



Variation in Measure Specifications: Sources and Mitigation Strategies: Second Draft Report

September 6, 2016

This report is funded by the Department of Health and Human Services under contract HHSM-500-2012 -000091 Task Order HHSM-500-T0000.

Introduction

As the U.S. healthcare system has become increasingly focused on issues of quality, cost, and efficiency, the use of quality measurement in healthcare has grown rapidly in both scope and importance. This has led to a proliferation of measures across a diverse range of clinical areas, settings, data sources, and programs, and there is growing recognition that performance measures used in various programs (e.g., at the federal, state, and community levels) are often not well-aligned. For example, different programs, while intending to address the same fundamental quality issue, may use slightly and/or significantly modified versions of the same measure. An analysis conducted by Bailit and Associates in 2013¹ identified 1,367 quality measures in use across 48 different state and regional programs. An examination of the measures revealed that there were only 509 “distinct” measures, i.e. measures used in multiple programs without changes and/or variations in specifications. After removing all duplicates and distinct measures, the remaining 800+ measures were overlapping or similar in focus, with one or more variations in the specifications.

This lack of consistency across measurement programs is viewed by many as unnecessary variation, contributing to a number of challenges facing providers, patients, health plans, regulators, and others who use measures to assess relative performance. For clinicians, hospitals, and other healthcare providers, variation may lead to an increase in data collection and reporting requirements without attendant increases in value. For those seeking to use measure results to inform decisions about and comparisons of healthcare, variation may diminish the value of measurement results. Specifically, if measures have been changed in a way that compromises the comparability of results, users may not be able to draw accurate conclusions about the differential performance of those being measured.

Purpose and Objectives

The purpose of this project is to identify reasons for and impact of variation in measure specifications, as well as provide guidance on ways to mitigate and/or prevent variation.

Throughout this project, NQF has worked with a multi-stakeholder Expert Panel to identify how, where, and why variation in measure specifications occurs; develop consensus definitions to facilitate common understanding around key terms, concepts, and measure components to help standardize measurement efforts and minimize unnecessary variation; and create a taxonomy for understanding and interpreting the contributing factors of variation and the implications of this variation. The taxonomy, along with feedback related to challenges faced by key informants, was then used to develop a framework and point-of-use decision guide on how to address variation during implementation and development of measures. Through the use of an environmental scan the project has explored the many facets of variation and developed practical guidance to help stakeholders in identifying, and addressing reason for variation in measure specifications.

This second draft report synthesizes and summarizes all the findings and concepts considered by the Expert Panel and builds on the preliminary recommendations in the first draft report. Key informant interview themes are incorporated into the document to illuminate practical considerations and reasons

NATIONAL QUALITY FORUM

for variation. In addition, the consensus definitions, taxonomy and framework included in this second draft will be finalized by the Expert Panel based on additional feedback from member and public comments.

Expert Panel

Through a public call for nominations, NQF convened a multi-stakeholder Expert Panel to accomplish the objectives of the project. The Panel includes experts in measure development and implementation, health informaticists, provider groups representing those who are being measured, purchasers, payers, and others who use measure results (see [Appendix C](#)).

The Panel provided input on all phases of the work (e.g., guidance on the environmental scan, framework development, and recommendations) through a series of in-person meetings and conference calls. The Panel also provided practical examples that exemplified and elaborated on the extent and impact of variation (see [Appendix A](#)).

Methodology

Environmental Scan

A critical component of this project was the environmental scan, which assessed the landscape of variation in measure specifications to help clarify and illuminate the nature and extent of variation. This environmental scan focused on how, where, and why variation is occurring across the healthcare system, and included a literature review and key informant interviews.

Literature Review

The literature review portion of the environmental scan included a review of both peer-reviewed published literature and non-peer reviewed (i.e. gray) literature from the past 10 years. The literature review was informed by input from Expert Panel members, key informant interview suggestions, and feedback received through a public comment period on the first draft report. To identify relevant literature, NQF staff used a combination of the following medical subject heading (MeSH) search terms: variation, change, measure specification, guideline implementation, the Healthcare Effectiveness Data and Information Set (HEDIS), measure, variance, performance, quality, metrics, methodology, quality improvement, burden, implementation science, translational science, point-of-care changes in measures, measure collection and volume. The MeSH searches and combination of search terms used led to a very small number of relevant articles only articles that empirically compared two or more sets of measure specifications, directly addressed the reasoning supporting a change to measure specifications, and/or described in qualitative or quantitative terms differences in burden due to variation in measure specifications were reviewed. Staff identified additional articles by conducting a Google search using simple terms such as “measure,” “change,” and “clinical burden.” The results of these searches demonstrated the relative scarcity of literature on this topic. Searches conducted at the beginning and throughout the project identified only 65 articles and reports, many of which were only tangentially related to measure variation. This absence of a large body of literature underscores the

importance of “real-world” experiential data, which was sought through examples and case-studies from measure implementers, payers, purchasers, providers, health informaticists and federal partners.

Key Informant Interviews

Data gathering for the project included interviewing key informants who represent a wide variety of healthcare stakeholders and perspectives. Interviews were conducted using an NQF-developed key informant interview guide which consisted of a standard set of questions related to measure variation along with a sub-set of questions that were modified to address unique perspectives and experiences of each key informant. Key informants included representation from the federal government, payers, measure implementers, quality collaboratives, consumers, and electronic health record vendors. The interview guide is available in [Appendix B](#). Key informants are meant to be a representative sample of those who develop as well as use measures. The key informant interviews provided information used to enhance and corroborate themes from the literature search and are not necessarily exhaustive in nature.

Environmental Scan Results

Literature Review

Literature on this topic was relatively limited; selected studies and their findings are presented below:

- *Different definitions of clinical concepts:* One study evaluated the effect of alternative specifications for a measure assessing the persistence of beta-blocker treatment after a heart attack: The authors compared results of the measure in instances where the concept of ‘adherence following myocardial infarction’ was defined and specified in different ways; the goal of this analysis was to determine whether any of the definitions more accurately predicted a composite of post-myocardial infarction outcomes. While this study was an assessment of the measure’s predictive validity at the patient level, and did not assess facility-level performance scores, the number of patients categorized as ‘adherent’ varied substantially (between 7 percent and 73 percent) depending on the definition of adherence that was used.²
- *Different exclusion rules:* Another study assessed different exclusion rules to identify whether complications were present on admission or hospital-acquired. The different exclusion rules varied substantially in their ability to correctly identify present-on-admission complications; the authors also noted that rates of mortality and length of stay were significantly higher for patients with hospital-acquired complications.³
- *Different measure timeframes:* A study found that varying the timeframe for a measure of the rate of Veterans Health Administration Medical Centers’ patients diagnosed with an alcohol abuse disorder who were receiving pharmacotherapy led to differences in facility rankings of as many as 24 percentile points when changing the measurement period from one year to two, and 29 percentile points when the measure was modified to focus on those receiving treatment for the first time.⁴
- *Different populations:* A study found that changing the denominator of a measure of glycemic outcomes from all patients in Veterans Health Administration Medical Centers to only those who are receiving a complex glycemic regimen led to “markedly different” facility rankings.⁵

- *Different data sources:* One study found that, when compared to manual review of data, electronic reporting significantly underestimated rates of appropriate asthma medication administration and pneumococcal vaccination and overestimated rates of cholesterol control in patients with diabetes, though nine other measures were relatively comparable.⁶ Another study of more than 200 commercial health plans found that nearly three-quarters of plans had a greater than 10% point difference in the prevalence of beta-blocker use after myocardial infarction when administrative data were supplemented with medical record data, compared with using administrative data alone.⁷ Similar research was conducted with measures of body mass index and immunization in children, with virtually identical findings.⁸ A comparison of the EHR-based version and the claims-based version of one measure found that the claims-based measure significantly underperformed the EHR-based measure when compared to a physician reviewed “gold standard”, as measured by the number of diabetics correctly identified as patients. In this case, the EHR was virtually identical to the physician review.⁹
- *Incomplete measure specifications:* An examination of ten NQF-endorsed eMeasures found that “literal implementation of specifications was not feasible due to incomplete specification and data availability issues in four instances.” When researchers adapted the measures to fit data elements available in their electronic health record system, “results substantially varied from those expected”.¹⁰
- *Difference in guidelines:* A study described the potential for variation in clinical guidelines to contribute to “downstream” variation in measure specifications that are derived from those guidelines. This study compared seven clinical practice guidelines or consensus statements related to inpatient glucose management, finding significant differences in content, depth of detail, and other characteristics, particularly with respect to process recommendations.¹¹

