NATIONAL QUALITY FORUM

Variation in Measure Specifications Project 2015-2016

FIRST DRAFT REPORT

April 19, 2016

This report is funded by the Department of Health and Human Services under contract HHSM-500-2012-00009I 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 also 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 being used in various programs (e.g., at the federal, state, and community levels) are often not well-aligned. For example, different programs, intending to address the same fundamental quality issue, frequently will use slightly different or modified versions of the same measure. An analysis conducted by Bailit and Associates in 2014 demonstrated that there were 1,367 quality measures in use across 48 different programs. An examination of the "distinct" measures across all of these programs, after removing all duplicates, revealed that there were only 509 distinct measures in use and more than 800 measures that were overlapping or similar in focus. The analysis found that when these programs shared measures, they often modified the existing specifications for these measures.

This lack of consistency across measurement programs creates what many view 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, as the measures address the same area of focus. For those seeking to use measure results to inform decisions about healthcare, variation may diminish the value of measurement data. Specifically, if measures have been changed in a way that compromises the comparability of measurement results, users may not be able to draw accurate conclusions about the differential performance of those being measured.

Purpose and Objectives

Through this project, NQF will identify how, where, and why variation in measure specification is occurring; create a framework for understanding and interpreting the different types of variation and the implications of this variation; and develop a common understanding around key terms, concepts, and measure components to help standardize measurement efforts and minimize unnecessary variation. Through the use of an environmental scan, an expert panel, and key informant interviews, the project will explore the many facets of variation and develop practical guidance and recommendations to help stakeholders in identifying, understanding, and addressing variation in measure specifications. A timeline and key activities of the project are included in Appendix D.

This report presents preliminary findings and concepts under consideration by the Expert Panel. Building on the evolving environmental scan and additional Panel deliberations, these findings will be refined and further developed over the course of the project; a second report will present the final results of this work.

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, provider groups representing those who are being measured, purchasers, health plans, and others who use measure results (see Appendix C).

The Panel will provide 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.

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. This project will help to clarify definitions, principles, and types of variation.

This work should be a useful adjunct to multiple efforts focused on development of core measures For example, the Institute of Medicine's Vital Signs report highlighted the need for coordination in measurement efforts and called for a set of core metrics that could be used consistently across programs.[1] In addition, a collaborative effort by America's Health Insurance Plans (AHIP), the Centers for Medicare and Medicaid Services (CMS), NQF, and other groups recently proposed core measure sets across several topic areas to promote alignment and harmonization of measure use and collection across payers in both the public and private sectors.[2]

Variation in measure specifications

Even when there is alignment across programs at a conceptual level, the measures used in these programs may vary by specific constructs, including definitions of clinical concepts, target populations, and risk-adjustment strategies.

As an illustration, two programs may both 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 PROMIS mental health;
- Different scales: 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 18 and older while the other is including all ages.

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

Definitions

To ensure that the Expert Panel as well as the wider healthcare quality field could have a common understanding of concepts relevant to variation in measure specifications, the Expert Panel has proposed definitions of some key terms related to variation.

Measure variation:

The Expert Panel defines **measure variation as *any deviation from a fixed set of reference measure specifications*.**

This definition recognizes that, for practical purposes, measure variation cannot be identified or assessed without first identifying an accepted 'reference' set of specifications 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 standardized sets (e.g., NQF-endorsed measures, HEDIS measures) as common reference points.

This definition includes different types of deviation 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 for their own purposes. Alternatively, an organization may implement a measure in a way that is inconsistent with the letter or intent of the original specifications, resulting in a *de facto* change to the measure specifications even if there was not a conscious decision made by the implementing organization to modify the original measure. Each of these scenarios qualifies as a measure variation under the definition proposed above.

NQF consensus-based technical definitions for **harmonization** and **alignment**, taken from NQF Guidance for Measure Harmonization: A Consensus Report, were provided to the Expert Panel for consideration:

> **Alignment**: Encouraging the use of similar, standardized performance measures across and within public and private sector efforts.
>
> **Harmonization**: The standardization of specifications for related measures with the same measure focus (e.g., influenza immunization of patient 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.

The Expert Panel considered these definitions during their deliberations and discussed their relevance and applicability to variation in measure specifications. The Panel will consider these definitions in the context of greater precision regarding measure variation.

