)*
Unmet Need	Percent of families that experience difficulties or delays in obtaining health care or do not receive needed care for one or more family members	MEPS/MCBS (national database)*
Mental Health/Substance Abuse	Percent of adults with serious mental illness who received treatment	NHSDA (national database)*

*These measures are a part of AHRQ's preliminary measure set, National Healthcare Disparities Report, 2002: http://archive.ahrq.gov/research/findings/nhqrdr/nhdr02/premeasurea.html.

[6] NQF 2689: Ambulatory Care Sensitive Emergency Department Visits for Dental Caries in Children
[7] NQF 2695: Follow-Up after Emergency Department Visit by Children for Dental Caries
[8] NQF 1516: Well-Child Visits in the Third, Fourth, Fifth, and Sixth Years of Life
[9] NQF 1392: Well-Child Visits in the First 15 Months of Life
[10] NQF 0280: Dehydration Rate (PQI 10)
[11] NQF 1999: Late HIV Diagnoses

Performance measures are evaluated against NQF's measure evaluation criteria, which are used to determine suitability of measures for use in both quality improvement efforts and for accountability purposes. The five major criteria[12] are:

1) Importance to measure and report - This criterion allows for a distinction between things that are important to do (or outcomes of importance) versus those processes, structures or outcomes that rise to the level of importance required for a national performance measure. Importance has two key subcriterion: Evidence and Performance Gap. Evidence is the extent to which the specific measure focus is evidence-based and can drive significant gains in healthcare quality. Performance gap denotes there is variation in performance among measured entities or that disparities (e.g., by race or ethnicity) exist even if a "macro-level" analysis appears a measure is topped out.
2) Scientific acceptability of measure properties – This reflects NQF's view that performance measures must demonstrate sound measurement science—that is, they must be both reliable and valid.
3) Feasibility - The Feasibility criterion reflects the extent to which the data required to compute a measure are readily available and retrievable without undue burden, as well as the ease of implementation for performance measurement.
4) Usability and use - NQF-endorsed measures are considered suitable for both accountability and quality improvement purposes and the expectation is that endorsed measures not only will be used, but also ultimately will lead to improved patient outcomes.
5) Comparison to related or competing measures – Since there is an abundance of measures, this criterion requires a careful consideration of such similar measures, with the goal of endorsing only the best measures—or, if there isn't a "best" measure, endorsing measures that are consistent to the extent possible.

In addition to this evaluation guidance, NQF has provided additional guidance on composite, appropriate use, resource use, population health, and patient-reported outcome measures. For population health measures, for example, NQF's guidance[13] document notes that the core criteria remain the same, but the language and direction are tailored.

NQF has defined access as the "ability to obtain needed healthcare services in a timely manner including the perceptions and experiences of people regarding their ease of reaching health services or health facilities in terms of proximity, location, time, and ease of approach. Examples may include, but are not limited to, measures that address the timeliness of response or services, time until next available appointment, and availability of services within a community."[14] From this, a minimum scope of access measures could be inferred as addressing timeliness and

[12] More detail on these criteria can be found in the Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement Document.
[13] The complete population health guidance document can be found at this link: http://www.qualityforum.org/Publications/2012/06/An_Environmental_Scan_of_Integrated_Approaches_for_Defining_and_Measuring_Total_Population_Health.aspx.
[14] National Quality Forum (NQF). *Glossary of terms*. Washington, DC: NQF 2013.

availability.

- *Discussion for Workgroup: Does this definition of access generally suffice as the starting point for guidance on access measures? If not, what concepts are not captured—i.e., we do not intend for "wordsmithing," but at this point are looking to see what concepts the Workgroup thinks need to be added or excluded.*

Currently, the NQF portfolio lacks a robust set of measures related to access (as defined by any means). Based on experience with other classes of measures, specific guidance on how NQF Committees should evaluate access measures can, in turn, provide clarity to developers on nuances of developing such measures and NQF's expectations for them.

Table 2 sets forth NQF's general evaluation criteria and the guidance specifically approved for population health measures.

