Technical Expert Panel on Electronic Health Record Data Quality Best Practices for Increased Scientific Acceptability

DRAFT REPORT FOR PUBLIC COMMENT

This report is funded by the Centers for Medicare and Medicaid Services under contract HHSM-500-2017-00060I HHSM-500-0000 for Task Order 75FCMC19F0004.

Executive Summary

The Electronic Health Record (EHR) Data Quality Best Practices for Increased Scientific Acceptability project aims to identify challenges that impact the development, endorsement, and implementation of healthcare performance measures that use EHR data and recommend actions for mitigating those challenges. For the purposes of this project, "data quality" refers to how well EHR data support clinical quality measurement, including electronic clinical quality measures (eCQMs) as well as other EHR-sourced measures (i.e., clinical quality measures that use data derived from EHRs).

NQF convened a multistakeholder technical expert panel (TEP) to inform the development of two reports: an environmental scan report that identifies the current-state of EHR data quality issues and this final recommendations report. The environmental scan report summarizes existing literature and TEP guidance to describe challenges related to implementation of eCQMs, unstructured data, and NQF endorsement, as well as existing frameworks for assessing EHR and guidance from standard-setting bodies. This final recommendation report identifies solutions to address major challenges from the environmental scan, broken down by key phases or processes of the measure lifecycle: development, endorsement, and implementation.

To address measure development challenges, the TEP suggested opportunities for broadening the availability and access to electronic data, as well as more cross-agency interactions and potential federal initiatives around national testing collaborative and test bed efforts. The TEP recommended that the Department of Health and Human Services (HHS) should offer credit to providers and health IT vendors in federal programs to support measure development and create recognition programs to support measure development efforts, and that CMS should consider developing more measures that align across multiple care settings and programs.

Through its measure evaluation criteria and Consensus Development Process (CDP), NQF endorses measures for use in accountability programs and public reporting. The TEP offered several recommendations to NQF that aim to address challenges during the measure endorsement process: The NQF Scientific Methods Panel should develop eCQM-specific guidance for evaluating scientific acceptability of eCQMs; NQF should assess and update (if necessary) both the measure evaluation criteria and the measure evaluation process, with a focus on improving clarity and eliminating conflicting criteria for eCQMs; and NQF should provide updated information to standing committees and measure developers.

For the measure implementation phase, the TEP recommended that the Centers for Medicare and Medicaid Services (CMS) and/or the Office of the National Coordinator for Health Information Technology (ONC) should establish and award grants to help vendors who serve post-acute care (PAC) settings and other specialty areas hire dedicated staff members to incorporate eCQMs and EHR-sourced measures into their products.

The TEP also identified opportunities that warrant future consideration. Opportunities exist to articulate the cost and return on investment for supporting measure testing, the curation of a data element catalogue, and the utilization of existing user groups. Another significant opportunity is the use of pilot projects involving Fast Healthcare Interoperability Resources (FHIR), Observational Medical Outcomes

Partnership (OMOP), the CMS PACIO project, and the Health Level Seven International (HL7) Gravity project. There should be continued development and use of existing and new frameworks related to EHR data quality, such as the NQF Feasibility Scorecard and the FHIR Maturity Model. The TEP also discussed opportunities related to expanding the concept of hybrid measures beyond EHR and claims data to include manually and electronically abstracted data.

Introduction

One of the promises of electronic health records (EHRs) is that they enable automated clinical quality measure reporting. EHR systems are primarily designed to support patient care and billing, not necessarily capture additional data for primarily to support quality measurement.[1] However, since EHR data routinely collected for patient care can be used for clinical quality measures, they can be reused to reduce provider burden associated with public reporting and value-based purchasing programs.[2,3] Despite high adoption rates in multiple care settings, the promises of EHRs haven't yet been fully realized because of considerable variation in data quality, due to a number of factors as described below.

The National Quality Forum (NQF) defines electronic clinical quality measures (eCQMs) as measures that are specified using the industry accepted eCQM technical specifications, which include but are not limited to health quality measure format (HQMF), the Quality Data Model (QDM), Clinical Quality Language (CQL), and value sets vetted through the National Library of Medicine's Value Set Authority Center (VSAC).[4] Using EHRs as a source of data, eCQMs were designed to enable automated reporting of measures using structured data. With the use of structured data, eCQMs have the potential to provide timely and accurate information pertinent to clinical decision support and facilitate timely and regular monitoring of service utilization and health outcomes.[5] Currently, NQF has endorsed nearly 540 healthcare performance measures with only 34 of these being eCQMs. Although the number of endorsed eCQMs is low, several measures in NQF's portfolio are quality measures that rely on data that come from an EHR, which NQF refers to as EHR-sourced measures. NQF has identified the ability of EHR systems to connect and exchange data as an important aspect of quality healthcare. However, eCQMs and EHR data are not enough to enable automated quality measurement. To better understand the potential of improving quality measurement with the use of EHR data for clinical quality measures, it is important to examine the current state of EHR data quality.

In November 2019, under a contract with the Centers for Medicare and Medicaid Services (CMS), NQF convened a multistakeholder technical expert panel (TEP) over a series of web meetings to identify the causes, nature, and extent of EHR data quality issues, particularly as they relate to measure development, endorsement, and implementation. This multistep effort was aimed at identifying a set of strategies for addressing issues hindering EHR data quality and also focused on how well EHR data can be used to support automated clinical quality measurement.

Prior to convening the EHR Data Quality TEP, NQF performed an environmental scan to identify currently available information on EHR data quality issues, current efforts to address these issues, and key stakeholders' perspectives and input based on their experiences. The current state assessment from the environmental scan set the foundation for the development of this report, which offers recommendations on how to advance EHR data in ways that better support the development, endorsement, and implementation of eCQMs.

Through an open and transparent nomination process, TEP members were selected to ensure representation from a variety of stakeholders, including experts in EHR data quality and eCQMs. Experts included clinicians with an informatics, EHR, or eCQM background; representatives of healthcare facilities; eCQM measure developers; and EHR vendors. As part of this process, NQF drew from its membership of over 360 diverse stakeholders and other organizations with informatics, EHR, and eCQM expertise. NQF also conducted targeted outreach to pertinent organizations and individuals who have served previously on NQF committees, as well as to other desired stakeholders, including individuals on CMS Listservs, individuals with experience in implementing electronic clinical quality measures, subject matter experts on electronic health record data standards and vocabularies, and clinicians with experience in electronic clinical quality measurement.

Terminology and Scope

At the onset of this project, the TEP determined that it was important to have a common understanding of what constitutes a data quality issue and to recognize the both the strengths and challenges of EHR data quality. To establish parameters and to help define the scope of this project, members of the TEP agreed that data quality for this project referred to how well EHR data (structured and unstructured) support clinical quality measurement, including eCQMs, other electronic measurement (such as standardized assessment tools used in post-acute care) and data collected by systems ancillary to EHRs, such as Anesthesia Information Management Systems, Lab Information Systems, and Radiological Information Systems. Data quality for this project did not refer to how well EHRs collect data for the primary purpose of supporting delivery of care. To provide clarity on the terminology around the subject matter of this report, a glossary of key terms is included in Appendix B.

Environmental Scan Findings

As a first step in this project, an environmental scan was conducted to compile current research around EHR data quality and examine the extent to which EHRs are being used as a data source for clinical quality measurements. Both peer-reviewed and grey literature are sources for the environmental scan literature review, and PubMed and Google Scholar were the primary tools used during the literature search. The literature review also included technical documentation around eCQM standards and subject matter expertise from organizations such as Health Level Seven International (HL7), NQF, and HHS, particularly the Centers for Medicare & Medicaid Services (CMS) and the Office of the National Coordinator for Health Information Technology (ONC).

After examining the literature, the TEP identified the following challenges:

- Implementation of EHR-sourced measures in inpatient and outpatient settings is challenging, marked by a limited number of EHR systems that can fully automate the implementation of eCQM technical specifications, and eCQM data element criteria often do not currently align with EHR data.
- Implementation in PAC settings is challenging, despite wide adoption of EHR systems in PAC settings. Performance measures are not integrated into existing PAC EHR systems and there are currently no PAC specific eCQMs.
- The use of unstructured EHR data presents a challenge as it is used extensively for clinical documentation, but it is difficult to ensure consistency across clinicians and consequently to use unstructured data with automated processes NLP or machine learning.