Environmental Scan

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

Methodology

Literature Review

The literature review portion of the environmental scan will include a review of both peer-reviewed published literature and non-peer reviewed (i.e. gray) literature from the past 5 to 10 years. For a preliminary literature review, NQF staff conducted MeSH searches using a combination of the following search terms: variation, change, guideline implementation, HEDIS, measure, performance, quality, metrics, methodology, and quality improvement. The MeSH searches led to a very small number of relevant articles. Staff identified additional articles by conducting a Google search using simple terms such as “measure” and “change”. It should be noted that the peer-reviewed and published literature on this topic is limited. This initial literature review identified only 56 published articles and reports (peer-reviewed and gray), many of which were only tangentially related to measure variation. A fuller literature review is in progress, informed by input from Expert Panel members and leads identified during the preliminary review. Feedback from the public comment period will also serve to inform a more comprehensive scan for information relevant to measure variation. However, the absence of literature identified in the preliminary review underscores the importance of “real-world” experiential data, which will be sought through examples and case-studies from measure implementers, payers, purchasers, providers, and federal partners.

Expert Panel Input

The information gathered through the initial literature review was presented to the Expert Panel for discussion and input during its first in-person meeting. The preliminary literature review will be refined and expanded based on Expert Panel input. The Panel also provided practical examples that exemplified and elaborated on the extent and impact of variation (see Appendix A).

Key Informant Interviews

The next phase of the project will involve interviewing key informants who represent a wide variety of healthcare stakeholders and perspectives. Key informants will be interviewed using a common set of questions that are modified to focus on and gather information specific to each interviewee's healthcare sector niche. Potential interviewees may include providers, payers, measure implementers, consumer representatives, federal representatives, and electronic health/medical record vendors.

Preliminary Findings – Literature Review

Over 50 potentially-relevant articles were identified in a preliminary literature review. However, the body of literature on this topic was relatively limited, and findings were sometimes contradictory. Selected findings, including studies related to the impact of different types of variation, are presented below; the literature review will be expanded and enhanced with additional guidance from the Expert Panel and other stakeholders.

- *Different measure definitions*: A study evaluated 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 when 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.[3]
- *Different exclusion rules:* The study assessed different exclusion rules to identify whether complications were present on admission or hospital-acquired. The different 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.[4]
- *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 are receiving pharmacotherapy led to differences 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[5].
- *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.[6]
- *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.[7] 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.[8] Similar research was conducted with measures of body mass index and immunization in children, with virtually identical findings.[9] 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. [10]

- *Incomplete measure specifications*: An examination of ten NQF-endorsed eMeasures found that 'literal implementation of specifications was not feasible 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'.[11]
- *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.[12]

Initial Components of Framework

With the aim of developing a conceptual framework for identifying and assessing measure variation, the Expert Panel, collectively, discussed the general phenomenon of variation as well as specific examples of variation offered by Panel members. A number of potential approaches to categorizing measure variation emerged from this discussion.

Measure specifications

Following the definition of measure variation presented above, variations are manifested as either formal or *de facto* changes to specifications of a given reference measure. For this reason, Expert Panel members suggested that it may be useful to 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 could potentially be varied is presented in Table 1, along with examples of variation that might occur in each category.

Table 1: Examples of variation in measure specifications

Measure specification element	Example of variation
Numerator/measure focus	• Differences in definitions, coding, or documentation of clinical concepts (e.g., 'encounter', 'adherence', etc.) • Differences in performance thresholds or criteria
Denominator/target population	• Differences in definitions, coding, or documentation of clinical concepts • Measure intended for adults applied to pediatric population
Exclusions from denominator/	• Differences in acceptable exclusions (e.g., specific medical

target population	conditions vs. unspecified "medical reasons")
Risk adjustment	• Differences in variables included in risk adjustment models • 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	• Use of administrative claims vs. registry reporting
Care setting	• Measure intended to be applied to hospitals is applied to ambulatory care facilities
Level of analysis or attribution strategy	• Measure intended to evaluate health plan performance is used to evaluate individual clinician performance

Intended use

The significance of variation depends substantially on whether measures are being used for internal quality improvement (QI) or accountability[a] 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 likely will be different 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.

[a] For the purposes of this project, we are defining accountability as the use of measure results for public reporting, payment, or other external decision-making purposes.

Type of variation

The Expert Panel noted that in order to fully understand variations and determine how to address them, it is necessary to understand why those variations are occurring. Panel members agreed that there are a wide range of types and causes of variation, but that many could be grouped into three overarching categories:

Formal 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). Or, for quality improvement purposes, a provider may want to tighten the timeframe of a specific measure to drive change within their organization. This type of variation is done deliberately to address a particular need or preference of the measure user.

Variation arising from incomplete or ambiguous measure specifications or a lack of operational guidance

Gaps in specifications were an area of great concern for many Expert Panel members, particularly 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. In such instances, measure users may "fill in gaps" in specifications by creating their own definitions of concepts or developing their own interpretations of vague specifications. This type of variation is done out of necessity, as the measure user does not consider the measure to be usable without modification (i.e., additional detail).

