

# Variation in Measure Specifications: Sources and Mitigation Strategies

FINAL REPORT

DECEMBER 21, 2016



NATIONAL  
QUALITY FORUM

This report is funded by the Department of Health and Human Services under contract HHSM-500-2012-00009I, Task Order HHSM-500-T0020.

# CONTENTS

EXECUTIVE SUMMARY	2
INTRODUCTION	3
PURPOSE AND OBJECTIVES	4
EXPERT PANEL	4
METHODOLOGY	5
CONSENSUS DEFINITIONS OF KEY TERMS	9
VARIATION TAXONOMY: IDENTIFYING REASONS AND IMPACT	15
STRATEGIES FOR ADDRESSING VARIATION	18
A FRAMEWORK FOR PREVENTING AND MITIGATING VARIATION	20
NEXT STEPS	25
APPENDIX A: Real World Examples of Variation	27
APPENDIX B: Key Informant Interview Guide	29
APPENDIX C: Key Informants	30
APPENDIX D: Search Strategy	31
APPENDIX E: Expert Panel	33

## EXECUTIVE SUMMARY

The focus on quality and valuation of care quality has increased the need for measures and measurement in healthcare. As a result, performance measures have proliferated across a diverse range of clinical areas, settings, data sources, and programs.

Over time, experience in measurement activities has revealed that measure use is often not well-aligned across programs and that measure users sometimes vary the specifications of measures unnecessarily. The lack of consistency arising from unnecessary variation lessens the usefulness of measure results and poses challenges for providers, patients, health plans, regulators, and others who use measures to assess performance. This project defined variation, explored its causes and its effects, and provided guidance on how to mitigate or prevent variation.

The project comprised two phases. The first phase was a synthesis of available evidence on the phenomenon of variation in measure specifications and development of standard definitions of key terms used in measure development and implementation. Primary causes of variation were identified and organized in a taxonomy, and strategies and guiding principles to prevent or mitigate the negative consequences of variation were developed. The second phase combined all of these elements in a framework for measure developers and implementers. All this work was completed under the guidance of a multistakeholder Expert Panel which represented experts in measure development and implementation, health informaticists, provider groups, purchasers, payers, and others who use measure results.

The Expert Panel defined measure variation as any deviation from a reference measure's specifications. A lexicon was developed for related terms based on this foundational

concept. The Panel found that variation often related to data availability, measure complexity, and communication. As a result, strategies to mitigate or prevent variation focused on access to measures, data collection strategies, implementation guidance, benchmarking against set standards, transparency regarding variation introduced during implementation, as well as communication and collaboration through feedback loops and open forums.

The decision logic proposed in the report provides implementers, developers, regulators, and other entities that select measures with a methodical approach to improve the comparability and interpretability of measure results, while reducing the burden of duplicative measures.

The report also identifies future opportunities to pilot test and operationalize ways to reduce the incidence and impact of variation in measure specifications.

# INTRODUCTION

As the U.S. healthcare system has increasingly focused on 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 or significantly modified versions of the same measure. An analysis conducted by Bailit and Associates in 2013<sup>1</sup> identified 1,367 quality measures in use across 48 different state and regional programs. Only 509 were “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 overlapped or had a similar focus, with one or more variations in the specifications.

Many healthcare stakeholders view this lack of consistency as unnecessary variation that poses challenges for providers, patients, health plans, regulators, and others who use measures to assess performance. For clinicians, hospitals, and other healthcare providers, variation may increase data collection and reporting requirements without attendant increases in value. Variation may diminish the value of measurement results for those who use the results to inform decisions about and comparisons of healthcare. Specifically, changing measures in a way that compromises the comparability of results may prevent users from drawing accurate conclusions about the differential performance of what is being measured.