- On the NQF side, the measure evaluation criteria and guidance specific for the evaluation of eCQMs is unclear, resulting in confusion for measure developers and standing committee members.
- NQF's existing data management system was implemented prior to the advent of eCQMs, and the capability for a robust analysis of those measures that are or are not endorsed (such as a count of how many submitted eCQMs did not pass the endorsement process), is not yet available. NQF is in the process of upgrading its data management system in ways that will allow for such capabilities.
- From the developer perspective, NQF endorsement is often challenging because developers are unable to find an acceptable number of more than one EHR system willing or able to participate in testing as described in the NQF evaluation criteria (Appendix C) .
- There is a lack of readily available, accessible, and affordable EHR data to support eCQM measure development and testing efforts.
- Human curation and NLP are often needed to map to the actual data elements which is both difficult to regulate and challenging for interoperability. Even when adherent to the data standards, customization of EHRs for the end-user leads to inconsistencies, resulting in non-standardized data abstracted from the EHR.
- While there are frameworks for assessing EHR data quality, those frameworks are inconsistent and rely on institutional interpretation rather than generalizable approaches to standardized data quality constructs.[6]
- There is a need for accrediting organization or regulatory bodies to support measure developers in testing eCQMs by setting standards for the quality of EHR data used for measurement.

The TEP used the results of the environmental scan to spur discussion and identification of consensus recommendations for promoting data quality as it relates to the development, endorsement, and implementation of eCQMs and EHR-sourced measures.

Recommendations for Improving Use of EHR Data in Measures

The TEP identified recommendations for improving data quality in the use of EHR data. The recommendations are categorized based on three high-level phases or processes of the measure lifecycle that apply to most measures. The measure development phase refers to when measures are specified and tested. The measure endorsement process is when measures are evaluated against NQF's measure evaluation criteria[4] and vetted for use. The measure implementation phase refers to when measures are implemented by healthcare providers and health IT vendors. Details of these recommendations are further discussed in subsequent sections.

Table 1. EHR Data Quality Technical Expert Panel Recommendations

Phase	Recommendations	Impact
Development	• HHS should offer credit to providers and health IT vendors in federal programs for supporting measure development • HHS should create recognition programs around supporting measure development efforts • CMS should consider developing more measures that align across multiple care settings across various programs	Provide incentives to participate in measure testing to increase availability of test sites
Endorsement	• NQF's Scientific Methods Panel should develop guidance specifically for EHR-sourced measures • NQF should determine if changes are needed to measure evaluation criteria • NQF should determine if changes should be made to measure evaluation process • NQF should provide updated criteria, guidance, and education to NQF committees and measure developers • NQF Standing Committee members should play a role in Scientific Methods Panel review	Provide appropriate guidance on how EHR-sourced measures are evaluated to improve likelihood of successful endorsement
Implementation	• CMS should consider grants to fund dedicated full-time equivalents (FTEs) to provide support for vendors in understanding and incorporating measurement into their products in the PAC and other important care setting that were not supported under ARRA/Meaningful Use program funding.	Increase use of EHR-sourced measures in PAC settings and other care settings

Measure Development Challenges

The TEP identified the lack of availability of healthcare providers and EHR vendors to support measure testing efforts as a major challenge impacting EHR-sourced measure development. Measure development and testing not only requires significant time and resources from measure developers to do the development but also from healthcare providers and health IT vendors to source the data for measure development and testing. To help increase the availability of providers and vendors that can provide measure testing support, the TEP recommended offering incentives to expand participation in measure testing by using existing healthcare reporting programs and potentially new recognition programs.

The TEP identified existing barriers such as challenges for healthcare providers and vendors to dedicate time and resources to measure testing. Existing payment programs provide incentives for providers to implement measures. Therefore, these same incentives are not applicable when measures are still in the development phase and are not yet linked to a program that providers and vendors are actively participating in. The TEP suggested that federal programs could offer incentives, not only for measure implementation, but also to dedicate time and resources to support measure testing before measures are finalized and included in federal reporting programs.

Recommendations

HHS should offer credit to providers and health IT vendors in federal programs for supporting measure development.

The range of federal reporting programs offers opportunities to incentivize providers and vendors to participate in measure testing. As one example, the Quality Payment Program (QPP) determines either a positive, neutral or negative payment adjustment for Merit-based Incentive Payment System (MIPS) participants who care for Medicare patients through a multi-factor payment design that includes four performance categories: quality, promoting interoperability, cost and improvement activities. As The TEP recommended targeting the Promoting Interoperability program (a follow-on to the Medicare EHR Incentive Program [Meaningful Use] that focuses on electronic exchange of health information) and Improvement Activities (a catalog of activities to assess clinicians' efforts toward care coordination, patient engagement and access) categories as mechanisms to offer potential incentives for providers to support measure testing efforts.[7] The TEP cited newer Improvement Activity measures such as IA-ERP_3, *COVID-19 Clinical Trials*, which offers MIPS credit for providers participating in a COVID-19 clinical trial or submitting data for a patient in their care who has been infected and is using a drug or biological product to treat the infection. Similarly, the TEP suggested that a new improvement activity or promoting interoperability measure could be targeted for providers who provide data to measure stewards developing new measures for use in federal programs. Finally, the TEP suggested that CMS consider offer bonus points under the MIPS quality category to participants who support eCQM measure testing efforts.

For vendors, the TEP suggested the ONC's Health IT Certification program as a potential mechanism to offer incentives as it already targets vendors seeking health IT functionality certification in their products. The TEP suggested ONC consider granting certification to vendors who provide data to support measures in the development and testing phase, instead of conducting the existing certification tests for published measures. However, ONC clarified that the program only certifies general eCQM functionality and does not certify vendor products for supporting specific eCQMs. And while ONC's certification program verifies a product's ability to collect and sort through quality data, CMS is the responsible party for all quality measure development, testing, and adoption. Another approach is to target CMS' registry self-nomination process, which approves Qualified Clinical Data Registries (QCDRs) every year for MIPS. CMS may be able to offer credit to a QCDR for supporting measure testing of its own eCQMs as part of the approval process.

New measures and improvement activities are nominated for MIPS via the annual call for measures. As a first step to proposing changes to the programs mentioned previously, NQF could first monitor the decisions of the previous annual call for measures cycle to see if new changes related to granting providers for supporting measure testing are already being considered. If no such changes are being

considered, NQF could propose and submit new improvement activities and promoting interoperability measures during 2021 MIPS annual call for measures. The timing of this report should enable NQF to gather additional stakeholder input to inform feedback during the upcoming call for measures cycle that closes in July 2021.

HHS should create recognition programs around supporting measure development efforts.

The TEP recommended that HHS consider recognition programs for providers and vendors to incentivize participation during measure development. The TEP noted examples of existing recognition programs such as the Malcolm Baldrige National Quality Award and the Joint Commission's Pioneers in Quality Award.[8,9] The Baldridge Award helps healthcare organizations distinguish themselves from competitors in their local markets by exceeding the Baldridge Frameworks measures of healthcare quality. The Pioneers in Quality program both highlights hospitals that successfully use health IT and eCQMs to drive quality improvement efforts and facilitates the sharing of best practices with other hospitals and health systems. The TEP suggested modeling a recognition program that highlights providers and vendors who participate in measure testing efforts. For example, modeling after an award such as the Baldridge Award would include recognition around the significant amount of work and criteria for consideration, thus the award is less of a "rubber stamp" but rather a meaningful reward for supporting measure testing. The recognition and prestige could demonstrate that the providers and vendors are committed to quality improvement and could help these organizations be more prominent in the marketplace. The TEP considered this incentive as one that could help justify the significant resources and time needed to support measure testing efforts.

CMS should consider developing measures that align across multiple care settings in various programs.

The TEP noted that measure development contracts are often linked to payment programs for which the measures are developed. This can lead to measures developed with the singular focus on the care setting of the underlying program. To promote measure development across settings, the TEP suggested more coordination at the CMS contracting level. The TEP noted existing efforts for coordination of data elements and other measure components across settings. The Continuity Assessment Record and Evaluation item set (CARE tool) aimed to coordinate elements across different post-acute care settings. The CMS Measures Under Consideration (MUC) List also takes a broad look at measures across Medicare programs. The TEP identified falls and pressure ulcers as a measurement area that can benefit from cross-setting measures, as both occur in multiple care settings. However, the TEP also raised concerns that while the data used across settings may be similar, the nuances of the stage of care in each setting can make reuse of data elements challenging. Examining the contract vehicles used for measure development for potential added coordination could lead to more cross-setting measures that meet the needs of the different payment programs.