Variation due to implementation challenges (e.g. data or resource limitations)

The Expert Panel noted that many instances of variation occur because of implementation challenges, such as the type of data a measure implementer has access to (e.g., registry data vs. data from administrative claims), or differences in capabilities across 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). Panel members observed that even within a single health system, different EHR systems or versions of an EHR system may be used, and one version may have the capacity to capture the measure as specified while another does not.

Guiding Principles for Considering Variation

As a preliminary step toward developing a model for assessment of variation and its effects, the Expert Panel identified a number of principles that may serve as guidance when considering a given instance of variation or the concept of variation in general:

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.

Next Phase of Project Work

The next phase of the project will involve refinement and expansion of the environmental scan, including interviews with key informants who represent a wide variety of perspectives. The Expert Panel also will continue to consider potential frameworks for understanding and interpreting measure variation.

References

[1] IOM (Institute of Medicine). 2015. *Vital signs: Core metrics for health and health care progress.* Washington, DC: The National Academies Press.

[2] https://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-releases/2016-Press-releases-items/2016-02-16.html

[3] Sanfélix-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.0000000000000148.

[4] 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.

[5] 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.

[6] 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.

[7] Kern LM, Malhotra S., Barrón Y., Quaresimo J., Dhopeshwarkar 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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/amiajnl-2014-002865.

[12] 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.

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	• Encounter • Adherence • Care Transition	• Lack of common or standardized definitions	• 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	• Variation in codes, fields, or problem sets used to indicate a clinical condition • Granularity of information	• Data source change • Differences in available fields (e.g., claims vs. registry)	• 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.
Changes in implementation	• Time intervention • Exclusions	• Updated evidence • Adaptation of measure for	• 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.,

Type of variation	Examples	Potential reasons for variation	Real World Examples
		implementation	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	• SDS factors added to clinical risk-adjustment model	• Different approaches to risk adjustment	• 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	• Changing target population based on data needs and or intent/ purpose of measurement	• Program needs (e.g., measure being used as part of a pediatric quality initiative)	• A readmissions measure specified for patients of all ages being applied to the Medicare population.
Level of analysis or attribution	• Change in attribution strategy	• Program needs (e.g., measure being used as part of a clinician quality initiative) • Data availability	• HEDIS breast cancer screening measure attributes patients to measured clinicians by including patients who have *any* 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 *primary care* visit with the measured clinician.

Type of variation	Examples	Potential reasons for variation	Real World Examples
Data source	• National registry versus chart data • Electronic vs. manual abstraction	• 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 Categories

Types of Key Informants:

1. Payers/Insurance
2. Developers
3. Implementers
4. Federal Liaisons
5. Providers
6. EHR/EMR vendors

Appendix C: Expert Panel

Panel Co-Chairs

Andrew Baskin, MD
National Medical Director- Aetna
Philadelphia, Pennsylvania

Blackford Middleton, MD, MPH, MSc
Instructor - Harvard T.H. Chan School of Public Health
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 Elliot, 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

Appendix D: Timeline and Schedule of Deliverables

Variation Timeline and Schedule of Deliverables (15 month POP: 9/30/15 - 12/30/16)

Task	Timeframe
Call for Nominations	10/23/15-12/4/15
Expert Panel Roster comment period additional outreach if needed.	1/6/16-1/19/16
Adjudication of roster comments	1/20/16-1/26/16
Expert Panel Orientation- **Webinar #1**	1/22/16 2:00-4:00 PM ET
Post final roster	1/27/16
Host public facing Web Meeting - **Webinar #2**	1/28/16 2:00-4:00 PM ET
In-Person Meeting #1 (1 day)	2/23/16 All Day
Webinar #3 – Co-Chair Review of Draft Report	3/31/16 2:00-3:00 PM ET
Public Commenting on Draft Report	4/19/16- 5/18/16
Key Informant Interviews	4/20/16- 5/31/16
Adjudicate comments, prep for webinar	5/5/16-5/24/16
Webinar #4 – Review Comments and Initiate Next Set of Activities	5/25/16 2:00-4:00 PM ET
Deliverable: First Draft Report to CMS	6/3/2016
In-Person Meeting #2 (1 day)	6/29/16 All Day
Webinar #5 – Review Comments and Refine/Reconcile Reports	9/8/16 2:00-4:00 PM ET
Public and member commenting- Draft #2	9/30/16-10/29/16
Deliverable: Second Draft Report to CMS	9/30/2016
Expert Panel post commenting: **Webinar #6**	11/3/16 2:00-4:00 PM ET
CSAC Review	11/9/16-11/10/16
Final report to CMS	12/21/16