- *Discussion for Workgroup: We have annotated the access evaluation column for your review and discussion.*

Table 2. NQF Criteria, Population Health Measure Criteria, and Access Measure Criteria Guidance

NQF Measure Evaluation Criteria	*Population Health Measure Evaluation: Additional Guidance and Context**	*Access Measure Evaluation Criteria*
Conditions for Consideration Several conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards. **If any of the conditions are not met, the measure will not be accepted for consideration.**	**Conditions for Consideration** Several conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards. **If any of the conditions are not met, the measure will not be accepted for consideration.**	**Conditions for Consideration** Several conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards. **If any of the conditions are not met, the measure will not be accepted for consideration.**
A. The measure is in the public domain or a measure steward agreement is signed.	**A.** No change.	**A.** No change. (Here and hereafter "no change" refers to no change from the general criteria.)
B. The measure owner/steward verifies there is an identified responsible entity and a process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every three years.	**B.** The measure owner/steward verifies there is an identified responsible entity or multi-stakeholder entities and a process to maintain and update the measure on a schedule that is commensurate with the rate of **population health innovation**, but at least every three years.	**B.** The measure owner/steward verifies there is an identified responsible entity or multi-stakeholder entities and a process to maintain and update the measure on a schedule that is commensurate with the rate of **policy- or structural-related access innovation**, but at least every three years.
C. The intended use of the measure includes both public reporting and quality improvement.	**C.** The intended use of the measure includes both public reporting and improvement in efforts to improve population health.	**C.** The intended use of the measure includes both public reporting and **improvement in efforts to improve access**.
D. The measure is fully specified and tested for reliability and validity.[1]	**D.** No change.	**D.** No change.
E. The measure developer/steward attests that	**E.** The measure developer/steward attests that	**E.** No change.

harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate. **F.** The requested measure submission information is complete and responsive to the questions so that all the information needed to evaluate all criteria is provided. **Note** **1.** A measure that has not been tested for reliability and validity is only potentially eligible for time-limited endorsement if all of the following conditions are met: 1) the measure topic is not addressed by an endorsed measure; 2) it is relevant to a critical timeline (e.g., legislative mandate) for implementing endorsed measures; 3) the measure is not complex (requiring risk adjustment or a composite); and 4) the measure steward verifies that testing will be completed within 12 months of endorsement.	harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate. **Harmonization of related measures at the provider and population levels measures has been considered and addressed.** **F.** No change.	 **F.** No change.
Criteria for Evaluation If all conditions for consideration are met, candidate measures are evaluated for their suitability based on four sets of standardized criteria in the following order: *Importance to*	**Criteria for Evaluation** No change.	**Criteria for Evaluation** No change.

Measure and Report, *Scientific Acceptability of Measure Properties*, *Usability*, and *Feasibility*. Not all acceptable measures will be equally strong among each set of criteria. The assessment of each criterion is a matter of degree. However, if a measure is not judged to have met minimum requirements for *Importance to Measure and Report* or *Scientific Acceptability of Measure Properties*, it cannot be recommended for endorsement and will not be evaluated against the remaining criteria.		
1. Impact, Opportunity, Evidence—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-impact aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to*** *meet all three subcriteria to pass this criterion and be evaluated against the remaining criteria.* **1a. High Impact** The measure focus addresses: • a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF;	**1. Impact, Opportunity, Evidence—Importance to Measure and Report:** Extent to which the specific measure focus is evidence-based, important to making significant gains in **population health, improving determinants of health and health outcomes of a population** for a high-impact aspect of **health** where there is variation in **(including geographic variation)** or overall less-than-optimal performance. ***Measures must be judged to*** *meet all three subcriteria to pass this criterion and be evaluated against the remaining criteria.* **1a. High Impact** **Note: For population health measures, high impact would also be identified by the National Prevention Strategy and the DHHS Consensus Statement on Quality in Public Health.**	**1. Impact, Opportunity, Evidence—Importance to Measure and Report:** Extent to which the specific measure focus is evidence-based, important to making significant gains in **access to care leading to improved health outcomes** for a high-impact aspect of healthcare **or health** where there is variation in **(including geographic variation)** or overall less-than-optimal performance. ***Measures must be judged to*** *meet all three subcriteria to pass this criterion and be evaluated against the remaining criteria.* **1a. High Impact** No change