When discussing burden with respect to quality measures, the literature is scarce and focuses on the burden of measurement for providers and does not address the burden that arises from measure variation. The focus is on the volume and the number of measures that a provider has to report on for accountability purposes. One study reports that “dealing with these measures imposes a considerable burden on physician practices in terms of understanding the measures, providing performance data, and understanding performance reports from payers, but that the extent of the burden has not been quantified.”¹²

Key Informant Interviews

Key informants consistently addressed the following three interrelated areas: data, measure complexity and clarity, and transparency. These areas were addressed as either a contributing factor that caused variation and/or a strategy to address variation so as to mitigate its impact.

- All of the key informants identified data as either a cause of variation and/or a strategy for mitigating variation. Measure users and implementers-as key informants-emphasized that the need for variation is most frequently based on what data are available, the completeness of the data, as well as access to all the necessary data elements. The interviewees noted that when necessary data are available, some implementers need to aggregate the data due to sample size

problems, and in the process change measure specifications, which become “drivers” of variation. However, when data are available but not readily accessible, collaborative efforts such as data sharing agreements can actually mitigate the need for variation by availing data already in existence. Data was the only topic with this dichotomous nature of being a contributor as well as mitigator/minimizer of variation.

- Measure complexity was emphasized as another cause of variation. Key informants called out risk adjustment, case mix adjustment and exclusions as areas that are most complex for front-line providers such as physicians, nurses, and nurse practitioners. Misunderstanding of measure constructs, along with a lack of training in quality measurement, creates an environment where the healthcare workforce, especially frontline providers, are ill-equipped to understand the fundamentals of a measure and appropriately capture data required to compute the measure. Secondly, as measurement science evolves and becomes more complex, the complexity itself becomes a driver of variation based on user’s interpretation of the measure. Additionally, individuals abstracting data from EHRs lack training to do so; incorrect data aggregation and analysis also affects comparability and introduces another level of variation post-implementation of measures. A number of key informants cited measures with complex case mix and or risk adjustments as examples of measure complexity.
- Measure clarity was highlighted as an additional reason for variation, specifically measure descriptions for numerators and denominators that are either unclear and/or incomplete. Key informants noted that some deviations happen due to misinterpretations of specifications stemming from a lack of clarity. Misunderstanding of the measure parts creates an environment where frontline providers incorrectly capture data required to compute the measure.
- Lack of transparency regarding measure variation was the most commonly cited concern voiced by the key informants. Transparency could include acknowledgement that a measure has been changed and, if possible, disclosure of the extent and type of variation (i.e., explicitly identifying what was changed and how) as well as the impact of the variation. During the key informant interviews, some interviewees noted that any deliberate change in measures should be accompanied by a before and after calculation that captures the magnitude of the impact of variation on measure results. This information should then be made available, along with a justification for creating the variant. Key informants noted that for proprietary measures such as HEDIS measures with licensing requirements, transparency is circumscribed to only acknowledging that a change was made to the measure being reported and cannot address the exact nature of the variation.

These themes are incorporated throughout the rest of the report and can be found in the sections discussing reasons for variation as well as mitigation strategies used to prevent and/or minimize variation.

Consensus Definitions of Key Terms Needed to Address Variation in Measure Specifications

Defining Variation

With a need for greater coordination and consistency across healthcare quality programs, terms such as “alignment” and “harmonization” are increasingly used to describe activities that are intended to address variation in measurement. However, these terms are not always used in a clear or consistent manner. Moreover, variation itself is a term that has multiple connotations in the healthcare context, and can be thought of in a number of different ways. Finally, specific elements of measure specifications are themselves inconsistently defined. This project will help to clarify definitions, principles, and types of variation.

Variation in measure specifications

Even when there is alignment across programs at a conceptual level, the measures used in these programs may vary in different ways, e.g., with different definitions of clinical concepts, target populations, and risk-adjustment strategies.

In a hypothetical illustration, two programs could be measuring whether patients with depression are achieving improvement, but are specifying their measures differently, for example:

- Different tools: One is using the PHQ-9 tool to quantify and measure improvement, while the other is using the PROMIS Depression Short Form;
- Different numerators: Both are using PHQ-9, but are defining “improvement” differently (a 3-point improvement vs. a 6-point improvement, or a 3-point improvement vs. a score less than 5);
- Different timeframes: One is using a six-month timeframe and the other is using a twelve-month timeframe;
- Different populations: One is focusing on patients 65 and older while the other is including all adults.

Variation at this level—where measures that have the same conceptual focus vary in their specifications—is the primary concern of this project.

To ensure that the healthcare quality field has a common understanding of concepts relevant to variation in measure specifications, the Expert Panel has proposed definitions of some key terms related to variation. In many instances, definitions, particularly definitions of measure specification elements, are derived from those routinely used in NQF’s work.

Measure variation:

The Expert Panel defines **measure variation as any deviation from reference measure specifications.**

This definition recognizes that, for practical purposes, measure variation cannot be identified or assessed without first identifying an accepted point of reference from which other specifications are deviating. Any measure may be used as a reference point, but the Expert Panel suggested that it would be preferable to use measures from the universe of well-established sets such as NQF-endorsed measures and/or HEDIS measures as common reference points.

This definition includes different types of deviations resulting in both intentional and unintentional variation. For example, an organization implementing a particular measure may determine that the reference specifications are not suitable for their needs and may modify the specifications accordingly. Alternatively, an organization, as the result of ambiguous specifications, may inadvertently, implement a measure in a way that is inconsistent with the intent of the original specifications, resulting in unintentional variation. Each of these scenarios qualifies as a measure variation under the definition proposed above.

Additionally, it should be noted that an instance of variation and/or a variant can be introduced both at the development and/or implementation stage. For example, the measure can demonstrate reliability and validity during its development and the measure can be endorsed, but unless there is the ability to look for variation in the implementation stage, it cannot be known if variation has occurred post-development. Therefore, audits can play an important role in identifying and addressing variant/variation creations post-implementation.

Types of variation

According to the definition of measure variation presented above, variations manifest as either inadvertent or intentional changes to specifications of a given reference measure. For this reason, Expert Panel members suggested that users should identify the particular specification that has been varied as a first-order question when assessing a specific instance of variation. A list of measure specifications that are commonly varied is presented in Table 1, along with examples of variation that might occur in each type. This is not a comprehensive list and any aspect of a measure's specifications can be altered to create a variant.

Table 1: Examples of variation in measure specifications

Measure specification element	Example of variation
Numerator	<ul style="list-style-type: none"> Differences in definitions, coding, or documentation of clinical concepts (e.g., 'encounter', 'adherence', etc.) Differences in performance thresholds or criteria
Denominator	<ul style="list-style-type: none"> Differences in definitions, coding, or documentation of clinical concepts Measure applied to an age group different from the age group in the reference measure specifications
Exclusions from denominator	<ul style="list-style-type: none"> Differences in acceptable exclusions (e.g., specific medical conditions vs. unspecified "medical reasons")
Risk adjustment	<ul style="list-style-type: none"> Differences in variables included in risk adjustment models

	<ul style="list-style-type: none"> • Adjustment for clinical factors only vs. adjustment for clinical plus sociodemographic factors • Differences in risk-adjustment strategy (e.g., logistic vs. hierarchical modeling)
Data source or collection instrument	<ul style="list-style-type: none"> • Use of administrative claims vs. registry reporting
Care setting	<ul style="list-style-type: none"> • Measure intended to be applied to hospitals is applied to ambulatory care facilities
Level of analysis	<ul style="list-style-type: none"> • Measure intended to evaluate health plan performance is used to evaluate individual clinician performance
Attribution strategy	<ul style="list-style-type: none"> • Attribute performance to the provider most often seen versus a particular provider type such as family med/internist/GP

Key Definitions

Alignment

Measures are aligned when they target the same outcome or care process in the same target population, but may not be completely identical with respect to specific measure element characteristics.