Limitations and Impact

The recommendations related to measure development are all impacted by the federal regulatory rulemaking process as that is the primary mechanism to propose changes to federal payment programs. Although the rulemaking process takes time, it allows for more stakeholders to provide input and enables HHS to collect adequate feedback prior to incorporating measures into new and existing programs. Even with these limitations, the recommendations discuss different types of incentives for both healthcare providers and health IT vendors to allocate resources to measure testing. Increasing the availability of test sites would help to address the challenge of measure developers finding enough,

diverse sites to complete measure testing, one of the most significant challenges with developing EHR-sourced measures and meeting criteria for measure endorsement.

Measure Endorsement Challenges

The TEP highlighted the measure endorsement process as an opportunity to improve use of EHR data in measures. NQF endorses measures intended for use in accountability applications as well as quality improvement. Accountability applications are uses of performance results about identifiable, accountable entities to make judgments and decisions as a consequence of 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). NQF applies its measure evaluation criteria for the Consensus Development Process (CDP) to carry out the review of performance measures for endorsement and review of eCQMs for trial approval status. The standard evaluation criteria foster consistency and predictability for measure developers and for those using NQF-endorsed measures.

The TEP reviewed NQF's measure evaluation criteria guidance that apply specifically to eCQMs and noted several areas that are challenging or, at the very least, needing clarification. Given the low volume of data compared to claims based measures, it has been noted that the NQF measure evaluation criteria and submission requirements for eCQMs are often too challenging to meet the expectations of NQF standing committees.[10] The TEP recommended NQF revisit the measure evaluation criteria and associated guidance to consider practical challenges faced by measure stewards and developers, while still addressing the concerns of the criteria. For evaluation criteria, guidance, and measure submission requirements that are simply not clear, the TEP recommended NQF clarify expectations and guidance for measure developers and NQF Standing Committee members. The recommendations related to measure endorsement only apply to eCQMs but are intended to apply to all EHR-sourced measures.

The TEP identified multiple challenges when submitting EHR-sourced measures for endorsement or approval for trial use. The lack of available test sites and data during measure development presents challenges for measure developers to meet the scientific acceptability criteria. Additionally, NQF's feasibility requirements are not always clear and the timing of NQF's feasibility review within the endorsement process can, at times, seem counterproductive. Lastly, the TEP highlighted the challenge that measure developers face when the Standing Committee's discussion and review of one criterion often impacts the review of other criteria. While the criteria are all related in some ways, it can be frustrating to measure developers when the review of one criterion impacts another in unexpected ways. Unclear guidance on the scope and definition of each criterion hinders a standing committee's ability to be more focused in their discussion of evaluation criteria.

Scientific acceptability of measure properties

The TEP raised concerns about ongoing challenges with applying NQF's measure evaluation criteria for scientific acceptability to EHR-sourced measures. These challenges are based on the lack of available data that are needed for testing reliability and validity. NQF has supported two approaches to dealing with the lack of data available for testing, including the use of data sets or data aggregators to meet measure testing requirements and by continuing to offer the Approval for Trial Use (ATU) program as an alternative pathway to endorsement.

NQF requires measure developers using data sets instead of operational EHR systems to document how the data is aggregated, what schemas are used in the normalized data set, what fields have been normalized and how, and how providers are sampled in the data set. The concern is that the testing should not only demonstrate that the measure works using transformed data in aggregated datasets, but also demonstrates that the measure works in live EHR systems. The challenge is that NQF's guidance on using a data set or a data aggregator in place of an operational EHR system is not clear and leads to inconsistent evaluations across different NQF standing committees.

In 2015, NQF attempted to address this challenge by offering the ATU status, which is granted for measures that meet the importance to measure and report, feasibility, and use and usability criteria, but delays the evaluation of scientific acceptability for up to three years after designated as approved for trial use. The goal of ATU was to encourage implementation of a measure on a limited basis, which would help facilitate the availability of more sites to support the testing necessary to demonstrate scientific acceptability within the three-year trial approval time frame. The TEP highlighted that even when granted approval for trial use status, measure stewards still faced the same challenges with a lack of test sites and available data for their measures during the subsequent trial approval period. When the trial period expired and the measures were expected to be reviewed for scientific acceptability, many developers still could not find sites to participate in the testing needed. Unless another steward or developer can take ownership and has the necessary funding to perform the required testing, the result is often orphaned measures.

Feasibility

For eCQMs and EHR-sourced measures in general, feasibility matters more than measures using other data sources. EHR-sourced measures promise to reduce burden by enabling electronic abstraction, and eCQMs go further by targeting the automated calculation of the measures. The TEP noted that the NQF Feasibility Scorecard and the timing of the feasibility review present challenges in ensuring EHR-sourced measures deliver on these promises of reducing burden. The TEP also noted that avoiding additional burden is not always possible in cases when the complexity of the measure or measure type inherently requires changes to workflow, data capture or eCQM standards. Although there is an attempt to minimize burden when developing complex or high-priority measures, there must be acknowledgement of the needed short-term burden increase that is offset by long-term gains by having a given measure fill a needed measure or specialty/sub-specialty gap area. The TEP discussed the need to consider what level of increased shorter-term burden is acceptable for the longer-term benefit when assessing feasibility.

In 2013, NQF's eMeasure Feasibility Assessment Technical Expert Panel recommended use of a standard score card for assessing feasibility of data elements in an EHR-sourced measure. In 2018, the scorecard was updated to reflect stakeholder feedback based on its initial implementation. Even with the updates, stakeholders have continued to express that there is a lack of clarity around it. For example, the scorecard is used is to help identify unstructured fields used in the measure. If an EHR-sourced measure uses unstructured fields, then data element testing must meet reliability requirements. The TEP noted that it is not clear to measure stewards how the feasibility assessment impacts the level of testing needed for reliability criterion, especially since feasibility is reviewed after reliability and validity in the NQF submission forms and in the actual review with the standing committees.

The other challenge with the NQF feasibility criterion is the sequence of criterion review. Feasibility is currently reviewed after the importance to measure and report and scientific acceptability criteria, both of which are "must-pass" criteria. Feasibility is not a must-pass criterion, so if there are glaring feasibility concerns, there is not a mechanism to fail the measure based on that alone. Discussing feasibility late in the evaluation process may not be the best use of resources. If measures are not considered feasible earlier in the process, there is no need to use additional resources trying to prove that they are scientifically acceptable.

Related and overlapping criteria

Related to the feasibility challenge, the evaluation and discussion of one criterion can influence the evaluation, discussion, and voting of subsequent criteria. This can be concerning because the TEP noted that this issue not only happens under the discussion of feasibility impacting reliability, but also occurs between other criteria such as between evidence and validity or between feasibility and validity. While it's generally understood that all of the criteria impact each other, when it comes to the review, discussion, and voting, the TEP noted that how the criteria impact each other when evaluating EHR-sourced measures could be made more clear to developers who are preparing submissions and to committees that are carrying out the reviews.

Recommendations

NQF's Scientific Methods Panel could provide specific guidance on EHR-sourced measures.

The first recommendation to address the NQF evaluation process of EHR-sourced measures is to utilize NQF's Scientific Methods Panel (SMP) to provide updated guidance on how the scientific acceptability criteria should be applied to EHR-sourced measures. Scientific acceptability is a critical and complex component of NQF's measure evaluation criteria and the SMP was created in 2017 specifically to address complex methodological issues. The goals of the SMP were to promote more consistent evaluations of the scientific acceptability criterion, reduce standing committee burden, and promote greater participation of consumers, patients, and purchasers on NQF standing committees. The panel's charge is to evaluate the scientific acceptability of complex measures and to serve in an advisory capacity to NQF regarding methodologic issues. In 2011, the NQF Measure Testing Task Force produced a report to provide guidance to NQF stakeholders on methodological issues related to measure testing. However, there are still nuances in special cases that require additional clarification. For example, the SMP published a white paper to provide targeted guidance for scientific acceptability of risk-adjusted clinical outcome measures. The TEP recommended that the nuances of EHR-sourced measures are significant enough to also warrant its own targeted guidance. Specifically, the SMP could provide input on validity and reliability testing using an operational EHR system versus a data set, thresholds for validity and reliability testing, and the relationship between other NQF criteria (i.e. feasibility and evidence).

NQF should determine if changes are needed to measure evaluation criteria.

The SMP's work would be crucial to determine if changes to NQF EHR-sourced measure evaluation criteria are needed. However, the SMP does not govern the NQF criteria; the Consensus Standards Approval Committee (CSAC) does. The SMP's input could help craft detailed proposals of modified or new criteria for the CSAC to consider. Proposed changes can be presented to CSAC and presented to NQF stakeholders through a comment period for broader feedback on updates to the criteria.

NQF should determine if changes are needed to EHR-sourced measure evaluation process.