## PURPOSE AND OBJECTIVES

This project identified reasons for variation in measure specifications and the impact of such variation, as well as provided guidance on ways to mitigate or prevent variation. To do so, the project synthesized available evidence on the phenomenon of variation in measure specifications, particularly as applied in accountability programs. It then addressed the phenomenon: it proposed standard definitions of key terms used in measure development and implementation, identified the primary causes of variation in a taxonomy, developed strategies and guiding principles to prevent or mitigate negative consequences of variation, and combined all of these elements in a usable framework for measure developers and implementers.

Throughout this project, NQF has worked with a multistakeholder 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 factors contributing to variation and the implications of this variation. The taxonomy, along with feedback related to challenges faced by key informants, informed the development of 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 many facets of variation and developed practical guidance to help stakeholders to identify and address reasons for variation in measure specifications.

This final report synthesizes and summarizes all the findings and concepts considered by the Expert Panel and builds on the preliminary recommendations in the prior two draft reports. Key informant interview themes incorporated into the document illuminate practical considerations and reasons for variation. In addition, the Expert Panel reviewed and finalized the consensus definitions, taxonomy, and framework included in this report in light of additional feedback from member and public comments.

## EXPERT PANEL

Through a public call for nominations, NQF convened the multistakeholder Variation in Measure Specifications Expert Panel to accomplish the objectives of the project. The Panel included 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 E](#)).

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 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 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 previous two draft reports. To identify relevant literature, NQF staff used a combination of the following medical subject heading (MeSH) search terms: variation, change, measure specification, guideline implementation, 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, and measure collection and volume.

The MeSH searches and combination of search terms used led to a very small number of relevant articles. The search was limited to 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. 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.

[Appendix D](#) provides detailed information regarding the search strategy and search term groupings used to conduct the literature searches.

---

### 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 subset 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](#). The key informants interviewed represent a sample of those who develop as well as use measures. These interviews provided information used to enhance and corroborate themes from the literature search and are not necessarily exhaustive.

## 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 used.<sup>2</sup>

- **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 identify appropriate and correct 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.<sup>3</sup>

- **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 and receiving pharmacotherapy led to differences in facility rankings. This difference was as pronounced as 24 percentage points when changing the

measurement period from one year to two, and 29 percentage points when the measure was modified to focus on those receiving treatment for the first time.<sup>4</sup>

- **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.<sup>5</sup>

- **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.<sup>6</sup> Another study of more than 200 commercial health plans found that nearly three-quarters of plans had a greater than 10 percentage 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.<sup>7</sup> Similar research using measures of body mass index and immunization in children resulted in virtually identical findings.<sup>8</sup> A comparison of the EHR-based version and the claims-based version of one measure found that the claims-based measure significantly underperformed versus 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.<sup>9</sup>

- **Incomplete measure specifications.** An examination of 10 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.”<sup>10</sup>

- **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.<sup>11</sup>

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, or burden as a driver of 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.”<sup>12</sup>

---

### 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 and mitigate its impact.

- All of the key informants identified the availability, i.e. quantity and quality, of 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 depends on data availability, data completeness, 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. Limited data leads to variation arising from efforts focused on increasing the sample size and or completeness of data elements. However, when data are available but not readily accessible, collaborative efforts such as data sharing agreements can actually mitigate the need for variation by enabling access to existing data.

- 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 frontline 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 users’ interpretation of the measure. Additionally, individuals abstracting data from EHRs may lack training to do so; incorrect data aggregation and analysis also affects comparability and introduces another level of variation after implementation of measures. Some 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 concern most commonly cited 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 limited: a measure developer may acknowledge that a change was made to a measure but cannot describe the exact methodology of the variation without compromising proprietary information.