OR • a demonstrated high-impact aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality). **AND** **1b. Performance Gap** Demonstration of quality problems and opportunity for improvement, i.e., data[2] demonstrating considerable variation, or overall less-than-optimal performance, in the quality of care across providers and/or population groups (disparities in care). **AND** **1c. Evidence to Support the Measure Focus**	**OR** • a demonstrated high-impact aspect of **health** (e.g., affects **large population** and/or has a substantial impact for a smaller population; **source of significant health disparities**; leading cause of morbidity/mortality; functional health; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality). **AND** **1b. Performance Gap** Demonstration of opportunity for improvement in **health**, i.e., data[2] demonstrating considerable variation, or overall less-than-optimal performance, in **health across providers (healthcare, public health, and other partners) and/or population groups, (including but not limited to disparities in care).** **AND** **1c. Evidence to Support the Measure Focus**	**OR** • a demonstrated high-impact aspect of **healthcare or health** (e.g., affects **large population** and/or has a substantial impact for a smaller population; **source of significant health disparities**; leading cause of morbidity/mortality; functional health; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality). **AND** **1b. Performance Gap** Demonstration of opportunity for improvement in **access**, i.e., data[2] demonstrating considerable variation, or overall less-than-optimal performance, in **access across providers (healthcare, public health, and other partners) and/or population groups, (including but not limited to disparities in care).** **AND** **1c. Evidence to Support the Measure Focus**

The measure focus is a health outcome or is evidence-based, demonstrated as follows: • Health outcome:[3] a rationale supports the relationship of the health outcome to processes or structures of care. • Intermediate clinical outcome, Process,[4] or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome. • Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes. • Efficiency:[6] evidence for the quality component as noted above. **Notes** **2.** Examples of data on opportunity for improvement include, but are not limited to:	The measure focus is a health outcome or is evidence-based, demonstrated as follows: • Health outcome:[3] a rationale supports the relationship of **the health outcomes in the population to strategies to improve health.** • **Health determinant**, Intermediate outcome, Process, or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome. • **Experience with care, services or other health determinants**: evidence that the measured aspects of care are those valued by **people and populations** and for which the **respondent** is the best and/or only source of information OR that experience is correlated with desired outcomes. • Efficiency:[6] evidence for the quality component as noted above.	The measure focus is evidence-based, demonstrated as follows: ~~Health outcome~~: **Not applicable.** ***Question for Workgroup*: Is access itself an outcome so only a rationale for the measure (and not more detailed evidence review) is sufficient? Recall that the bar for evidence (see attached algorithm) is different for health outcome measures)?** • Intermediate outcome, Process, or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a improved access to care and a desired health outcome. • **Experience with access to care or services**: evidence that the measured aspects are those valued by **people and populations** and for which the **respondent** is the best and/or only source of information OR that experience is correlated with desired outcomes. • Efficiency:[6] evidence for the quality **and access** component as noted above. **Notes** **2.** Examples of data on opportunity for

prior studies, epidemiologic data, or data from pilot testing or implementation of the proposed measure. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem. **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.** 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. **5.** The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines. **6.** Measures of efficiency combine the concepts	**4. Population health determinants typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with stakeholder input) → provide intervention → evaluate impact on population health status. If the measure focus is one step in such a multistep process, the steps with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement.** **5.** No change. **6.** No change.	improvement include, but are not limited to: prior studies, epidemiologic data, or data from pilot testing or implementation of the proposed measure. If data are not available, the measure focus is systematically assessed (e.g., expert panel rating) and judged to be a quality problem. ~~**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.** Access typically includes several leverage points: access to payment coverage; covered services; access to (timely) services; receipt of services; quality of service received → improved outcome. If the measure focus is less proximal to the receipt of services and quality, the step with the strongest evidence for the link to improved access should be selected as the focus of measurement. **5.** No change. **6.** No change.

of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).		
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 subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.* **2a. Reliability** **2a1.** The measure is well defined and precisely specified[7] so it can be implemented consistently within and across organizations and allow for comparability. EHR measure specifications are based on the quality data model (QDM).[8] **2a2.** Reliability testing[9] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise.	**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 subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.* **2a. Reliability** **2a1.** The measure is well defined and precisely specified[7] so it can be implemented consistently within and across organizations, **multistakeholder groups, populations or entities with shared accountability for health** and allow for comparability. **2a2.** No change.	**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 subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.* **2a. Reliability** **2a1.** No change. [8] **2a2.** No change.