While changes, such as requiring feasibility as a must-pass criterion, are directly related to the measure evaluation criteria, changes to the sequence in which feasibility is assessed are directly tied to NQF's measure evaluation process. NQF evaluates measures according to its CDP. Changes to the CDP involve soliciting feedback from its staff, membership, committees, measure users, measure developers and other stakeholders. The TEP recommended considering the evaluation of feasibility earlier in the CDP process, such as the intent to submit period, prior to measure submission. This will facilitate timely feedback for measure stewards on whether, and how best, to use resources on a measure with significant feasibility concerns. Instead, developers could address these issues prior to investing resources in testing the measure for reliability and validity.

NQF should disseminate Scientific Methods Panel guidance to standing committees and measure developer community.

If the SMP recommends that changes be made to the NQF Measure Evaluation Criteria and CSAC approves, NQF can utilize its existing measure developer webinars as educational opportunities to review updates within the measure evaluation criteria, guidance document, and measure developer guidebook. NQF could use the Standing Committee Guidebook and standing committee orientation meetings to share updated guidance to NQF committees and ensure newer guidance is applied across committees. All changes to criteria and guidance can also be shared via public commenting periods to ensure broader stakeholder input. If the SMP guidance leads to NQF criteria changes, then all these mechanisms can be used to share updated NQF guidance.

NQF Standing Committee members could play a role in Scientific Method Panel review.

NQF's standing committees carry out the evaluation, conduct voting, and make recommendations for the endorsement of measures. The SMP reviews measures that are considered complex for scientific acceptability prior to the standing committee evaluation and provides feedback for the standing committees to use during its deliberations. The TEP recommended to have a member of the standing committee participate in the SMP meeting and give perspective as needed, if he or she is more familiar with the submitted measure and the portfolio of measures being reviewed this cycle Alternatively, the TEP suggested that each standing committee include some level of EHR, eCQM, and/or HL7 standards expertise to provide input during the evaluation of eCQMs and EHR-sourced measures.

Impact and Limitations

The recommendations for the measure endorsement phase are subject to review and feedback from NQF stakeholders. Changes to measure evaluation criteria require feedback from external stakeholders, including measure developers, measure implementers, other NQF committees, and the CSAC. Changes to the NQF's Consensus Development Process may require coordination with CMS. Consequently, proposed changes will take time to gather feedback and adjust to process. To update endorsement criteria or processes, they must first be presented for consideration to NQF governance bodies. Finally, improvements to NQF's criteria and processes may only impact NQF's review of measures submitted for endorsement consideration. They may not directly impact measures under development, how they are implemented, or what programs they may be used in. For measures that are submitted to NQF, the recommendations can provide additional clarity for measure stewards developing EHR-sourced measures and increase the odds of endorsing stronger and more feasible ones.

Measure Implementation Challenges

The TEP highlighted the incorporation of EHR-sourced measures into PAC-setting EHR systems as an area to improve implementation of measures using EHR data. All EHR systems can benefit from deeper integration with measures; inpatient and outpatient setting EHR systems already offer some level of EHR-sourced measure integration by providing eCQM-related functionality. Vendors in PAC settings do not currently provide eCQM functionality because eCQMs are not currently required in PAC federal reporting programs. Regardless, the TEP noted that deeper integration of EHR-sourced measures in PAC settings is possible and can reduce the burden of implementing measures.

The environmental scan conducted as part of this work identified different incentives promoting the use of EHR-sourced measures in PAC settings compared to those in ambulatory and acute settings. MIPS, Hospital Acquired Condition Reduction Program (HACRP) and Inpatient Quality Reporting (IQR) programs have used eCQMs as an approach for EHR-sourced measures for the ambulatory and acute settings. At the same time, the ONC Health IT Certification program has incorporated eCQM-related functionality to ensure vendors can support EHR-sourced measurement of eCQMs. While EHR adoption may be lower in PAC settings than in acute and ambulatory ones, the TEP discussed that EHR systems are still widely used across PAC settings, despite the lack of similar incentives that are offered to EHR systems serving the acute and ambulatory ones. One challenge that vendors face in the PAC settings is a lack of expertise in the complex reporting programs. Consequently, the TEP suggested that focusing on incorporating EHR-sourced measurement directly into PAC-setting and other specialty EHR systems can improve EHR data use in measurement.

Recommendations

CMS should explore grants to fund dedicated FTEs to provide support for vendors in understanding and incorporating measurement into their products in the LTPAC and other important care setting that were not supported under ARRA/Meaningful Use program funding.

The TEP described a need for CMS to provide specific funding towards implementation support for EHR-sourced measures. One approach is for CMS to offer grants to fund experts dedicated to providing support to PAC and other specialty providers and vendors for implementing and incorporating EHR-sourced measures into EHR systems. The TEP noted that vendors and providers of all settings could benefit from this approach and that similar funding mechanisms that CMS has used to support measure development. CMS offered cooperative agreement funding to specialty societies and associations under MACRA section 102 to support measure development. While MACRA section 102 funding supported measure development, the TEP recommended CMS consider using a similar approach for funding FTEs to support measure implementation.

While dedicated FTEs to support measure implementation benefits providers and vendors in all settings, this recommendation is particularly relevant to providers and vendors in the PAC settings for two reasons. Firstly, providers and vendors do not always have the staff support to manage the administrative requirements involving payers, alternative payment models, and contracts. Secondly, providers and vendors in PAC settings have concerns with having to adapt to standards, tools, and measures that are primarily hospital focused, without having the opportunity to specify a PAC-tailored approach to EHR-sourced measurement. Several PAC groups, including National Association for the Support of Long Term Care and LTPAC Health IT Collaborative, representing a significant portion of the vendor market, have expressed interest in this approach and are willing to support measure testing

efforts in addition to measure implementation. By providing funded experts to vendors and PAC sites, CMS could address the challenge of availability of test sites and reduce the burden of implementing EHR-sourced measures in PAC settings. The TEP also suggested looking at existing contracts where CMS has funded measure development under MIDS for PAC settings in terms of engaging vendors and PAC sites.

Limitations and Impact

The recommendations to address measure implementation challenges primarily depend on the availability of funding to support the potential grants. Additionally, HHS would need to consider regulatory, policy, and program mechanisms needed to offer such a grant program. If the program is implemented, much needed support could be provided to vendors and providers in PAC settings. Implementation support could relieve some of the burden of using measures and testing measures, which may open the possibility for seeing enhanced development of measures in this space and better utilization of EHR data.

Potential Areas for Further Consideration

The TEP discussed and identified additional areas for consideration that do not fall under the measure development, endorsement, implementation phases, but may still have an impact on improving use of EHR data for measurement. These areas may improve EHR data which may have subsequent benefit for use in quality measurement. While not as explicit as the previous recommendations, these opportunities highlight potential areas worth continued exploration to improve the use of EHR data in measurement.

Articulate the cost and return on investment for supporting measure testing.

The TEP noted that supporting measure testing requires considerable resources from providers and vendors. One of the challenges is that measure testing typically requires efforts from clinical staff and IT who are usually not solely dedicated to measure implementation or measure testing work. The TEP noted that cost can vary widely based on the number of measures and the extent of manual data abstraction needed rather than electronic data abstraction. The TEP suggested that it would be beneficial to collect information on the cost of supporting measure testing efforts to help executives plan and make decisions on when they can support measure testing efforts. Information like the number of hours, type of staff, and type of activities would be useful for planning purposes. CMS measure development contracts, health systems that use measures internally, and registries all develop measures and can provide valuable perspectives to understanding the costs for testing measures. Additionally, the TEP suggested that executive leaders, particularly in the areas of health information technology and quality, would have an innate understanding of the downstream benefits of measure testing efforts as they pertain to future data collection and reporting. The ability to qualitatively and, where possible, quantitatively demonstrate the value of these efforts could help leaders strategically advocate for testing resources within competitive organizational budgets.

Data element catalog.

The TEP noted there is still a need for a data element catalog that can help aid implementation and development of measures. While there are multiple existing data element catalogs like the CMS PAC Data Element Library (DEL) and the eCQI Resource Center Data Element Repository, these catalogs do not contain information related to implementing the data elements. For example, the TEP suggested that adding information from the NQF Feasibility Scorecard would be useful for each data element to

get a sense of feasibility challenges. The TEP noted that some measures are more feasible and other measures require significant workflow changes to encourage new behavior. This information is useful to implementers when deciding on implementing a measure but is also useful for measure developers as they are creating measure specifications.

Utilize existing user groups.