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

## 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 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 within the healthcare context. 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, 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 improving, but could be 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 three-point improvement versus a six-point improvement, or a three-point improvement versus a score less than five).
- **Different timeframes.** One is using a six-month timeframe and the other is using a 12-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 with 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 defined *measure variation* as *any deviation from a reference measure’s 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 the 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. 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
<b>Numerator</b>	<ul style="list-style-type: none"> <li>• Differences in definitions, coding, or documentation of clinical concepts (e.g., 'encounter', 'adherence', etc.)</li> <li>• Differences in performance thresholds or criteria</li> </ul>
<b>Denominator</b>	<ul style="list-style-type: none"> <li>• Differences in definitions, coding, or documentation of clinical concepts</li> <li>• Measure applied to an age group different from the age group in the reference measure specifications</li> </ul>
<b>Exclusions from denominator</b>	<ul style="list-style-type: none"> <li>• Differences in acceptable exclusions (e.g., specific medical conditions versus unspecified "medical reasons")</li> </ul>
<b>Risk adjustment</b>	<ul style="list-style-type: none"> <li>• Differences in variables included in risk-adjustment models</li> <li>• Adjustment for clinical factors only versus adjustment for clinical plus sociodemographic factors</li> <li>• Differences in risk-adjustment strategy (e.g., logistic versus hierarchical modeling)</li> </ul>
<b>Data source or collection instrument</b>	<ul style="list-style-type: none"> <li>• Use of administrative claims versus registry reporting</li> </ul>
<b>Care setting</b>	<ul style="list-style-type: none"> <li>• Measure intended to be applied to hospitals is applied to ambulatory care facilities</li> </ul>
<b>Level of analysis</b>	<ul style="list-style-type: none"> <li>• Measure intended to evaluate health plan performance is used to evaluate individual clinician performance</li> </ul>
<b>Attribution strategy</b>	<ul style="list-style-type: none"> <li>• Attribute performance to the provider most often seen versus a particular provider type such as family medicine/internist/GP</li> </ul>

## Key Definitions

### Alignment

Measures are aligned when they target the same outcome or care process within the same target population; aligned measures may contain variations in their specifications or calculation.

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 include numerator, denominator, exclusions, calculation, data source, and collection instructions. The extent (i.e., degree) 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). Standardization is accomplished by degrees, depending on what is achievable or desirable with the ultimate goal being complete standardization where the specifications are identical.

The definition provided above has been adapted from NQF's 2010 [Guidance for Measure Harmonization](#), a consensus report establishing a definition and approach to harmonization.

### Reference Measure

A reference measure is the “parent” measure from which a variant has been created. If the measure was not directly created as a modification of an existing measure, the reference measure is an endorsed measure that predates the variant and targets the same outcome or care process in the same target population, and thus serves as a de facto parent.

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 multistakeholder consensus-based entity using 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 multistakeholder consensus-based entity using 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 multistakeholder consensus-based entity using 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, if 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 a 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. Panel members 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 otherwise be 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.”<sup>13</sup> 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 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, [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.”<sup>14</sup>

Possible patient-related factors include clinical severity, conditions present at the start of care and socioeconomic as well as sociodemographic 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., medication, administered). 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 about which specifications have changed
- Showing the impact of the changes on the measure result

---

## 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 steward’s 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) all contribute 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 explicitly parse 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 AND IMPACT

A classification system for identifying variation and assessing its effects was developed based on feedback from the Expert Panel and key informants, along with public comments received during two separate commenting periods. As a preliminary step toward developing a framework for assessing 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 and stages in the measure lifecycle.

**Intended use.** The significance of variation substantially depends on whether measures are being used for internal quality improvement (QI) or accountability purposes. If a measure is modified by a healthcare provider for its 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 focus on measure variation intended for accountability purposes only.

**Stages in the measure lifecycle.** Measure variation can present at any stage 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.

## Reasons for Variation

In order to understand and address variation, one must understand why developers or implementers create variants. Both Panel members and key informants agreed that there are numerous 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 that provider's 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 results when developers and implementers unknowingly create 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. Some 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**

---

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 versus 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 the 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 be able to capture the measure as specified while another may not.

---

### **Need to Coordinate Multiple External Requests for Measures**

---

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 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.

### **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 in most cases. However, they did acknowledge that some forms of variation are beneficial and warranted, such as

those introduced through innovation. This section discusses the potential impact of variation.