2b. Validity **2b1.** The measure specifications[7] are consistent with the evidence presented to support the focus of measurement under criterion 1c. The measure is specified to capture the most inclusive target population indicated by the evidence, and exclusions are supported by the evidence. **2b2.** Validity testing[10] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. **2b3.** Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;[11] **AND** If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately,	**2b. Validity.** **2b1.** No change. **2b2.** Validity testing[10] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the effect of interventions to **improve population health**, adequately identifying differences in effectiveness. **2b3.** Exclusions are supported by the **evidence**; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; **AND** If **individual or subgroup** preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure **or variation**; in such cases, the measure must be specified so that the information about **individual or subgroup** preference and the effect on the measure is transparent (e.g., numerator category computed separately,	**2b. Validity** **2b1.** No change. **2b2.** Validity testing[10] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in **access**. **2b3.** Exclusions are supported by the ~~clinical~~ evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion;[11] **AND** No change.

denominator exclusion category computed separately).[12] **2b4.** For outcome measures and other measures when indicated (e.g., resource use): • an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care;[13,14] and has demonstrated adequate discrimination and calibration **OR** • rationale/data support no risk adjustment/ stratification. **2b5.** Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful[15] differences in performance; **OR** there is evidence of overall less-than-optimal performance.	denominator exclusion category computed separately). **2b4.** For outcome measures and other measures when indicated (e.g., resource use): • an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on factors that influence the measured outcome (but not factors related to **disparities in population health or health interventions**) and are present at start of care;[13,14] and has demonstrated adequate discrimination and calibration **2b5.** Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and meaningful[15] differences in performance **or variation across populations in improving health.** **OR** there is evidence of overall less-than-optimal performance **or significant variation across**	**2b4.** When indicated: • ~~an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care;[13,14] and has demonstrated adequate discrimination and calibration.~~ ***QUESTION TO WORKGROUP:*** **When is risk adjustment appropriate for access measures, if ever?** **OR** • ~~rationale/data support no risk adjustment/ stratification.~~ **2b5.** Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and ~~practically/clinically~~ meaningful[15] differences in performance **(i.e., access)**; **OR** there is evidence of overall less-than-optimal performance **(i.e., access)**.

	populations.	
2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.	**2b6.** No change.	**2b6.** No change.
2c. Disparities If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);	**2c. Disparities** If **health disparities** have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);	**2c. Disparities** If disparities in **access to** care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender);
OR	**OR**	**OR**
rationale/data justifies why stratification is not necessary or not feasible.	**No option for justification for lack of stratification.**	No change.
Notes **7.** Measure specifications include the target population (denominator) to whom the measure applies, identification of those from the target population who achieved the specific measure focus (numerator, target condition, event, outcome), measurement time window, exclusions, risk adjustment/stratification, definitions, data source, code lists with descriptors, sampling, scoring/computation.	**Notes** **7.** No change	**Notes** **7.** No change.

8. EHR measure specifications include data type from the QDM, code lists, EHR field, measure logic, original source of the data, recorder, and setting.	**8.** N/A	**8.** EHR measure specifications include data type from the QDM, code lists, EHR field, measure logic, original source of the data, recorder, and setting.
9. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).	**9.** No change.	**9.** No change.
10. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as	**10.** No change.	**10.** No change.