The TEP recommended that existing user groups should provide a forum for measure developers to learn about the state of data being collected from implementers. The TEP note that vendor user groups can help serve this purpose since the membership consists of the vendors and users of the vendor products. Inviting measure developers to these meetings can help increase dialogue on challenging data, such as PROs, while measures are under development. The TEP suggested several groups including the vendor association groups like the Electronic Health Record Association to take this on but also encouraged Healthcare Information and Management Systems Society (HIMSS) , the Quality Data Model users group, and HL7 to consider using their existing groups as well.

Create pilots using existing frameworks, models, and standards to make progress on urgent use cases.

The TEP agreed that pilots are a useful way to move the field forward in challenging areas. It identified the following projects that are in early development but are promising: Fast Healthcare Interoperability Resources (FHIR) as a format for exchanging data not just in clinical applications but in patient-facing applications; Observational Medical Outcomes Partnership (OMOP) as another approach to creating a common data model for exchanging medical data across multiple sources; CMS PACIO project focuses on improving interoperability in PAC settings using FHIR implementation guides; and the Health Level Seven International (HL7) Gravity project for handling social determinants of health. The TEP discussed the benefits these efforts can have by focusing on specific and relevant use cases. However, the relevancy of use cases can change over time, and the feasibility of use cases have been limited by the tools and standards available. As standards evolve, the TEP highlighted that it is important to revisit previous use cases that were not possible. For example, CMS and measure developers have not been able to create risk adjusted or composite eCQMs because the standards have not supported the complexity around specifying the risk adjustment model or the composite scoring of from the component measures. In one instance, the workaround for the eCQM composite scoring limitations was to create an eCQM for each component measure with the composite scoring algorithm documented in the HQMF header and in additional documentation. With advances in Clinical Quality Language (CQL) that improve expressive capability of eCQMs, the feasibility of composite eCQMs may be more possible and worth revisiting.

Revisit applicability of existing frameworks and guidance on assessing how EHR data is used in measures.

The literature review of the environmental scan identified several frameworks that related to EHR data quality. The TEP also discussed other relevant frameworks such as the ONC Interoperability Standards Advisory, which categorizes standards for different types of data and ranks them based on the maturity of the development and adoption of the standard in the industry. The TEP noted that the NQF Feasibility Scorecard is a framework for assessing feasibility of data elements during the development and endorsement. The FHIR specification uses the FHIR Maturity Model to rank the maturity of its resources. While the different frameworks are useful, the TEP noted that harmonizing frameworks or new frameworks may be needed to address specific challenges of using EHR-sourced measures.

Create measures that use manually abstracted data and electronically abstracted data.

The TEP discussed measure developers creating measures that use a combination of manually abstracted and electronically abstracted data. The rationale is that aspirational measures may require data elements that are not currently feasible. Rather than limiting the measure concepts that are pursued to fully feasible measures, this approach aims for a balance, with the intention that using manually abstracted data elements would be the exception and not the rule. The TEP was hesitant on fully recommending that measure developers adopt this approach because using manually abstracted measures increases the burden for implementors. If hybrids of manual and electronically abstracted measures are pursued, it should be done judiciously with these caveats in mind.

Conclusion

Examining the current state of EHR data quality is an important first step to better understanding the potential of improving quality measurement with the alternative use of EHR data. Through an environmental scan and input from a multidisciplinary Technical Expert Panel, NQF highlighted challenges that impact the development, endorsement, and implementation of measures that use EHR data and outlined recommendations for mitigating those challenges. Focusing efforts on improving EHR data in ways that support healthcare performance measures could impact and enhance quality measurement and improve clinical care and patient health outcomes.

The recommendations in this report related to NQF measure evaluation criteria and guidance include an assessment of NQF's eCQM evaluation criteria, guidance and evaluation processes and recommendations for improvements to better inform CDP topic-specific standing committees on eCQMs attributes and implementation considerations. This effort is expected to help address current challenges and provide all stakeholders with guidance and best practices on improving EHR data. Improvements in these areas are expected to increase the scientific acceptability and likelihood of endorsement of high-quality, meaningful EHR-sourced measures, and reduce provider burden for reporting to CMS' quality and reporting programs. For these recommendations to be implemented, NQF will present potential changes to guidance documents cited above during a future CSAC meeting.

An overarching issue of EHR data quality is the challenge of getting multiple stakeholders (e.g., vendors and providers) to participate with measure developers early and throughout the development lifecycle, in a way that balances the cost of participation with the downstream benefit of reducing workflow and implementation costs once the tested measure is in a given program. Although this report focuses on opportunities for HHS, CMS and NQF, it should be noted that additional work in this area does not only lie with these stakeholder groups. It is recommended that future work should focus on opportunities for other stakeholders who can impact EHR data quality issues beyond HHS, CMS and NQF. Until then, NQF will share the recommendations in this report with HHS, CMS and other external stakeholders for consideration and potential implementation.

References

1 Department of Health and Human Services, Centers for Medicare and Medicaid Services. *Medicare and Medicaid Programs; Patient Protection and Affordable Care Act; Interoperability and Patient Access for Medicare Advantage Organization and Medicaid Managed Care Plans, State Medicaid Agencies, CHIP Agencies and CHIP Managed Care Entities, Issuers of Qualified Health Plans in the Federally- Facilitated Exchanges and Health Care Providers*. Vol 45 CFR Part 156.; Proposed. https://www.govinfo.gov/content/pkg/FR-2019-03-04/pdf/2019-02200.pdf. Last accessed December 2019.

2 Eisenberg F, Lasome C, Advani A, et al. *A Study Of The Impact Of Meaningful Use Clinical Quality Measures*. American Hospital Association; 2013. http://assets.fiercemarkets.net/public/newsletter/fiercehealthit/13ehrchallenges-report.pdf. Last accessed November 2019.

3 Tang PC, Coye MJ, Bakken S, et al. *Key Capabilities of an Electronic Health Record System.* Washington, DC: Institute of Medicine of the National Academies; 2003. https://www.ncbi.nlm.nih.gov/books/NBK221802/pdf/Bookshelf_NBK221802.pdf. Last accessed November 2019.

4 National Quality Forum. Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement. September 2019. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=88439. Last accessed July 2020.

5 Bailey LC, Mistry KB, Tinoco A, et al. Addressing electronic clinical information in the construction of quality measures. *Acad Pediatr*. 2014;14(5 Suppl):S82-89.

6 Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. 2013;20(1):144-151.

7 Merit-based Incentive Payment System (MIPS) Overview - QPP. https://qpp.cms.gov/mips/overview. Last accessed September 2020.

8 U.S. Department of Commerce, National Institute of Standards and Technology. Malcolm Baldrige National Quality Award. National Institute of Standards and Technology Baldrige Performance Excellence Program. https://www.nist.gov/baldrige/baldrige-award. Published February 18, 2010. Last accessed July 2020.

9 The Joint Commission. Pioneers in Quality Recognition. Joint Commission Pioneers in Quality Recognition. https://www.jointcommission.org/measurement/pioneers-in-quality/Pioneers in Quality Recognition. Last accessed July 2020.

10 National Quality Forum. *EHR Data Quality Final Environmental Scan*. https://www.qualityforum.org/Publications/2020/05/EHR_Data_Quality_Final_Environmental_Scan.aspx. Last accessed August 2020.

11 National Quality Forum. *NQF Health Information Technology Glossary: A Guide to HIT Jargon*. Washington, DC: NQF https://www.qualityforum.org/Measuring_Performance/HIT_Glossary.aspx. Last accessed December 2019.

12 Clinical Quality Language (CQL). https://cql.hl7.org/. Last accessed April 2020.

13 Skyttberg N, Chen R, Blomqvist H, et al. Exploring Vital Sign Data Quality in Electronic Health Records with Focus on Emergency Care Warning Scores. *Appl Clin Inform*. 2017;8(3):880-892.

14 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.

15 Medicare C for, Baltimore MS 7500 SB, Usa M. Data Element Library Overview. Data Element Library Overview. https://del.cms.gov/DELWeb/pubHome. Published September 20, 2016. Last accessed December 2019.

16 HL7 International. Quality Measure Implementation Guide (v0.1.0) (STU Ballot). HL7 FHIR Quality Measure Implementation Guide. http://hl7.org/fhir/us/cqfmeasures/2019May/index.html. Last accessed December 2019.

17 National Quality Forum. *A Measurement Framework to Assess Nationwide Progress Related to Interoperable Health Information Exchange to Support the National Quality Strategy*. Washington, DC: NQF; 2017. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=85827.

18 Welcome to the Promoting Interoperability Programs Registration System. https://ehrincentives.cms.gov/hitech/loginCredentials.action. Last accessed December 2019.

19 4. Understanding Clinical Data and Workflow. https://www.phii.org/sites/default/files/resource/files/Understanding%20Clinical%20Data%20and%20Workflow%20Guide.pdf. Last accessed August 2020.