Alignment encourages the use of conceptually similar performance measures across and within public and private sector efforts.

Harmonization

Harmonization is the standardization of specifications for related measures with the same measure focus (e.g., influenza immunization of patients in hospitals or nursing homes), or related measures for 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.

Harmonizing measures reduces variations in measure elements and their specifications for similar measure concepts, and should be considered when measures are intended to address either the same measure focus—the target process, condition, event, outcome (e.g., numerator)—OR the same target population (e.g., denominator).

Reference Measure

A reference measure is the “parent” measure from which a variant has been created.

The concept of variation in measure specifications implies the existence of a reference point or “reference measure”. Measure variation is identified by comparing specifications that have been modified in the variant as compared to those of its reference measure (the “parent” measure). The entity (i.e., organization, individual, etc.) creating a variant should be the entity that identifies the reference measure—often this is the implementer of a particular variant, but it may also be a developer of a variant. The “pairing” of a reference measure and its variant reflects circumstances and choices made at a particular point in time. Because specifications of measures may change over time (e.g., due to updates in coding, available medications, changes in evidence, etc.), the creator of the variant may periodically decide to select a different reference measure (which implies that the variant itself also may change over time).

It is possible that there may be more than one potential reference measure at a particular point in time from which users could choose. The Expert Panel has identified three categories of potential reference measures, as follows:

Category 1: Measures that have been reviewed and approved by a multi-stakeholder consensus-based entity utilizing an evidence-based process. These measures may be used in an accountability program, and are publicly available.

Category 2: Measures that have not been reviewed and approved by a multi-stakeholder consensus-based entity utilizing an evidence-based process, but are used in an accountability program. These measures are publicly available.

Category 3: Measures that have not been reviewed and approved by a multi-stakeholder consensus-based entity utilizing an evidence-based process and are not used in an accountability program but are publicly available.

The Expert Panel recommended that when more than one potential reference measure is available, the implementer or developer should choose one from Category 1 when possible. Such measures—for example, those endorsed by NQF—are publicly available and have demonstrated merit regarding evidence, opportunity for improvement, reliability, validity, feasibility, and usability. If selecting a reference measure from Category 1 is not possible, the implementer or developer should select a measure from Category 2, assuming one is available. Creating a preferential order for categories reflects the Panel’s attempt to foster alignment, standardization, consistency in quality measurement, and recognition of the benefits of using well-vetted and/or established measures used by others in healthcare (e.g., potential to compare and better interpret their own performance results).

The Expert Panel also recommended that measures that are not publicly available such as unpublished or proprietary measures should not be selected as reference measures, unless no other measures are available.

While the Expert Panel recognized the hierarchical nature of the categories defined above, members were careful to note that particular categorization of a measure does not necessarily imply a value

judgement on its suitability for use. For example, developers of a measure in Category 3 may not have yet had the opportunity for NQF endorsement. In fact, the categories themselves emphasize the point-in-time pairing of a variant and its reference, as reference measures may “move” from one category to another over time. Members of the Panel also acknowledged that implementers or developers may have incomplete knowledge of available reference measures and their category placement. For example, an implementer may not realize that a particular measure is being used in state quality improvement programs or in programs created by private payers. These limitations notwithstanding, hierarchical categories may prompt a more thorough search and investigation of potential reference measures than might not otherwise have been done—which theoretically could result in identification of a measure that would not need to be modified after all.

Variant Measure

A variant is a measure that differs from the specifications of the reference measure. Variant is used to describe the measure and not the specific instance of variation.

Accountability Programs

Programs that use specific measure results to make judgments, comparisons and decisions based on performance, such as reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, licensure, professional certification, health information technology incentives, performance-based payment, and network inclusion/exclusion).

Performance Measure

According to the Institute of Medicine (IOM) definition, a performance measure is the “numeric quantification of healthcare quality.”¹³ The IOM defines quality as “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge.” Thus, performance measures can quantify healthcare processes, outcomes, patient perceptions, and/or organizational structure and/or systems that are associated with the provision of high-quality care.

Measure Specifications

Measure specifications are the technical instructions for how to build and calculate a measure. They describe a measure’s building blocks: numerator, denominator, exclusions, target population, how results might be split to show differences across groups (stratification scheme), risk adjustment methodology, how results are calculated (calculation algorithm), sampling methodology, data source, level of analysis, how data are attributed to providers and/or hospitals (attribution model), and care setting. Measure specification elements most commonly varied are defined below:

Numerator

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome, such as definitions, specific data collection items/responses, code/value sets.

Denominator

All information required to identify and calculate the target population, such as definitions, specific data collection items/responses, code/value sets.

Exclusions

Criteria applied before a measure is calculated in order to narrow the target population to remove any individuals for whom the treatment is not applicable. Exclusions are removed from the denominator before determining if numerator criteria are met.

Level of Analysis

The accountable entity whose performance is being measured (e.g. clinician, health plan, county population).

Care Setting

Any facility or office, including a discrete unit of care within such facility, that is organized, maintained, and operated for the diagnosis, prevention, treatment, rehabilitation, convalescence or other care of human illness or injury, physical or mental, including care during and after pregnancy.

Target Population

The group of care recipients for whom quality of care is being assessed.

Timeframe

The time period in which data will be aggregated to calculate the measure result.

Data Source

Source(s) from which data are obtained for measurement.

Risk Adjustment

Risk adjustment, as defined by the 2014 NQF report on “Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors, “ refers to statistical methods to control or account for patient-related factors when computing performance measure scores; methods include multivariable modeling, indirect standardization, direct standardization.”¹⁴

Possible patient-related factors include clinical severity, conditions present at the start of care and socio-economic as well as socio-demographic factors. Because patient-related factors can have important influence on patient outcomes, risk adjustment can improve the ability to make accurate and fair conclusions about the quality of care patients receive.

Data element

A data element is a single piece of information that is used in quality measures to describe part of the clinical care process, including both a clinical entity and/or its context of use (e.g., diagnosis, active).

Data elements are often patient-level information (e.g., blood pressure, lab value, medication, surgical procedure, death).

Measure score

A measure score is the numeric result that is computed by applying the measure specifications and scoring algorithm. The computed measure score represents an aggregation of all the appropriate patient-level data (e.g., proportion of patients who died, average lab value attained) for the entity being measured (e.g., hospital, health plan, home health agency, clinician, etc.).

Modification

Any change to one or more measure elements, regardless of ultimate impact on comparability or burden. Modification of a reference measure results in the creation of a measure variant.

Transparency

For the purposes of this project, the Expert Panel defines ‘transparency’ to mean the following ideal state of disclosure for users of variants:

- Disclosing that a change to the measure specifications has been made
- Being specific, if possible, about which specifications have changed
- Showing, if possible, the impact on the measure result of the changes

Feedback Loops

Feedback loops are a way to collect and share useful information between users, developers, and stewards of a measure. Examples include a user identifying an unclear specification and recommending a change to the measure steward, and the stewards’ promulgation of that new specification.

Mitigation

Mitigation refers to the reduction in the impact of variation along with a lessening of the negative effects resulting from variation in measure specifications.