---

### **Innovation**

---

Variation is not always 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. All parties experience this burden: 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 gathering and reporting on variants of quality measures can 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 assess accurately and compare the 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 outweigh the costs.

## Preventing 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 existing measures as well as those 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 the prevention of duplication of efforts.

**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 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 clarification and communication of measure-specific needs. 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 respond to end-user needs as well as be aware of measure variants created by end-users. The bidirectional 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.** Access to precise, unambiguous, and complete specifications should be available for all measure implementers to reduce variation. Measure stewards should respond to any requests for technical assistance and/or clarifications from measure implementers.

**Data collection strategies.** 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 abstraction 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.

**Data auditing.** Auditing can identify and address variation through measure compliance reviews, which may include assessment of reliability of the data source, coding, data abstraction, and inter-rater reliability of abstractors. For example, The HEDIS audit ensures a fair comparison of organizations for several programs, such as, accreditation, value-based initiatives, and the Centers for Medicare & Medicaid Services (CMS) Star Ratings Program. According to NCQA, the effectiveness of audits should be maximized through simultaneous data collection, which allows the auditor to detect errors in the data collection process while there is time for the organization to correct its methods and minimize the possibility of biased rates. The Expert Panel noted that requiring audits could potentially reduce variation by focusing attention on the reliability of data collection during measure implementation and reporting.

## Mitigating Variation

**Feedback loops.** Communication is fundamental to receiving clarifications and current measure-related information, and feedback loops can both mitigate and prevent variation. Sometimes measure specifications are changed because existing measures do not address end-use needs. When the steward of a measure learns why an implementer has introduced a variant, this can benefit the steward and the implementer. It can improve understanding of whether a change in specifications is relevant and/or necessary, and it can inform future 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 information may benefit other implementers who may be struggling with the same or similar issues. In addition, disclosing changes to specifications 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. Transparency makes users and consumers 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 “workarounds” that minimized variations or other 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 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. Figure 2 presents these decision points in a logic diagram that guides the user in deciding whether or not variation is needed and how to mitigate the quality-measurement related consequences 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, identical or harmonized 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.