Appendix A TEP Roster, CMS/ONC Representatives, NQF Staff

Technical Expert Panel

JohnMarc Alban, MS, RN, CPHIMS
Associate Director, Quality Measurement & Informatics, The Joint Commission
Oakbrook Terrace, Illinois

Zahid Butt, MD, FACG
President and Chief Executive Officer, Medisolv, Inc.
Columbia, Maryland

Cindy Cullen, MS, MBA, PMP
Director, Mathematica
Princeton, New Jersey

John Derr, RPh
JD & Associates Enterprises Inc.
Anacortes, Washington

Karen Dorsey, MD, PhD
Associate Research Scientist, Yale University
New Haven, Connecticut

Zabrina Gonzaga, MSN, RN, CPNP-PC
Lantana Consulting Group.
Wake Forest, North Carolina

Toby Heyn
Business Intelligence Developer, Epic Systems Corporation
Verona, Wisconsin

David Kendrick, MD, MPH
Principal Investigator and Chief Executive Officer, MyHealth Access Network
Tulsa, Oklahoma

Angela Kennedy, DC, MBA
Director, Measure Development, American Society of Clinical Oncology
Alexandria, Virginia

Joseph Kunisch, PhD, RN-BC, CPHQ
Enterprise Director of Clinical Quality Informatics, Memorial Hermann Health System
Houston, Texas

James Langabeer, PhD, MBA
Professor, University of Texas Health Center at Tyler
Houston, Texas

Jamie Lehner, MBA, CAPM
Associate Director, PCPI Foundation

Chicago, Illinois

Michael Lieberman, MD, MS
Medical Director, Oregon Community Health Information Network (OCHIN)
Portland, Oregon

Jacob Lynch, RN-BC
Technology Consultant at Medical Advantage Group
Tampa, Florida

Jana Malinowski
Solution Lead – Regulatory Quality Reporting, Cerner
Kansas City, Missouri

James McClay, MD, MS, FACEP
Associate Professor, Emergency Medicine and Medical Informatics, UNMC
Omaha, Nebraska

Shelly Nash, DO
CMIO – Ambulatory, AdventHealth
Altamonte Springs, Florida

Shea Polancich, PhD, RN
Assistant Dean, System Director Nursing Analytics and Improvement, University of Alabama at Birmingham
Birmingham, Alabama

Stan Rankins
Integration Architect, Telligen
West Des Moines, Iowa

Mike Sacca
Vice President, Healthcare Quality and Data Analytics, Reli Group Inc.
Baltimore, Maryland

Albert Taylor, MD
Medical Informatics Officer, Office of the National Coordinator for Health Information Technology (ONC)
Springfield, Virginia

NQF Staff

Sheri Winsper, RN, MSN, MSHA
Senior Vice President, Quality Measurement

Chuck Amos, MBA
Director

Kathryn Goodwin, MS
Director

Jean-Luc Tilly, MPA
Senior Project Manager

Bejier Edwards, PMP
Project Manager

Erin Buchanan, MPH
Manager

Ngozi Ihenacho, MPH
Analyst

Ameera Chaudhry, MS
Project Analyst

Christopher Millet
NQF Consultant

Appendix B Glossary of Key Terms

TERM	DEFINITION
Bonnie	Bonnie is a tool for testing electronic clinical quality measures (eCQMs). This tool is designed for use by measure developers as part of their development process and validates that the eCQM logic matches the measure's intent. Bonnie uses patient scenarios to represent each logic component of the measure specification such as the initial patient population (IPP), denominator, numerator, exclusions, etc. Health IT developers and implementers may also use the tool to evaluate measure implementation into their systems. Measure developers use both Bonnie and the Measure Authoring Tool (MAT) in concert to promote test driven development.[11]
Clinical Decision Support (CDS)	A process for enhancing health-related decisions and actions with pertinent, organized clinical knowledge and patient information to improve health and healthcare delivery. The information delivered can include general clinical knowledge and guidance, intelligently processed patient data, or a mixture of both. Information delivery formats can include data and order entry facilitators, filtered data displays, reference information, alerts, and others.[11]
Clinical Document Architecture (CDA)	The HL7 Version 3 Clinical Document Architecture (CDA®) is an HL7 standard in XML-based document markup standard that specifies the structure and semantics of "clinical documents" for the purpose of exchange between healthcare providers and patients. It defines a clinical document as having the following six characteristics: persistence, stewardship, potential for authentication, context, wholeness, and human readability. A CDA can contain any type of clinical content—typical CDA documents would be a Discharge Summary, Imaging Report, Admission & Physical, Pathology Report, etc. The most popular use is for inter-enterprise information exchange, such as is envisioned for a US Health Information Exchange (HIE).[11]
Clinical Quality Language (CQL)	A high-level, domain-specific language focused on clinical quality and targeted at measure and decision support artifact authors.[12]
Clinical quality measures (CQM)	Clinical quality measures are tools that help measure and monitor the quality of healthcare and the contribution of healthcare services towards improved health outcomes. In the past, quality measures primarily used data that came from claims, but as technology has improved and become more prominent in the healthcare setting, many quality measures now use data that comes from a provider's electronic health record (EHR). These electronic CQMs (eCQMs) use EHR data to measure health outcomes, clinical processes, patient safety, efficient use of healthcare resources, care coordination, patient engagement, and population and public health improvement. [11]
Completeness	Availability and accessibility of expected entries in the EHR.[13]
Computability	The extent to which an eCQM specification algorithm can be translated to programmable logic constructs and the availability of EHR data elements to implement the eCQM specified QDM data criteria.[14]

TERM	DEFINITION
Current Procedural Terminology (CPT)	This code set is maintained by the American Medical Association. The CPT code set describes medical, surgical, and diagnostic services and is designed to communicate uniform information about medical services and procedures among physicians, coders, patients, accreditation organizations, and payers for administrative, financial, and analytical purposes. CPT coding is similar to ICD-9 and ICD-10 coding, except that it identifies the services rendered rather than the diagnosis on the claim. (ICD code sets also contain procedure codes but these are only used in the inpatient setting.) CPT is currently identified by the Centers for Medicare and Medicaid Services as Level 1 of the Healthcare Common Procedure Coding System. [11]
Cypress	Cypress is an open source certification testing tool for evaluating the accuracy of clinical quality measure calculations in EHR systems and EHR modules. Cypress enables testing of an EHR's ability to accurately calculate eCQMs. Cypress serves as the official eCQM testing tool for the 2014 EHR Certification program by the Office of the National Coordinator for Health IT (ONC). [11]
Data Element Feasibility	The likelihood that data elements are available and a significant number of organizations can capture and access the data element in a consistent manner.[11]
CMS Data Element Library (DEL)	The CMS Data Element Library is the centralized resource for CMS assessment instrument data elements and their associated health information technology (IT) standards.[15]
Data Exchange	The process of sending and receiving data in such a manner that the information content or meaning assigned to the data is not altered during the transmission.[11]
Data Exchange for Quality Measures (DEQM)	The Data Exchange for Quality Measures Implementation Guide provides a framework that defines conformance profiles and guidance to enable the exchange of quality information and quality measure reporting. The DEQM expects to use quality measures specified in accordance with the Quality Measure Implementation Guide and QI-Core.[11]
Denominator	The denominator can be the same as the initial patient population or a subset of the initial patient population, to further constrain the population for the purpose of the eMeasure. Different measures within a set may have the same initial patient population but different denominators. Continuous Variable measures do not have a denominator, but instead define a Measure Population. For proportion or ratio measures, the verbiage "Equals Initial Patient Population" with no additional criteria indicates the denominator is identical to the initial patient population.[11]
Electronic Clinical Quality Measure (eCQM)	Electronic clinical quality measures are eMeasures specified for use in the Medicare and Medicaid EHR Incentive Programs. Eligible professionals, eligible hospitals and critical access hospitals (CAHs) are required to submit CQM data from certified EHR technology to help measure and track the quality of health care services provided within the health care system. These measures use data associated with providers' ability to deliver high-quality care or relate to long term goals for quality health care.[11]