Access / Availability / Completeness of Data

The terms ‘access to data’, ‘availability of data’, and ‘completeness of data’ all refer to the degree to which data elements necessary to capture a measure are readily available. Appropriate data are a necessary precursor to implementing a measure as specified. The inability to access data (e.g. prohibitive cost), the unavailability of data (e.g. treatment information is not collected in a claim or health record), or the incompleteness of data (e.g. only adults 65 and older are captured in the dataset) are all contributors to limitations of data, which challenge efforts to correctly calculate a measure result consistent with specifications. For the purposes of this project, discussions of data will not be explicitly parsing out each nuance, but data issues will be addressed as permutations of the possibilities described above.

Benchmarking

The process of comparing performance of an institution, entity, and/or other provider with that of their peers. The process of comparing measure results against accepted best practice thresholds.

Burden

Additional costs or resources associated with healthcare measurement activities, which may include the collection, management, reporting, and analysis of data for measurement purposes. Measurement burden may also include work by consumers of measure information (e.g., patients, payers, etc.) to understand and interpret measure results.

Comparability

Comparability refers to whether measure results can be used to make fair and valid comparisons between measured entities. Variation in measure specifications can diminish comparability, as seemingly minor changes in specifications may lead to significant differences in measure results.

Variation Taxonomy: Identifying Reasons for and Impact of Variation

A classification system for identifying variation and assessing its effects was developed based on feedback from the Expert Panel, Key Informants, and those who commented on the first draft report. As a preliminary step toward developing a framework for assessment of variation and its effects, two main principles were identified that served as guidance when considering a given instance of variation or the concept of variation in general:

Intended use

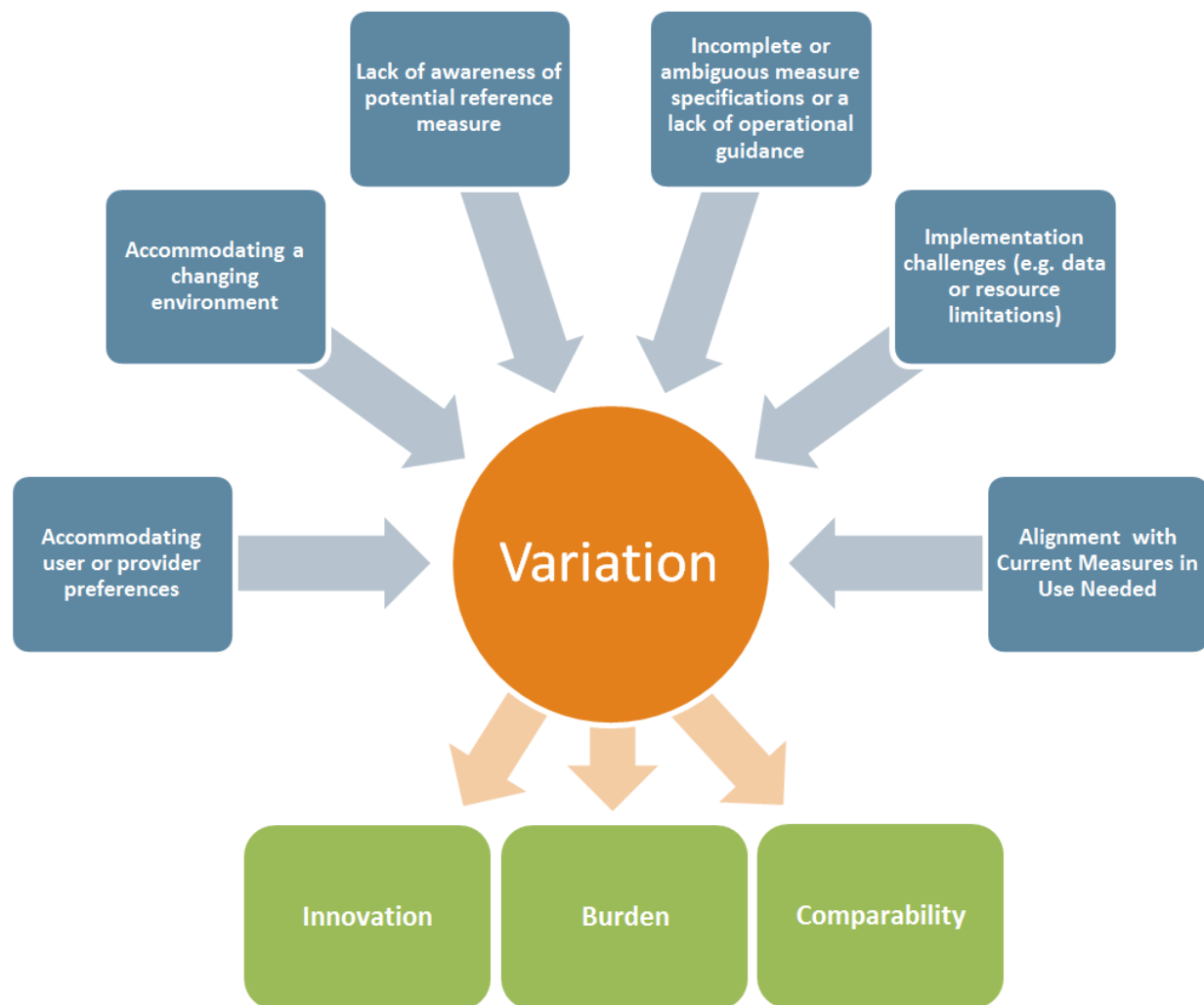
The significance of variation depends substantially on whether measures are being used for internal quality improvement (QI) or accountability purposes. If a measure is modified by a healthcare provider for their own QI efforts, this variation is likely to have little impact on any party other than that particular organization, as the results are only being used internally. However, if a measure is being used in external accountability programs (e.g., if the measure results are being publicly-reported by a state or regional collaborative), then a healthcare provider's modification of that measure may undermine the comparability of measure results, since other measured providers are not modifying the measure in the same way. Because the impact of variation is likely to be higher when measures are used for accountability applications, and because of NQF's focus on accountability, this project will largely focus on measure variation intended for accountability purposes.

Stages in the measure lifecycle

Measure variation can present at any and all stages of a measure's lifecycle (e.g., ideation, development, selection, implementation and use, reporting, etc.). Interventions to mitigate unnecessary variation or increase transparency around necessary variation will differ depending on where and when the variation is occurring. For example, variation arising from development of redundant measures might be addressed by increasing access to information on existing measures, while variation arising from

modification of measures during implementation might be addressed by providing additional implementation guidance or by working to increase awareness of the impact of such modifications.

The diagram below illustrates the factors that lead to measure variation, as well as the potential impacts of variation.



Taxonomy: Reasons for Variation

In order to fully understand and address variation, it is necessary to understand why developers or implementers create variants. Both panel members and key informants agreed that there are a number of reasons for variation; these have been grouped into the following overarching categories:

Modification of existing specifications to accommodate user or provider preferences

Specifications may be changed to better capture measure results for a specific patient population (e.g. changing the numerator or denominator to look at a subpopulation or to define the population of interest in a different way). For example, a provider may want to tighten the timeframe of a specific

measure to drive change within their organization. The measure may be specified for one level of analysis but the user would like to gather data at an alternative level. This type of variation is often done deliberately to address a particular need or preference of the measure user.

Modification of existing specifications to accommodate changing science

Sometimes the clinical science underlying a particular measure changes, necessitating changes to the measure. Depending, however, on the timing of the evidence change, an implementer may update the measure (e.g., to conform to new clinical guidelines) prior to (or concurrent with) revision of the measure by the developer.

Lack of awareness of existing measures that would meet user needs

This type of variation is the result of developers and implementers unknowingly creating duplicative measures because an existing measure that would meet the needs of the user was not sought or was not found. The user may not be aware of how to find existing measures, or perhaps such measures are not easily accessible. This can result in the creation of a “new measure” that is, in fact, a variant (although in this case, no reference measure is identified by the creator of the variant).