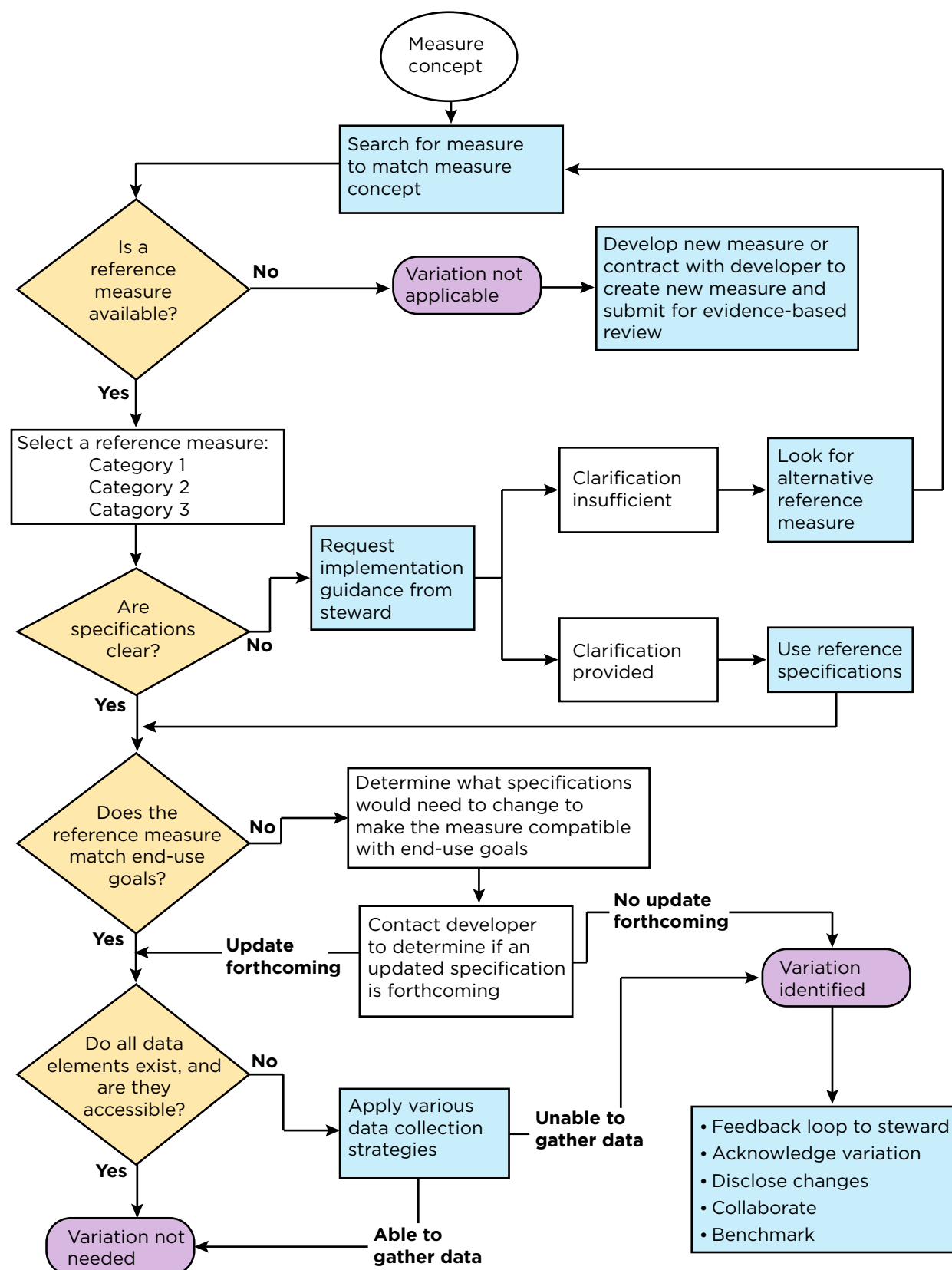
**Specificity.** To ensure consistency in implementation, measures used for accountability should include precise, unambiguous, and complete specifications that minimize the need for interpretation or additional specification by measure users.

**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 respond 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.

FIGURE 1. FRAMEWORK FOR PREVENTING AND MITIGATING VARIATION





## Using the Framework to Address Variation

The Framework presented here will serve as a guide for those seeking to prevent and mitigate the effects of measure variation, including parties selecting measures for use, measure developers and/or stewards, and regulators using measure results. 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 and justified, 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 a measure 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 measures 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 implement the measure confidently. 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:

- The measure can be improved to increase reliability and validity, such as incorporating risk adjustment in order to accurately capture provider performance.
- 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.
- 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. These strategies include 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 unfeasible, users should use the same set of mitigation strategies mentioned for measures that do not meet user end-use goals.

## NEXT STEPS

This framework report explores opportunities to reduce the incidence and impact of variation in measure specifications. Implementers, developers, regulators, and other entities that select measures are provided with an approach to improve the comparability and interpretability of measure results, while reducing the burden of duplicative measures. In order to execute on the guiding principles and framework detailed in this report, a series of potential next steps is recommended as follows.

**Model best practices.** Development of best practice recommendations will facilitate the effective use of the decision logic to address variation so that users can easily adopt and follow the decision logic at the point of measure implementation. The best practice recommendations will relate to measure specification descriptions, level of depth and accuracy of implementation guidance, disclosure of changes when the measure is a variant, and attestation when measure specifications are implemented without varying the measure steward's specifications. These best practices will also elucidate the implementation challenges that may lead to variation as well as suggest ways to prevent and/or minimize implementation challenges. Issues considered could include various permutations of the following: measure use at inappropriate settings based on the measure's level of analysis, a lack of transparency

and collaboration among measure developers, sharing of measures in development, and fostering economies of scale in measure development. In addition, pilot testing of the decision logic can identify strategies to reduce variation.