TERM	DEFINITION
Electronic Health Record (EHR)	An electronic health record is a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports. The EHR automates and streamlines the clinician's workflow. The EHR has the ability to generate a complete record of a clinical patient encounter, as well as supporting other care-related activities directly or indirectly via interface, including evidence-based decision support, quality management, and outcomes reporting.[11]
EHR-sourced Measures	Quality measures that rely on data that come from an electronic health record.
Expression Logic Model (ELM)	A machine-readable canonical representation of CQL targeted at implementations and designed to enable sharing of clinical knowledge.[12]
Electronic Medical Record (EMR)	A digital version of a paper chart that contains all of a patient's medical history from one practice. An EMR is mostly used by providers for diagnosis and treatment. The difference between an EMR and an EHR is that an EHR is designed to share information with other health care providers, such as laboratories and specialists. The National Alliance for Health Information Technology stated that EHR data "can be created, managed, and consulted by authorized clinicians and staff across more than one healthcare organization".[11]
Fast Healthcare Interoperability Resources (FHIR)	Fast Healthcare Interoperability Resources is a next-generation standards framework created by HL7 (hl7.org/fhir). FHIR combines the best features of HL7's Version 2, Version 3, and CDA® product lines while leveraging the latest web standards and applying a tight focus on implementability. FHIR solutions are built from a set of modular components called "resources." These resources can easily be assembled into working systems that solve real world clinical and administrative problems at a fraction of the price of existing alternatives. FHIR is suitable for use in a wide variety of contexts, including mobile phone apps, cloud communications, EHR-sourced data sharing, server communication in large institutional healthcare providers, and much more. [11]
FHIR Quality Measure Implementation Guide (QMIG)	The FHIR Quality Measure Implementation Guide defines conformance profiles and guidance focused on the specification of quality measures using the FHIR Measure and Library resources. The QMIG does not standardize the content of any particular measure, rather it defines the standard approach to the representation of that content so that quality measure specifiers can define and share standardized FHIR-based electronic clinical quality measures (eCQMs).[16]
Health Information Exchange (HIE)	A term used to describe both the sharing of health information electronically among two or more entities and also an organization which provides services that enable the sharing electronically of health information.[11]
Health Information Technology for Economic and Clinical Health (HITECH) Act	The HITECH Act provides HHS with the authority to establish programs to improve health care quality, safety, and efficiency through the promotion of Health IT, including EHRs and private and secure electronic health information exchange.[11]

TERM	DEFINITION
Health Insurance Portability and Accountability Act (HIPAA)	HIPAA provides federal protections for personal health information held by covered entities and gives patients an array of rights with respect to that information.[11]
Health IT Policy Committee (HITPC)	A Federal Advisory Committee that coordinates industry and provider input regarding the Medicare and Medicaid Incentive Programs, as well as in consideration of current program data for the Medicare and Medicaid EHR Incentive Programs.[11]
Healthcare Common Procedure Coding System (HCPCS)	A set of health care procedure codes based on the American Medical Association's Current Procedural Terminology (CPT). HCPCS was established in 1978 to provide a standardized coding system for describing the specific items and services provided in the delivery of health care necessary for Medicare, Medicaid, and other health insurance programs to ensure that insurance claims are processed in an orderly and consistent manner. With the implementation of the Health Insurance Portability and Accountability Act of 1996 (HIPAA), use of the HCPCS for transactions involving health care information became mandatory. HCPCS is divided into two principal subsystems, referred to as Level I and Level II. Level I is comprised of the CPT-4 to identify medical services and procedures furnished by physicians and other health care professionals. The Level II HCPCS is a standardized coding system that is used primarily to identify products, supplies, and services not included in the CPT-4 codes. It is maintained and distributed by CMS.[11]
Healthcare Quality Measures Format (HQMF)	A Health Level 7 (HL7) international standard that serves as a wrapper into which a health quality measure using the QDM can be placed. The HQMF serves as a means to share and distribute a clinical quality measure as an electronic document.[11]
Human readable	Each eCQM exported from the Measure Authoring Tool (MAT) includes the measure specifications in an HTML human readable format so that the user can understand both how the elements are defined and the underlying logic used to calculate the measure.[11]
International Classification of Diseases (ICD)	The ICD terminology is maintained by the World Health Organization, the directing and coordinating authority for health within the United Nations System. The ICD is designed as a health care classification system, providing a system of diagnostic codes for classifying diseases, including nuanced classifications of a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or disease. Diagnosis codes are key for determining coverage and are used in treatment decisions. From plan design to statistical tracking of disease, these codes are a crucial part of the way health plans—including State Medicaid agencies—run their programs. Current health plan systems and health care providers are required by HIPAA to use a standard code set to indicate diagnoses and procedures on transactions.[11]
Interoperability	The ability of a system to exchange electronic health information with and use electronic health information from other systems without special effort on the part of the user.[17]

TERM	DEFINITION
Initial Patient Population (IPP)	The initial patient population refers to all patients to be evaluated by a specific performance eMeasure. These patients share a common set of specified characteristics within a specific measurement set to which a given measure belongs. This initial patient population is present regardless of the measure scoring type; i.e., proportion, ratio, and continuous variable measures all have an initial patient population section. Details often include information based upon specific age groups, diagnoses, diagnostic and procedure codes, and enrollment periods.[11]
Logical Observation Identifiers Names and Codes (LOINC)	LOINC is a database and universal standard for identifying medical laboratory observations. It was developed in 1994 and is maintained by the Regenstrief Institute, a US non-profit medical research organization. LOINC was created in response to the demand for an electronic database for clinical care and management and is publicly available at no cost.[11]
Medicare and Medicaid EHR Incentive Programs	The American Recovery and Reinvestment Act of 2009 authorized CMS to provide incentive payments to eligible professionals (EPs) and hospitals who adopt, implement, upgrade, or demonstrate meaningful use of certified EHR technology. Commonly referred to as "Meaningful Use."[11] Currently known as the Promoting Interoperability Programs.
Measure Scoring	Indicates how a calculation is performed for the eMeasure (e.g., proportion, continuous variable, and ratio).[11]
Measure Type	Indicates whether the eMeasure is used to examine a process or an outcome over time (e.g., Structure, Process, and Outcome).[11]
Measurement Period	The time period for which the eMeasure applies.[11]
Measure Population	Measure population is used only in continuous variable eMeasures. It is a narrative description of the eMeasure population. (e.g., all patients seen in the Emergency Department during the measurement period).[11]
Numerator	Numerators are used in proportion and ratio eMeasures. In proportion measures the numerator criteria are the processes or outcomes expected for each patient, procedure, or other unit of measurement defined in the denominator. In ratio measures the numerator is related, but not directly derived from the denominator (e.g., a numerator listing the number of central line blood stream infections and a denominator indicating the days per thousand of central line usage in a specific time period).[11]
Promoting Interoperability Programs	The Promoting Interoperability Programs (previously known as The Medicare and Medicaid EHR Incentive Programs) provides incentive payments to eligible professionals and eligible hospitals as they demonstrate adoption, implementation, upgrading, or meaningful use of certified EHR technology. These interoperability programs are designed to support providers in this period of Health IT transition and instill the use of EHRs in meaningful ways to help our nation to improve the quality, safety, and efficiency of patient health care.[18]

TERM	DEFINITION
Quality Data Model (QDM)	A QDM element is a discrete unit of information used in quality measurement to describe part of the clinical care process, including a clinical entity and its context of use. It can include criteria for any relevant metadata about a clinical or administrative concept relevant to quality measurement. A QDM element provides an unambiguous definition and enables consistent capture and use of data for quality measurement. It may be defined for any given measure and reused when the same information is required for another measure. Reuse encourages standardization of quality measures and reduces the generation of additional software requirements for every new measure.[11]
Quality Improvement Core Implementation Guide (QI-Core)	The Quality Improvement Core Implementation Guide defines a set of FHIR profiles with extensions and bindings needed to create interoperable, quality-focused applications. Importantly, the scope of QI-Core includes both quality measurement and decision support to ensure that knowledge expressed can be shared across both domains. QI-Core is derived from US-Core, meaning that where possible, QI-Core profiles are based on US-Core to ensure alignment with and support for quality improvement data within healthcare systems in the US Realm.[16]
Quality Reporting Document Architecture (QRDA)	The Health Level Seven International (HL7) Quality Reporting Document Architecture is a standard document format for the exchange of eCQM data. QRDA reports: • Contain data extracted from EHRs and other health information technology systems. • Can be used to exchange eCQM data between systems. • Are the data submission standards for a variety of quality measurement and reporting initiatives. • Were adopted by the ONC as the standard to support both QRDA Category I (individual patient) and QRDA Category III (provider's aggregate) data submission approaches for Stage 2 of Meaningful Use.[11]
RxNorm	RxNorm is a non-proprietary drug vocabulary maintained and distributed by the National Library of Medicine. It has been identified as the vocabulary-of-choice to be incorporated into government systems as they are updated. RxNorm provides normalized names for clinical drugs and links its names to many of the drug vocabularies commonly used in pharmacy management and drug interaction software.[11]
State Health Information Exchange	The state HIE program promotes innovative approaches to the secure exchange of health information within and across states. It also works to ensure that health care providers and hospitals meet national standards and Meaningful Use requirements. The Office of the National Coordinator for Health Information Technology (ONC) funds the State Health Information Exchange (HIE) Cooperative Agreement Program.[11]
Structured Data	Structured data follows a prescribed data model and value set, constraining the users to only be able to entering or choose pre-determined values. Computers can readily process structured data, which it often stores in databases. Data sent by medical devices to EHRs will typically send structured data.[19]