a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.		
11. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.	**11.** Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, **multistakeholder groups, and populations** and sensitivity analyses with and without the exclusion.	**11.** No change.
12. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.	**12.** N/A	**12.** No change.
13. Risk factors that influence outcomes should not be specified as exclusions.	**13.** Risk factors that influence outcomes should not be specified as exclusions.	**13.** Risk factors that influence **access** should not be specified as exclusions.
14. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic	**14.** Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in health determinants, such as race, socioeconomic status, or gender (e.g., poorer **health outcomes** of African American men with prostate cancer or inequalities in CVD risk factors between men and women). It is preferable to stratify measures by race and	**14.** Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in **access to** care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and

status rather than to adjust out the differences. **15.** With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.	socioeconomic status rather than to adjust out the differences. **15.** With large enough sample sizes, small differences that are statistically significant may or may not be **practically meaningful**. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of **people** who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is meaningful; or whether a statistically significant difference of $25 in cost for an intervention (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers **or populations**.	socioeconomic status rather than to adjust out the differences. **15.** With large enough sample sizes, small differences that are statistically significant may or may not be practically ~~or clinically~~ meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of **people** who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers. ***QUESTION FOR THE WORKGROUP*: Is this note appropriate for access measures?**
3. Usability: Extent to which intended audiences (e.g., consumers, purchasers, providers, policymakers) can understand the results of the measure and find them useful for decision-making. **3a**. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audiences for public reporting (e.g., focus group, cognitive testing) or	**3. Usability: Note: intended audiences can include community members and coalitions.** **3a.** Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audiences for public reporting (e.g., focus group, cognitive testing) or	**3. Usability:** No change. **3a**. No change.

rationale; **AND** **3b.** Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audiences for informing quality improvement[16] (e.g., quality improvement initiatives) or rationale. **Note** **16.** An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.	rationale; **AND** **3b.** Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audiences for informing improvement[16] in **health determinants and/or population health** or rationale. **Note** **16.** An important outcome that may not have an identified improvement strategy still can be useful for informing **improvement in quality and/or population health** by identifying the need for and stimulating new approaches to improvement.	**AND** **3b.** Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audiences for informing improvement in **access** or rationale. **Note** **16.** An important **measure** that may not have an identified improvement strategy still can be useful for informing **improved access** by identifying the need for and stimulating new approaches to improvement.
4. Feasibility: Extent to which the required data are readily available or could be captured without undue burden and can be implemented for performance measurement. **4a.** For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).	**4. Feasibility:** No change. **4a.** No change for clinically oriented measures.	**4. Feasibility:** Extent to which the required data are readily available or could be captured without undue burden and can be implemented for performance measurement. **4a.** No change for clinically oriented measures.

4b. 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.	**4b.** The required data elements are available in electronic health records, **personal health records, health information exchanges, population data bases**, or other electronic sources. If the required data are not available **in existing electronic sources**, a credible, near-term path to electronic collection is specified.	**4b.** The required data elements are available in electronic health records, **personal health records, health information exchanges, population health data bases,** or other electronic sources. If the required data are not available **in existing electronic sources**, a credible, near-term path to electronic collection is specified.
4c. Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified.	**4c**. Susceptibility to inaccuracies, errors, **inappropriate comparison across populations,** or unintended consequences and the ability to audit the data items to detect such problems are identified.	**4c.** No change.
4d. Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality,[17] etc.) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use).	**4d.** No change.	**4d.** No change.
Note **17.** All data collection must conform to laws regarding protected health information. Patient confidentiality is of particular concern with measures based on patient surveys and when there are small numbers of patients.	**Note** **17.** All data collection must conform to laws regarding protected health information. **Confidentiality** is of particular concern with measures based on **individual** surveys and **for small populations**.	**Note** **17.** All data collection must conform to laws regarding protected health information. Patient confidentiality is of particular concern with measures based on patient surveys and when there are small numbers of patients.