**Develop a repository of measure variants and measures under development.** In order to increase transparency, the quality measurement enterprise would benefit from a measure repository that includes a comprehensive database of measures under development and measure variants. This repository would be very useful to reduce variation of measures under development and selected for use.

**Adapt NQF's policies and procedures to address variation.** As the consensus-based entity charged with endorsing healthcare performance measures, NQF has a role to play in reducing variation through limiting the number of variants that receive endorsement. NQF has worked to harmonize measure variants as measures are evaluated in the Consensus Development Process. However, the assessment is likely too late in the development process. NQF will consider opportunities to evolve the measure endorsement and selection processes to consider new approaches to related and competing measures. Development of best practice recommendations will pilot test and facilitate the operationalization of the decision logic to address variation.

## ENDNOTES

---

- 1 Bazinsky K, Bailit M. *The Significant Lack of Alignment Across State and Regional Health Measure Sets*. Needham, MA: Bailit Health Purchasing, LLC; 2013. Available at [http://www.bailit-health.com/articles/091113\\_bhp\\_measuresbrief.pdf](http://www.bailit-health.com/articles/091113_bhp_measuresbrief.pdf). Last accessed October 2016.
- 2 Sanf  lix-Gimeno G, Franklin JM, Shrank WH, et al. Did HEDIS get it right? Evaluating the quality of a quality measure: adherence to B-blockers and cardiovascular outcomes after myocardial infarction. *Med Care*. 2014;52(7):669-676.
- 3 Needleman J, Buerhaus PI, Vanderboom C, et al. Using present-on-admission coding to improve exclusion rules for quality metrics: the case of failure-to-rescue. *Med Care*. 2013;51(8):722-730.
- 4 Fernandes-Taylor S, Harris AH. Comparing alternative specifications of quality measures: access to pharmacotherapy for alcohol use disorders. *J Subst Abuse Treat*. 2012;42(1):102-117.
- 5 Pogach LM, Rajan M, Maney M, et al. Hidden complexities in assessment of glycemic outcomes: are quality rankings aligned with treatment? *Diabetes Care*. 2010;33(10):2133-2139.
- 6 Kern LM, Malhotra S, Barr  n Y, et al. Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study. *Ann Intern Med*. 2013;158(2):77-83.
- 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, et al. Variation in outcomes of quality measurement by data source. *Pediatrics*. 2014;133(6):E1676-E182.
- 9 Tang PC, Ralston M, Arrigotti MF, et al. Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures. *J Am Med Inform Assoc*. 2007;14(1):10-15.
- 10 Amster A, Jentzsch J, Pasupuleti H, et al. Completeness, accuracy, and computability of National Quality Forum-specified eMeasures. *J Am Med Inform Assoc*. 2015;22(2):409-416.
- 11 Mathioudakis N, Golden SH. A comparison of inpatient glucose management guidelines: implications for patient safety and quality. *Curr Diab Rep*. 2015;15(3):13.
- 12 Meyer GS, Nelson EC, Pryor DB, et al. More quality measures versus measuring what matters: a call for balance and parsimony. *BMJ Qual Saf*. 2012;21(11):964-968.
- 13 Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*, Washington, DC: National Academy Press; 2001. Available at <http://www.nationalacademies.org/hmd/Global/News%20Announcements/Crossing-the-Quality-Chasm-The-IOM-Health-Care-Quality-Initiative.aspx>. Last accessed October 2016.
- 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](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
<b>Definitions of clinical concepts and/or terms</b>	<ul style="list-style-type: none"> <li>• Encounter</li> <li>• Adherence</li> <li>• Care transition</li> </ul>	<ul style="list-style-type: none"> <li>• Lack of common or standardized definitions</li> </ul>	<ul style="list-style-type: none"> <li>• 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.</li> <li>• 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.</li> <li>• 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 versus a clinician's estimate of gestational age).</li> </ul>
<b>Coding or documentation of clinical concepts</b>	<ul style="list-style-type: none"> <li>• Variation in codes, fields, or problem sets used to indicate a clinical condition</li> <li>• Granularity of information</li> </ul>	<ul style="list-style-type: none"> <li>• Differences in available fields (e.g., claims versus registry)</li> </ul>	<ul style="list-style-type: none"> <li>• 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 data being reported depending on whether the registry/claims-based measure or the eMeasure was being used.</li> </ul>
<b>Changes in implementation</b>	<ul style="list-style-type: none"> <li>• Time intervention</li> <li>• Exclusions</li> </ul>	<ul style="list-style-type: none"> <li>• Updated evidence</li> <li>• Adaptation of measure for implementation</li> </ul>	<ul style="list-style-type: none"> <li>• Implementation of sepsis bundle: includes changes to the timeframe for the administration of crystalloid fluids for septic shock patients (i.e., 3 hours for CMS-adapted measure versus 6 hours 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).</li> </ul>