Incomplete or ambiguous measure specifications or a lack of operational guidance

Imprecise or ambiguous specifications were an area of great concern for many Expert Panel members, particularly for those who implement measures for Medicaid or at the state level. A number of Panel members suggested that measures can be poorly constructed or lack sufficient specificity to allow for consistent implementation. For example, at the state level, Medicaid core set measure users may “fill in gaps” in specifications by creating their own definitions of concepts and/or developing their own interpretations of vague specifications. This type of variation is done out of necessity, as the measure cannot be calculated otherwise.

Implementation challenges (e.g. data or resource limitations)

The Expert Panel and Key Informants noted that many instances of variation occur because of implementation challenges. These can include the type of data a measure implementer has access to (e.g., registry data vs. data from administrative claims), data integrity issues (e.g. missing data elements, attribution of data, inaccurate reporting of data) and/or differences in capabilities across and within organizations (e.g., resource limitations, inability of an EHR to capture required data elements correctly, or a lack of interoperability between systems required to capture the data elements of a measure as specified). The challenge with the type and integrity of data, i.e. missing and/or incomplete data, is that the data available do not match data needed per the measure specifications. Based on information gathered in this project, this is fairly common and often affects validity of the measure results. The other challenge arises due to varying EHR capabilities, where even within a single health system, different EHR systems or versions of an EHR system may be used, and one version may have the capability to capture the measure as specified while another does not.

Alignment with current measures in use

Implementers often participate in several accountability programs (e.g., Medicare and Medicaid programs, health plan programs, other payer programs, etc.), and these programs may use measures that are aligned conceptually but not harmonized. Given the burden of reporting on multiple different measures, users may alter measures to align them with their current accountability program measures. This facilitates data gathering efforts and increases both participation and reporting rates as well.

Taxonomy: Impact of Variation

Many of the reasons for variation are based on resource constraints as well as practicalities of measuring quality in a continuously evolving healthcare system. Panel members agreed that variation should be avoided for most cases. However, they did acknowledge that some forms of variation are beneficial and warranted, such as those introduced through innovation. The intent of this section is to present and discuss potential impact of variation.

Innovation

As stated earlier, variation is not always negative and detrimental to quality measurement and improvement. Variation may result from new and innovative approaches to measurement. Developers and implementers should update measures as needed to match the changing healthcare environment. While temporary variants may be created, the ability to update and/or test alternative specifications based on new guidelines and policies or user feedback on the performance of the measure is an essential part of building a stronger set of quality measures. Innovative measures can become the new reference measure. However, innovative measures used alongside existing measures can result in additional burden and reduced comparability.

Burden

Variation creates burden through the use of multiple similar but different measures. This burden is experienced by all parties including providers, consumers and those involved in the development, maintenance, implementation, and review of measures and measure results. Users often struggle with competing quality reporting requirements from different payers which result in “double reporting,” reduced understanding of how to implement measures, and conflicting measure results. Time spent on gathering and reporting on variants of quality measures can often reduce the time providers, hospitals and other users have to provide effective care and improve quality of care and outcomes.

Comparability

Comparison of performance across healthcare providers is one of the primary goals of quality measurement. The ability to accurately assess and compare the differential performance of providers is essential for meaningful public reporting and payment programs. Variation in measure specifications may undermine this goal—changes in measure specifications can have a significant impact on measure results, so when measures are not being applied in a consistent, standardized way across providers, the users of measure results cannot be confident that those results reflect actual differences in performance rather than differences in the measures being used.

Strategies for Addressing Variation

The following strategies to address variation were developed in consideration of the various reasons for and effects of variation in measure specifications. Some of the strategies are intended to prevent variation from occurring in the first place, while others are intended to mitigate the effects of variation—these mitigation strategies should be applied when variation is unavoidable or if the benefits of variation have been determined to outweigh the costs.

Strategies to Prevent Variation

Access to Measures: The most direct way of preventing variation is to ensure access to measures and their specifications which includes access to measure specifications including regular updates from measure stewards regarding measures in existence as well as measures in development. This issue of accessibility can arise from measures not being publicly available and/or from difficulty in locating measures that address end-user needs. The responsibility of making measures accessible lies mostly with measure stewards. They are ultimately responsible for authoring readily interpretable measure specifications, while keeping those specifications up to date and being responsive to new clinical evidence, new data sources, and implementation feedback from measure users. Access to measures should also include information on measures under development. When development efforts are shared and different measure developers and/or stewards and implementers collaborate, variation is prevented through prevention of duplication of efforts.

Feedback Loop: Feedback loops are channels of communication between two or more entities that facilitate exchange of information and ultimately improve processes and outcomes. Specifically, feedback loops between measure implementers and measure stewards allow for clarifications to be provided and measure specific needs to be communicated. For example, when measure specifications appear unclear and measure implementers request clarifications, measure stewards can prevent variation by providing the necessary information required to use the measure as specified. If and when measure specifications do not address measure implementer needs and/or measures are unavailable to address end-user needs, providing feedback to the measure steward allows them to be responsive to end-user needs as well as be aware of measure variants created by end-users. The bi-directional exchange of information can also help prevent duplication of efforts- where both the measure steward and the measure implementer may be working on updating a current measure.

Implementation Guidance: Given that a lack of clear guidance on how to implement measures was one of the reasons provided as to why variation occurs, access to precise, unambiguous, and complete specifications should be available for all measure implementers. Measure stewards should be responsive to any requests for technical assistance and/or clarifications from measure implementers.

Data Collection Strategies: Given that lack of data is a key driver of measure variation, measure implementers should strive to obtain the data needed to calculate the measure as specified rather than create a variant. Possible strategies for addressing data issues include aggregation of available data, data abstracter education, interagency agreements, and/or creation of databases. These strategies

address both data availability and completeness issues, which often lead to variation in measure specifications, where numerators and denominators are changed based on data collection feasibility.

Strategies to Mitigate Variation

Identifying Measures: Searching for and identifying measures that address end-user needs minimizes downstream variation. The Expert Panel emphasized that every end-user should start by searching through available measures. To this point, all measure implementers should be educated in ways to search for existing measures using established publicly available measure repositories such as [NQF's Quality Positioning System](#), [AHRQ's National Quality Measures Clearinghouse](#), as well as lists of measures used in Federal accountability programs. Identification of the existing measures should focus on the most relevant set of reference specifications with regards to the measure implementer end-use goal.

Feedback Loops: Given that communication is fundamental to sharing and receiving clarifications along with most current measure related information, feedback loops can be dually purposed as a way to both address and/or prevent variation. One of the reasons changes in measure specifications are introduced is because existing measures do not address end-use needs. When stewards of the original measure are informed of the reasons for introduction of a variant, this can help both the steward and the implementer better understand whether the change in specifications is relevant and/or necessary, and can inform future updates or improvements to the original measure.

Transparency: Acknowledge Variation: The Expert Panel agreed that the first step in addressing variation is to be transparent about any changes made to the specifications of the reference measure. Similarly, the key informants highlighted the importance of transparency as the most common strategy employed by measure implementers. The Expert Panel recognized that some measures are bound by licensing requirements where the exact nature of the change cannot be disclosed. In such cases, the Expert Panel encouraged measure implementers to, at minimum, acknowledge variation.

Transparency: Disclose Changes: The main purpose of transparency is to foster communication. Therefore, if creating a variant, measure developers or implementers should disclose the specific changes that were made, if possible. This allows other implementers to see what changes were made, because they may be struggling with the same or similar issues. In addition, this allows developers and measure stewards to address actual measurement needs and measure reporting constraints through subsequent measure improvements, clarifications and/or changes to the specifications. For users and consumers, this allows for them to be aware of limitations with regards to comparability. Disclosure of changes allows for all users of data to account for the changes while comparing across providers, as well as facilitate quality improvement for both the measure and the measurement enterprise.

Collaboration: The Expert Panel noted that transparency is an essential first step in addressing variation and that the utility of transparency could be maximized by sharing information in a forum or a collaborative. Such a forum would permit implementers to discuss their measurement needs, and their reasons for creating variants, as well as share “work-arounds” that minimized variations and/or lessons learned.