TERM	DEFINITION
Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT)	A comprehensive clinical terminology, owned, maintained, and distributed by the International Health Terminology Standards Development Organisation (IHTSDO).[11]
Taxonomy	A standard vocabulary or other classification system that can be used to define a QDM element's category. For the purpose of the QDM, taxonomy is synonymous with a code system (a collection of codes with associated designations and meanings). Specific taxonomies are used in applying the QDM to quality measures based on the recommendations of the HIT Standards Committee of the ONC and established certification rules for Meaningful Use.[11]
Unstructured Data	Unstructured data (also called "free text") does not follow a pre-defined set of values, allowing users to instead enter narrative information about data using their own words. This means of recording data provides the user with the most freedom for recording an entry, but because the same clinical event could be documented in myriad ways, computers can't easily process unstructured data, making errors more likely.[19]
Value Set	Previously referred to as code list, this is a set of values that contain specific codes derived from a particular taxonomy. Value sets are used to define an instance of a category used in a QDM element. A parent value set may also contain child (or nested) value sets that define the same category. The approach is consistent with the HL7 definition for a value set as "a uniquely identifiable set of valid concept representations, where any concept representation can be tested to determine whether or not it is a member of the value set…A sub-value set is a sub-set of a 'parent' value set…When a value set entry references another value set, the child value set is referred to as a nested value set. There is no preset limit to the level of nesting allowed within value sets. Value sets cannot contain themselves, or any of their ancestors (i.e. they cannot be defined recursively)." With respect to value sets, a value is a specific code defined by a given taxonomy. Values are included in value sets. In the context of QDM elements, some categories (e.g., laboratory test) have an attribute of "result." A result may be expressed as a value (numeric or alphanumeric).[11]
Value Set Authority Center (VSAC)	A service provided National Library of Medicine (NLM), in collaboration with the ONC and CMS. The VSAC provides downloadable access to all official versions of vocabulary value sets contained in Clinical Quality Measures (CQMs) used in federal programs. Each value set consists of the numerical values (codes) and human-readable names (terms), drawn from standard vocabularies such as SNOMED CT®, RxNorm, LOINC, and ICD-10-CM, which are used to define clinical concepts used in CQMs (e.g., patients with diabetes, clinical visit). The content of the VSAC will gradually expand to incorporate value sets for other use cases, as well as for new measures and updates to existing measures.[11]
XML (Extensible Markup Language)	This is a computer readable format which enables the automated creation of queries against an EHR or other operational data store for quality reporting. XML provides a basic syntax that can be used to share information among different computers, applications, and organizations without needing to pass through many layers of conversion.[11]

Appendix C NQF Guidance for Evaluating eCQMs

The following guidance addresses NQF measure evaluation criteria for the endorsement of eCQMs as stated in the Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement, September 2019.[4]

Specifications

- Measure specifications should use latest accepted versions of the following industry eCQM technical specifications: Health Quality Measure Format (HQMF), Quality Data Model (QDM), and Clinical Quality Language (CQL). Output from the CMS Measure Authoring Tool (MAT) ensures that the measure uses these technical specifications; however, the MAT is not required to produce HQMF.
- Value sets.
    - All eCQMs submitted to NQF must have published value sets within the VSAC as part of the measure.
    - If an eCQM does not have a published value set, then the measure developer must look to see if there is a published value set that aligns with the proposed value set within its measure.
    - If such a published value set does not exist, then the measure developer must demonstrate that the value set is in draft form and is awaiting publication to VSAC.

Each submitted eCQM undergoes a technical review by NQF staff before going to the Standing Committee for evaluation. For this technical review, NQF staff assess that the measure uses the industry accepted eCQM technical specifications; determine if value sets have been vetted through the VSAC; reviews the feasibility of each data element; and make sure the measure logic has been adequately unit tested using a simulated data set.

Feasibility Assessment

- A feasibility assessment (i.e., scorecard), as originally described in the eMeasure Feasibility Assessment report, is required for all eCQMs. The feasibility assessment includes a scorecard to addresses the data elements and an assessment of the measure logic against a simulated data set. All eCQMs should use the latest NQF Feasibility Scorecard that is available. For assessing measure logic, HTML output from the

CMS Bonnie tool can be used. Alternative unit testing results are acceptable, provided they also demonstrate 100% coverage of the measure logic using simulated data.

Testing for Reliability and Validity

To be considered for NQF endorsement, all eCQMs must be tested empirically using the HQMF specifications. **Beginning Summer 2019**, data element validation will be required for all eCQMs (demonstration of score-level validation is also encouraged). For eCQMs based solely on structured data fields, reliability testing will not be required if data element validation is demonstrated. If data element testing is not possible, justification is required and must be accepted by the Standing Committee.

- The minimum requirement is testing in **EHR systems from more than one EHR vendor**. Developers should test on the number of EHR systems they feel appropriate. It is highly desirable that measures are tested in systems from multiple vendors.
- In the description of the sample used for testing, indicate how the eCQM specifications were used to obtain the data.
- eCQMs specified in older HQMF releases that have previously been endorsed do not need to be retested for maintenance. They may, however, need to be updated to accommodate variations in the most current HQMF release. All newly developed measures should be tested using the most current eCQM technical specifications (HQMF, CQL, and QDM) specifications release format.
- Reliance on data from structured data fields is expected; otherwise, unstructured data must be shown to be both reliable and valid (and this must be demonstrated empirically).
- If a developer is testing an eCQM using any type of normalized EHR clinical data (e.g. from multiple EHR sources), NQF requires, at a minimum, supporting information of what schemas are included in the normalized data set and how they are calculated by the measure logic (i.e., what fields have been normalized and how, including any considerations of how this may affect the measure).
- **As of August 2019**, validity testing at the data element level will be required for all eCQMs. However, as with other measures, testing at the level of the performance measure score also is encouraged if data can be obtained from enough measured entities. If data element testing is not possible, justification is required and must be accepted by the Standing Committee.
    - If the testing is focused on validating the accuracy of the electronic data, analyze agreement between the electronic data obtained using the eCQM specifications and those obtained through abstraction of the entire electronic record (not just the fields used to obtain the electronic data), using statistical analyses such as sensitivity and specificity, positive predictive value, and negative predictive value. The guidance on measure testing allows this type of validity testing to also satisfy the requirement for reliability testing (see Algorithms 2 and 3).
    - Note that testing at the level of data elements requires that all critical data elements be tested (not just agreement of one final overall computation for all patients). At a minimum, the numerator, denominator, and exclusions (or exceptions) must be assessed and reported separately.
    - Use of a simulated data set (e.g. BONNIE) is no longer accepted for testing validity of data elements and is best suited for checking that the measure specifications and logic are working as intended and that value sets are included in the VSAC.
    - NQF's guidance has some flexibility; therefore, measure developers should consult with NQF staff if they think they have another reasonable approach to testing reliability and validity.
- The general guidance on samples for testing any measure also is relevant for eCQMs:
    - Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.
    - The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to

participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.
    - The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.
    - When possible, units of measurement and patients within units should be randomly selected.
- The following subcriteria under Scientific Acceptability of Measure Properties also apply to eCQMs.
    - Exclusion analysis (2b2). If exclusions (or exceptions) are not based on the clinical evidence, analyses should identify the overall frequency of occurrence of the exclusions as well as variability across the measured entities to demonstrate the need to specify exclusions.
    - Risk adjustment (2b3). Outcome and resource use measures require testing of the risk adjustment approach.
    - Differences in performance (2b4). This criterion is about using the measure as specified to distinguish differences in performance across the entities that are being measured. The performance measure scores should be computed for all accountable entities for which eCQM data are available (not just those on which reliability/validity testing was conducted) and then analyzed to identify differences in performance.
    - Because eCQMs are submitted as separate measures, even if the same or similar measures exist, comparability of performance measure scores if specified for multiple data sources (2b5) does not apply.
    - Analysis of missing data (2b6). Approved recommendations from the 2012 projects on eCQM feasibility assessment, composites, and patient-reported outcomes call for an assessment of missing data or nonresponses.