Type of variation	Examples	Potential reasons for variation	Real world examples
<b>Risk adjustment models/factors</b>	<ul style="list-style-type: none"> <li>• SDS factors added to clinical risk-adjustment model</li> <li>• Address risk arising from social and other vulnerabilities</li> </ul>	<ul style="list-style-type: none"> <li>• Different approaches to risk adjustment</li> </ul>	<ul style="list-style-type: none"> <li>• 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 publicly reports SDS-adjusted measure results on its website (along with results of the measure as calculated by CMS and unadjusted rates).</li> </ul>
<b>Target population</b>	<ul style="list-style-type: none"> <li>• Changing target population based on data needs and/or intent/purpose of measurement</li> </ul>	<ul style="list-style-type: none"> <li>• Program needs (e.g., measure being used as part of a pediatric quality initiative)</li> </ul>	<ul style="list-style-type: none"> <li>• A readmissions measure specified for patients of all ages being applied to the Medicare population.</li> </ul>
<b>Level of analysis or attribution</b>	<ul style="list-style-type: none"> <li>• Change in attribution strategy</li> </ul>	<ul style="list-style-type: none"> <li>• Program needs (e.g., measure being used as part of a clinician quality initiative)</li> <li>• Data availability</li> </ul>	<ul style="list-style-type: none"> <li>• 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.</li> </ul>
<b>Data source</b>	<ul style="list-style-type: none"> <li>• National registry versus chart data</li> <li>• Electronic versus manual abstraction</li> </ul>	<ul style="list-style-type: none"> <li>• Availability of data</li> <li>• Compliance or reporting requirements</li> </ul>	<ul style="list-style-type: none"> <li>• 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.</li> </ul>

## APPENDIX B:

### Key Informant Interview Guide

#### Variation in Measure Specifications

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?
  - a. 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: Key Informants

### **Washington Health Alliance**

Susie Dade, Deputy Director

### **Aetna**

Patricia McDermott, Sr. Operations/Program Manager

### **Consumer Reports**

John Bott

### **NextGen**

Dr. Sarah Corley, Chief Medical Officer

### **QCorp**

Betsy Boyd Flynn, Program Director

Meghan Haggard, Program Manager

Meredith Roberts Tomasi, Program Director

Cindi McElhaney, Senior Health Care Analyst

### **Centers for Medicare & Medicaid Services (CMS)**

Karen Matsuoka, PhD, CMCS Chief Quality Officer and  
Director, Division of Quality and Health Outcomes

Barbara Dailey, Deputy Director, Division of Quality  
and Health

Megan Thomas, Center for Medicaid and CHIP Services

## APPENDIX D:

### Search Strategy

This search strategy appendix outlines the MeSH definitions available during the searches. Please note that the MeSH definition for variation differs from the one used by the project.