Benchmarking: Benchmarking may allow measure implementers to assess the impact of variation, and determine whether changes are appropriate. Comparing results of the variant measure to regional or national performance benchmarks may help illuminate the extent to which variation of the measure specifications affects measure results, and whether the variant is being used in accordance with the intent of the original measure. This may also provide some insight into whether results of the reference and the variant measure are comparable. All benchmarking efforts should focus on the meaning and utility of the variant.

A Framework for Preventing and Mitigating Variation

The Expert Panel articulated a series of critical decision points experienced by both those developing measure concepts into full-fledged performance measures, as well as those implementing measures for accountability programs. The decision points are presented in a logic diagram that assists and guides the user in deciding whether variation is needed or not, and how to mitigate the negative externalities associated with using a measure variant. The decision logic diagram was developed based on the following guiding principles:

Promotion of comparability

Measures used for payment, public reporting, and other accountability purposes should provide information that enables meaningful comparison of measured entities. To the extent possible, consistency in specifications across measures with the same or similar focus should be pursued to promote comparability of measure results.

Reduction of burden

Hospitals, clinicians, and other healthcare providers often are required to report multiple quality measures to different entities, creating administrative complexity and adding data collection demands to organizations that have limited resources and many competing needs. While measurement is an essential activity that creates value for all healthcare stakeholders and warrants the use of resources, measurement efforts should be aligned, harmonized, and streamlined wherever possible to avoid redundant or unnecessary data collection and reporting burden for providers.

Protecting innovation

The field of healthcare quality measurement is evolving rapidly, and there will remain a need for continuous innovation and improvement in measure development and implementation. While alignment and harmonization of measurement activities will continue to be an important goal, efforts to reduce variation should not stifle innovation in measurement activities.

Meeting end-user needs

Healthcare organizations, purchasers, payers, and other stakeholders may have varying purposes, objectives, and priorities for measurement that require variation in measure specifications. End users of measures should be able to meet their needs with measurement, and efforts to reduce

measure variation should allow for sufficient flexibility in adaptation of measures where appropriate.

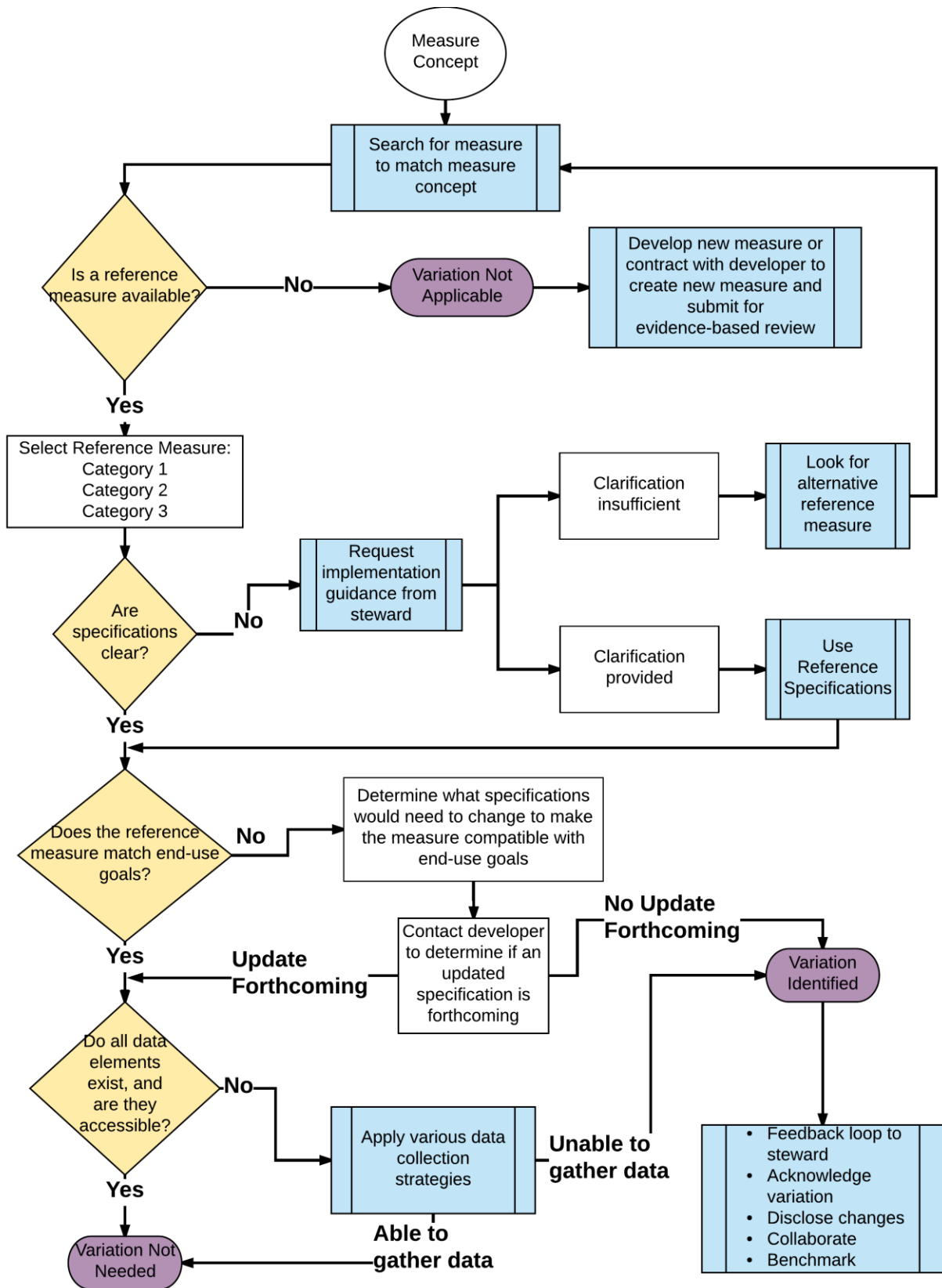
Specificity

To ensure consistency in implementation, measures used for accountability purposes should include precise, unambiguous, and complete specifications that minimize the need for interpretation or additional specification by measure users.

Transparency

Measure variants often are intentionally developed and implemented to meet particular needs of various stakeholders. Measure specifications may be modified to develop innovative new measures, to be responsive to the latest changes in clinical guidelines, and to finely tune measures to meet individual end-user needs and capabilities, including data availability. However, even if some types of variation are warranted, there is a need for increased information about the nature, scope, and impact of measure variation. Such transparency will help to identify where unnecessary variation is occurring so it can be avoided or mitigated, and to ensure necessary variation is clearly labeled and transparent, preventing misleading comparisons between similar (but not comparable) measure results. Recognizing that there are valid reasons for measure variation, and that variation cannot be avoided or mitigated in all situations, instances of variation in measure specifications should be fully and clearly disclosed to users of measure results, particularly where those measure results are used for public reporting, payment, or other accountability purposes. If possible, the parties who have introduced measure variation should also provide an assessment of the potential effects of that variation on measure results.

These principles should not be taken as strict rules or directives; indeed, members of the Expert Panel noted that there is tension between some of the principles, and some may even be in direct conflict. For example, the principle of meeting end-user needs may conflict with the principle of comparability if user needs require changes that lead to reduced comparability of measure results. The principles should be considered as guidance that should be balanced and taken into consideration as appropriate when applied in a particular context.



Using the Framework to Address Variation

The Framework presented here will serve as a guide for users seeking to prevent and mitigate the effects of measure variation. Users of the framework are able to proceed through each decision point, selecting the option that most closely matches their individual situation. When the user reaches a point where a need for variation is identified, the most relevant strategies for avoiding or reducing the impact of variation are provided. Each decision point of the framework will either suggest the applicable strategies to address and mitigate variation, or validate that variation is not applicable or needed.

Is an existing measure available?