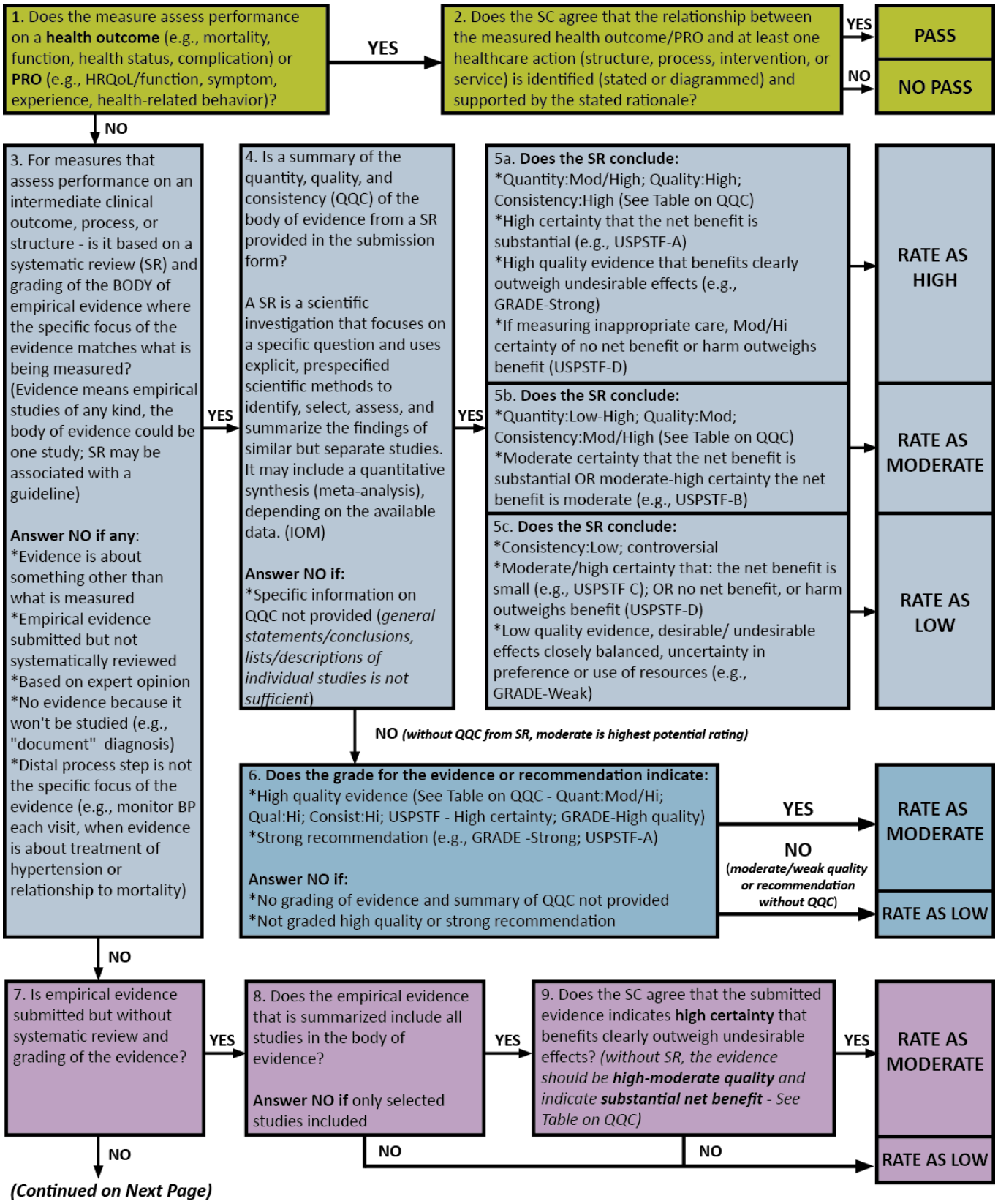
5. Comparison to Related or Competing Measures If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure. **5a.** The measure specifications are harmonized[18] with related measures; **OR** the differences in specifications are justified. **5b.** The measure is superior to competing measures (e.g., is a more valid or efficient way to measure); **OR** multiple measures are justified.	**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. **Note: Complementary measures that address different improvement strategies are not considered competing measures.** **OR** **5b.** No change.	**5. Comparison to Related or Competing Measures** No change. **5a.** The measure specifications are harmonized[18] with related measures. **Complementary measures that address different strategies to improve access are not considered competing measures. For example, a Medicaid program measure of access to X service and a system measure of availability (or delivery) of same service would be complementary and not competing.** **OR** the differences in specifications are justified. **5b.** No change.