#### Links to MeSH definitions

<https://www.ncbi.nlm.nih.gov/mesh/?term=variation>

<https://www.ncbi.nlm.nih.gov/mesh/?term=measure>

<https://www.ncbi.nlm.nih.gov/mesh/?term=measure+quality>

#### Selected Search Strategies

This section details the search strategies used to conduct the literature searches. Once these searches were completed, staff went through the search results to identify articles that were relevant to the project.

##### **Variation and measure specification:**

(variation[All Fields] AND ("weights and measures"[MeSH Terms] OR ("weights"[All Fields] AND "measures"[All Fields]) OR "weights and measures"[All Fields] OR "measure"[All Fields]) AND specification[All Fields]) AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and guideline implement:**

(variation[All Fields] AND ("guideline"[Publication Type] OR "guidelines as topic"[MeSH Terms] OR "guideline"[All Fields]) AND implement[All Fields]) AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and "Healthcare Effectiveness Data and Information Set":**

variation[All Fields] AND "Healthcare Effectiveness Data and Information Set"[All Fields] AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and Hedis:**

(variation[All Fields] AND Hedis[All Fields]) AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and "performance measure":**

variation[All Fields] AND "performance measure"[All Fields] AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and "quality measure":**

variation[All Fields] AND "quality measure"[All Fields] AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and "quality metric":**

variation[All Fields] AND "quality metric"[All Fields] AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and "quality improvement":**

variation[All Fields] AND "quality improvement"[All Fields] AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and "quality improvement and measure":**

variation[All Fields] AND "quality improvement"[All Fields] AND ("weights and measures"[MeSH Terms] OR ("weights"[All Fields] AND "measures"[All Fields]) OR "weights and measures"[All Fields] OR "measure"[All Fields]) AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and "implementation science":**

variation[All Fields] AND "implementation science"[All Fields] AND ("humans"[MeSH Terms] AND English[lang])

##### **Variation and "translational science":**

variation[All Fields] AND "translational science"[All Fields] AND ("humans"[MeSH Terms] AND English[lang])



**Variation and “point-of-care” measure:**

variation[All Fields] AND "point-of-care"[All Fields] AND ("weights and measures"[MeSH Terms] OR ("weights"[All Fields] AND "measures"[All Fields]) OR "weights and measures"[All Fields] OR "measure"[All Fields]) AND ("humans"[MeSH Terms] AND English[lang])

**Variation and “point-of-care” measure and quality:**

variation[All Fields] AND "point-of-care"[All Fields] AND (("weights and measures"[MeSH Terms] OR ("weights"[All Fields] AND "measures"[All Fields]) OR "weights and measures"[All Fields] OR "measure"[All Fields]) AND quality[All Fields]) AND ("humans"[MeSH Terms] AND English[lang])

**Variation and measure and collect And volume:**

(variation[All Fields] AND ("weights and measures"[MeSH Terms] OR ("weights"[All Fields] AND "measures"[All Fields]) OR "weights and measures"[All Fields] OR "measure"[All Fields]) AND collect[All Fields] AND volume[All Fields]) AND ("humans"[MeSH Terms] AND English[lang])

---

**Non-MeSH Basic Searches Using Google**

---

Performance Measure AND Variance

Measure AND Change

Measure AND Quality Metrics AND Variation AND Translation Science

Guideline AND Performance Measure AND Implementation

## APPENDIX E: 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, Electronic Clinical Quality Measures (eCQMs),  
Health Services Advisory Group, Inc.  
Chicago, Illinois

**Amy Moyer**

Manager of Value Measurement, The Alliance  
Madison, Wisconsin

**Allison Peel, DC, MHA, MPH, PMP**

Sr. Program Manager, General Dynamics Information  
Technology  
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**

U.S. 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

**Karen Johnson, MS**

Senior Director

**Andrew Lyzenga, MPP**

Senior Director

**Poonam Bal, MHA**

Senior Project Manager

**Jean-Luc Tilly**

Project Manager

NATIONAL QUALITY FORUM  
1030 15TH STREET, NW, SUITE 800  
WASHINGTON, DC 20005  
[www.qualityforum.org](http://www.qualityforum.org)