The development and implementation of measures invariably begins with the conceptualization of a performance measure. Elements that fit into that measure concept may include the level of analysis (provider, hospital, and community), the target population (diabetics), and an outcome or process that reflects quality care (hospital readmissions). Other elements, such as the data source or the period of performance, e.g. lookback period, may be undetermined at the time of measure conceptualization.

When a measure concept and a need for measurement are identified, users are encouraged to search for a set of measure specifications that match their end-use needs. If such a measure is found, this may serve as the “reference measure” against which decisions about variation should be considered. As described in the section above that defines a reference measure, users should consult the various categories of reference measure as guidance when making a selection. If no measure is available, users are encouraged to develop a new measure, and submit for evidence-based review so that others might adopt their specifications.

Are specifications clear?

In order to implement measure specifications in such a way as to generate results comparable with those obtained by other users, users must completely understand how to obtain the data for the measure and calculate the result. Users are encouraged to contact measure stewards, requesting clarification and/or additional explanations in order to confidently implement the measure. If clarification is insufficient or not forthcoming in a timely fashion, seek an alternative reference measure.

Does the measure match end-use goals?

Having selected a reference measure, and clearly understanding the process for gathering and analyzing the data as specified, users are encouraged to reflect on whether the specifications of the existing measure match their end-use goals. The causes of a mismatch between measure specifications are many, and could include:

- a. The measure can be improved to increase reliability and validity, such as incorporating risk adjustment in order to accurately capture provider performance
- b. The measure is specified for analysis at the health plan level, but the user would like to implement the measure to evaluate a team of clinicians
- c. The evidence behind the measure has changed, and/or the target threshold for success has changed

Having identified one or more discrepancies between the measure specifications and the implementer's end-use goals, users are encouraged to contact the developer to determine if an updated set of specifications is forthcoming. Measures endorsed by the National Quality Forum are updated annually, and many organizations issue periodic updates to measures they steward. If no update is forthcoming, a measure variant may be created.

Users creating measure variants are encouraged to apply mitigation strategies to reduce possible negative impacts of variation, including acknowledging that the measure as implemented has been varied from the reference, full disclosure of those changes including an estimate of the impact on the measure result, collaboration with other measure implementers to form a learning community, benchmarking the measure result as calculated against a nationally recognized performance target, and finally, offering feedback to the measure steward to foster and facilitate the development of innovative changes to the reference measure.

Do all data elements exist, and are they accessible?

When the reference measure specifications match the user's end-use goals, users naturally turn to the implications of practical implementation. To implement a measure, all data elements must exist, and be accessible to the implementer. In the absence of data, users are encouraged to apply mitigation strategies targeted towards avoiding variation, by striving to collect the data as specified by entering into interagency data-sharing agreements and developing databases. If these strategies prove to not be feasible, users should use the same set of mitigation strategies mentioned for measures that do not meet user end-use goals.

Next Steps

Commenting on the second draft report will close on October 5, 2016 at 6:00pm ET. The Expert Panel will meet to discuss the comments received and finalize content for the final report. The report will then be shared with the Consensus Standards Approval Committee (CSAC) and updated as necessary. The final report will be completed and posted in December 2016.

References

- 1 Bazinsky K., and Bailit M. 2013. The Significant Lack of Alignment Across State and Regional Health Measure Sets. Bailit Health Purchasing, LLC. http://www.bailit-health.com/articles/091113_bhp_measuresbrief.pdf
- 2 Sanf  lix-Gimeno G., Franklin JM, Shrank WH, Carlo M., Tong AY, Reisman L., Matlin OS, Brennan TA, and Choudhry NK. 2014. "Did HEDIS Get It Right? Evaluating the Quality of a Quality Measure: Adherence to B-Blockers and Cardiovascular Outcomes After Myocardial Infarction." *Medical Care* 52 (7): 669-76. doi:10.1097/MLR.000000000000148.
- 3 Needleman J., Buerhaus PI, Vanderboom C., and Harris M. 2013. "Using Present-On-Admission Coding to Improve Exclusion Rules for Quality Metrics: The Case of Failure-To-Rescue." *Medical Care* 51 (8): 722-30. doi:10.1097/MLR.0b013e31829808de.
- 4 Fernandes-Taylor, Sara, and Alex H. S. Harris. "Comparing Alternative Specifications of Quality Measures: Access to Pharmacotherapy for Alcohol Use Disorders." *Journal of Substance Abuse Treatment* 42, no. 1 (January 2012): 102–7. doi:10.1016/j.jsat.2011.07.005.
- 5 Pogach LM, Rajan M., Maney M., Tseng CL, and Aron DC. 2010. "Hidden Complexities in Assessment of Glycemic Outcomes: Are Quality Rankings Aligned with Treatment?" *Diabetes Care* 33 (10): 2133-9. doi:10.2337/dc09-1665..
- 6 Kern LM, Malhotra S., Barr  n Y., Quaresimo J., Dhopeswarkar R., Pichardo M., Edwards AM, and Kaushal R. 2013. "Accuracy of Electronically Reported "meaningful Use" Clinical Quality Measures: A Cross-Sectional Study." *Annals of Internal Medicine* 158 (2): 77-83. doi:10.7326/0003-4819-158-2-201301150-00001.
- 7 Pawlson LG, Scholle SH, Powers A. Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures. *Am J Manag Care*. 2007;13(10):553-558.
- 8 Angier H., Gold R., Gallia C., Casciato A., Tillotson CJ, Marino M., Mangione-Smith R., and DeVoe JE. 2014. "Variation in Outcomes of Quality Measurement by Data Source." *Pediatrics* 133 (6): E1676-82. doi:10.1542/peds.2013-4277.
- 9 Tang PC, Ralston M., Arrigotti MF, Qureshi L., and Graham J. 2007. "Comparison of Methodologies for Calculating Quality Measures Based on Administrative Data Versus Clinical Data from an Electronic Health Record System: Implications for Performance Measures." *Journal of the American Medical Informatics Association : JAMIA* 14 (1): 10-5.
- 10 Amster A., Jentzsch J., Pasupuleti H., and Subramanian KG. 2015. "Completeness, Accuracy, and Computability of National Quality Forum-Specified EMeasures." *Journal of the American Medical Informatics Association : JAMIA* 22 (2): 409-16. doi:10.1136/amiainl-2014-002865.
- 11 Mathioudakis N., and Golden SH. 2015. "A Comparison of Inpatient Glucose Management Guidelines: Implications for Patient Safety and Quality." *Current Diabetes Reports* 15 (3): 13. doi:10.1007/s11892-015-0583-8.
- 12 Meyer GS, Nelson EC, Pryor DB, James B, Swensen SJ, Kaplan GS, et al. "More quality measures versus measuring what matters: a call for balance and parsimony." *BMJ Qual Saf*. 2012;21(11):964–8.

¹³ Crossing the Quality Chasm: A New Health System for the 21st Century , IOM March 2001. Link <http://www.nationalacademies.org/hmd/Global/News%20Announcements/Crossing-the-Quality-Chasm-The-IOM-Health-Care-Quality-Initiative.aspx>

14 National Quality Forum (NQF). Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors. Technical Report. Washington, DC: NQF; 2014. Available at http://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx. Last accessed August 2016.