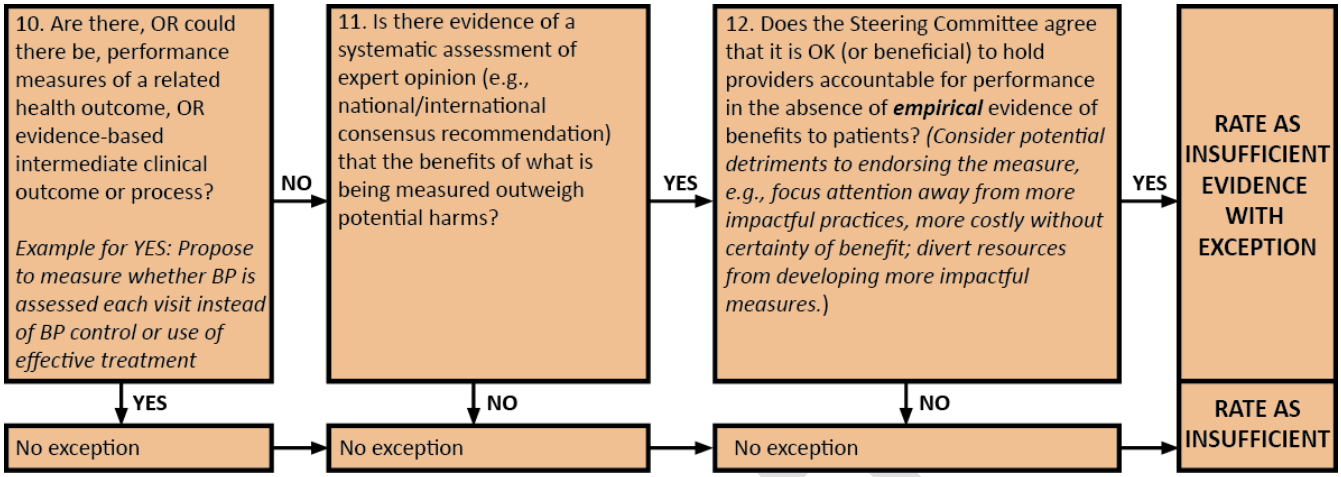
Note **18.** Measure harmonization refers to the standardization of specifications for related measures with the same measure focus (e.g., *influenza immunization* of patients in hospitals or nursing homes); related measures with the same target population (e.g., eye exam and HbA1c for *patients with diabetes*); or definitions applicable to many measures (e.g., age designation for children) so that they are uniform or compatible, unless differences are justified (e.g., dictated by the evidence). The dimensions of harmonization can include numerator, denominator, exclusions, calculation, and data source and collection instructions. The extent of harmonization depends on the relationship of the measures, the evidence for the specific measure focus, and differences in data sources.	**Note** **18. Additional conceptualization needed for harmonization between clinical and population-level measures.**	**Note** **18. Additional conceptualization needed for harmonization among clinical, population, resource use, appropriate use, and access measures (i.e., is a broader NQF portfolio issue).**

Finally, NQF's population measures guidance provides a few overarching recommendations: First, "total population health measures should use existing state and local-sponsored population health surveys, clinical care system administrative data, patient registries, and government public health surveillance." Additionally, population health measures should ensure the capacity to measure disparities using distributive metrics to assess disparities in care. An access measure that is too far upstream is less important than a more proximal access measure.

- *Discussion for Workgroup: Is there additional guidance beyond that embedded within the evaluation criteria the Workgroup wishes to include? Are there specific leverage points the Workgroup feels need to be emphasized—e.g., measuring access closer to delivery or delivery itself vs. an access point more distally. Should commentary be offered that the focus here is on healthcare, whereas guidance on measurement of access to social determinants is better provided through the population health framework and guidance?*

Algorithm 1. Guidance for Evaluating the Clinical Evidence




10. Are there, OR could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcome or process?
Example for YES: Propose to measure whether BP is assessed each visit instead of BP control or use of effective treatment
NO
11. Is there evidence of a systematic assessment of expert opinion (e.g., national/international consensus recommendation) that the benefits of what is being measured outweigh potential harms?
YES
12. Does the Steering Committee agree that it is OK (or beneficial) to hold providers accountable for performance in the absence of empirical evidence of benefits to patients? (Consider potential detriments to endorsing the measure, e.g., focus attention away from more impactful practices, more costly without certainty of benefit; divert resources from developing more impactful measures.)
YES
RATE AS INSUFFICIENT EVIDENCE WITH EXCEPTION
YES
No exception
NO
No exception
NO
No exception
RATE AS INSUFFICIENT