Appendix A: Real World Examples of Variation

Type of variation	Examples	Potential reasons for variation	Real World Examples
Definitions of clinical concepts and/or terms	<ul style="list-style-type: none"> • Encounter • Adherence • Care Transition 	<ul style="list-style-type: none"> • Lack of common or standardized definitions 	<ul style="list-style-type: none"> • For a measure of statin use in diabetic patients, a health plan may define diabetics differently depending on whether the measure is being used for accountability or quality improvement purposes. For use in accountability programs, patients are included in the denominator population only if they meet strict criteria (e.g., two diabetes-related ICD claims plus two diabetes-related pharmacy claims), while for QI purposes, the criteria are looser (e.g., one ICD or pharmacy claim may suffice). This represents measure modification based on the purpose/use of the measure. • Application of access to primary care HEDIS measure to the Medicaid Child Core Set: The measure defines primary care provider (PCP) as the individual acting as the primary care provider for the patient. However, when implementing this measure at the state level, the state does not track all of its Medicaid patients and their primary care providers unless the patient is in a primary care case management program. Variation could result from state interpretation of the technical specification. • A measure of early elective deliveries was found to have variation in implementation, with different implementers defining early delivery in different ways (e.g., using estimated delivery date vs. a clinician's estimate of gestational age).
Coding or documentation of clinical concepts	<ul style="list-style-type: none"> • Variation in codes, fields, or problem sets used to indicate a clinical condition • Granularity 	<ul style="list-style-type: none"> • Data source change • Differences in available fields (e.g., claims vs. registry) 	<ul style="list-style-type: none"> • A registry and claims-based measure related to optic nerve evaluation for patients with glaucoma was retooled as an eMeasure. While the registry/claims-based measure assesses only whether or not an optic nerve exam was performed, the eMeasure version collects data at a more granular level, assessing whether specific aspects of an optic nerve exam were performed. This resulted in different levels of

Type of variation	Examples	Potential reasons for variation	Real World Examples
	of information		data being reported depending on whether the registry/claims-based measure or the eMeasure was being used.
Changes in implementation	<ul style="list-style-type: none"> Time intervention Exclusions 	<ul style="list-style-type: none"> Updated evidence Adaptation of measure for implementation 	<ul style="list-style-type: none"> Implementation of sepsis bundle: includes changes to the timeframe for the administration of crystalloid fluids for septic shock patients (i.e., 3hrs for CMS-adapted measure versus 6hrs for NQF measure) and denominator exclusions (e.g., the measure as implemented by CMS excludes patients with length of stay greater than 120 days, while the NQF-endorsed version does not).
Risk adjustment models/factors	<ul style="list-style-type: none"> SDS factors added to clinical risk-adjustment model Address risk arising from social and other vulnerabilities 	<ul style="list-style-type: none"> Different approaches to risk adjustment 	<ul style="list-style-type: none"> Measures in CMS's Hospital Readmissions Reduction Program (HRRP) are adjusted for clinical factors only, not for sociodemographic factors. A statewide hospital association has developed a methodology for adjusting HRRP measures for sociodemographic (SDS) factors, and publically-reports SDS-adjusted measure results on their website (along with results of the measure as calculated by CMS and unadjusted rates).
Target population	<ul style="list-style-type: none"> Changing target population based on data needs and/or intent/purpose of measurement 	<ul style="list-style-type: none"> Program needs (e.g., measure being used as part of a pediatric quality initiative) 	<ul style="list-style-type: none"> A readmissions measure specified for patients of all ages being applied to the Medicare population.
Level of analysis or attribution	<ul style="list-style-type: none"> Change in attribution strategy 	<ul style="list-style-type: none"> Program needs (e.g., measure being used as part of a clinician quality initiative) Data 	<ul style="list-style-type: none"> HEDIS breast cancer screening measure attributes patients to measured clinicians by including patients who have <i>any</i> enrollment claim or encounter with the clinician in the denominator population. A state-based quality collaborative narrowed the denominator of this measure to include only patients who have a <i>primary care</i> visit with the measured clinician.

Type of variation	Examples	Potential reasons for variation	Real World Examples
		availability	
Data source	<ul style="list-style-type: none"> • National registry versus chart data • Electronic vs. manual abstraction 	<ul style="list-style-type: none"> • Availability of data • Compliance or reporting requirements 	A national group was implementing a measure of early elective deliveries based on medical record review. A state-based quality collaborative proposed using a similar measure based on birth certificate registry data to minimize reporting burden for providers in that state, since the providers were already reporting this data. Use of these data may have missed medical record exclusions.

Appendix B: Key Informant Interview Guide

Variation in Measure Specifications

Interview Guide

Thank you for agreeing to meet with the NQF staff on the Variation Project. We want to let you know how these interviews will be used. We are conducting at least eight key informant interviews that will be themed and added to our second draft report for CMS. Ultimately, they will also be included in our final report. If we want to use any specific quotes you provide in the report, we will reach out to you in advance to ensure we have your permission to use those quotes.

To start today, please tell us about your organization, how it relates to quality healthcare measures and/or measurement science and what role you have within the organization.

General Questions

1. Are you seeing variation in measure specifications?
2. Who/what occupations or roles are introducing variation?
 - a. If purpose mentioned, skip to 10, then resume at 3
3. What is the value of the variation to your organization?
4. What types of variation are you seeing?
 - Can you categorize and define the types of variations?
5. Why do you think variation is occurring?
6. What impact do you perceive this variation to have?
7. Do you try to reduce or mitigate the variation and/or its impact? If not, then why not?
8. What methods do you use to reduce and/or mitigate variation?
9. In your opinion, does variation happen based on the purpose/use of the measure (e.g. public reporting, internal quality improvement)?
10. Can you give me examples of instances where variation happened because of the purpose and/or use of the measure?

Framework for Categorizing Variation

1. Do you have any additional thoughts about measure specification elements/areas that we may have missed?
2. When developing a variation mitigating framework, what should be our top three considerations?
3. In your opinion, what are the key elements needed to organize the types of variation happening while capturing the impact of variation as well? We are looking to develop a way to organize variation as well as capture the impact of variation.

Appendix C: Expert Panel

Panel Co-Chairs

Andrew Baskin, MD

National Medical Director- Aetna
Philadelphia, Pennsylvania

Blackford Middleton, MD, MPH, MSc

Chief Informatics and Innovation Officer, Apervita
Harvard T.H. Chan School of Public Health
Nashville, Tennessee

Panel Members

Matt Austin, PhD

Assistant Professor- Armstrong Institute for Patient Safety and Quality at Johns Hopkins Medicine
Baltimore, Maryland

Mary Barton, MD, MPP

Vice President- National Committee for Quality Assurance
Washington, DC

Beverly Court, PhD

Research Manager- WA State Dept of Social & Health Services
Olympia, Washington

Hazel Crews, PT, MHA, MHS, CPHQ

Executive Program Director- Indiana University Health
Indianapolis, Indiana

Tricia Elliott, MBA, CPHQ

Director, Quality Measurement – The Joint Commission
Chicago, Illinois

Charles Gallia, PhD

Senior Policy Advisor- State of Oregon
Portland, Oregon

Jeff Geppert, PMP, EdM, JD

Senior Research Leader, Batelle Memorial Institute
Columbus, Ohio

Matt Gigot, MPH

Program Manager, Clinical Data Analysis and Reporting- Wisconsin Collaborative for Healthcare Quality
Middleton, Wisconsin

Kendra Hanley, MS

Director, Measure Specification Operations- PCPI
Chicago, Illinois

Amy Moyer

Manager of Value Measurement – The Alliance
Madison, Wisconsin

Allison Peel, DC, MHA, MPH, PMP

Sr. Program Manager- GDIT -HAFP
Des Moines, Iowa

Peter Robertson, MPA

Quality Measure Developer- American Academy of Ophthalmology
San Francisco, California

Patrick Romano, MD, MPH

Professor- UC Davis Health System
Sacramento, California

Federal Representative

Ann Page

US Department of Health and Human Services, ASPE

NQF Staff

Helen Burstin, MD, MPH

Chief Scientific Officer

Marcia Wilson, PhD, MBA

Senior Vice President
Quality Measurement

Elisa Munthali, MPH

Vice President
Quality Measurement

Debjani Mukherjee, MPH

Senior Director
Quality Measurement

Karen Johnson, MS

NATIONAL QUALITY FORUM

Senior Director
Quality Measurement

Andrew Lyzenga, MPP

Senior Director
Quality Measurement

Poonam Bal, MHA

Senior Project Manager
Quality Measurement

Jean-Luc Tilly

Project Manager
Quality Measurement