



TO: eMeasure Feasibility Technical Expert Panel

FR: Reva Winkler, Senior Director
Kathryn Streeter, Project Manager

SU: eMeasure Feasibility Assessment draft report: Post-Comment Call to Discuss Member and Public Comments

DA: February 22, 2013

The eMeasure Feasibility TEP will meet via conference call on Wednesday, February 27 to:

- review and discuss comments received during the Member and Public Comment period;
- provide input on responses to comments; and
- suggest changes to the [draft report](#) in response to the comments.

Conference Call: *Wednesday, February 27 2013, 11-1pm ET*

Please use the following information to access the conference call line and webinar:

Dial-in Number: (888) 799-5160

Confirmation Code: 11330922

Event Title: eMeasure Feasibility Expert Panel – Conference Call

Webinar: Register at <http://nqf.commpartners.com/se/Rd/Mt.aspx?738581>

All Committee and speaker phone lines will be open. Please place your phone on mute when not speaking. Please do not place your phone on hold during the call.

Attachments

- Comment table (excel spreadsheet)

NQF received 91 comments on the draft report from 10 public and NQF member organizations. In this memo, staff has grouped the comments into each of the report's recommendations for discussion by the TEP.

General comments

- It is not clear whether measure developers are testing eMeasures prior to determining the feasibility of capturing the data elements from the EHR. Although this problem may exist, how widespread is it that guidance is needed?
- The report is limited to data element testing and particularly testing within the boundary of the current existing data element pool. It is not eMeasure (measure level) feasibility assessment but rather eMeasure data element feasibility assessment based on currently existing data elements.
- In the overarching principles, there is a tendency to lump whether it may be possible for an EHR to capture a data element with whether it is clinically appropriate to capture the element, as well as with the difficulty that may be associated with data capture. These three elements need to be considered separately.
- Clarify the meaning of “not feasible.”
- It would be clearer to state that this report is not to address any issue related to creating new data elements but to focus on the feasibility testing using the existing data elements. There were a few places that mentioned the need to look for alternatives for capturing new data elements for new clinical concepts but there is no follow up or elaboration on this suggestion in the report.
- Details related to the high-level, general recommendations are lacking. Is the same process used at various points and what is the value of the outcome to the development process? How does this result inform the stakeholders to respond differently? Are all recommendations critical for success and must they be entirely implemented? Are certain recommendation weighted differently than others?
- The unique perspective of a small practice was not included in the environmental scan, and therefore, your results are somewhat skewed. eMeasures are currently unattainable for many providers, a large number of whom are in small or solo practices. Many neither have electronic health records (EHRs) nor do they report on quality measures. Their input in this assessment would be just as valuable as those of large health systems and would likely help you to develop a strategy to bridge the gap to those not currently participating.
- Why would feasibility requirements be different between re-tooled and de novo eMeasures? What is the TEP's rationale for separating these processes? Shouldn't one hold the developers and stewards to the same criteria regardless of whether the measure has already been created or is being created?
- There are really two aspects of feasibility that should be addressed - can a measurement be captured and passed on to CMS for data aggregation and the ability to drill down and allow for focus review of data to improved patient outcomes. The vendor variability among the certified EHR vendors and the true costs associated with a complete quality improvement oriented EHR requires analysis greater than the routine physician is accustomed.
- Data collection burden traditionally refers to burden on health care providers themselves. In this report, it appears that burden on EHR developers, or even measure developers themselves, is a factor in determining feasibility. The first “overarching principle” states the importance of the measure needs to be weighed against “the cost and the time required for development.” There is a similar statement in Recommendation 2.2. Is the burden on EHR vendors and/or measure developers considered in the feasibility assessment?
- The report lacks input from patients and public consumer advocates which is essential to having a thorough, comprehensive report.

Recommendations 1: Assess feasibility throughout eMeasure development

- Collaborations require active and ongoing communication with stakeholders as well as additional time and resources during the measure development process. Obtaining feedback multiple times may increase the amount of time needed at the beginning of the process to send requests for feedback, communicate with stakeholders, analyze feedback, and integrate suggestions into measure specifications. Although we agree that earlier identification of infeasible measures will be more efficient in the long run, these additional steps may impact the timeline for measurement development.
- This approach is focused on measures as the root of the issue while the fundamental underlying problem is the lack of semantic interoperability to enhance clinical care, care coordination, clinical decision support and measurement. It is important that the recommendations for eMeasure development also impact data usage within and among EHRs to enhance clinical care.
- Add the inclusion of NLM vocabulary experts in the early stages. Request for new codes from NLM and its supported standards have been made at the tail-end of the process. These requests are the canary in the coal mine for signaling problems with feasibility (if there is not an existing code, the data would never be in a current computer record.)
- Recommend that developers map to the QDM and codes early on to allow for time to request new codes, if necessary.

Recommendation 2: Framework for eMeasure feasibility assessment

- To promote an open dialogue between all stakeholders and allow for greater flexibility, this report should outline the standards and types of questions for feasibility assessment, rather than outlining a prescriptive approach to obtaining answers to those questions. Such guidance would better support the identification and discussion of the obstacles to feasibility than a potentially “game-able” scoring system.

Recommendation 2.1 Data element feasibility assessment (score card)

- It is unclear how the score card will be used to drive meaningful use of EHRs. A potential negative consequence of this tool may be the development of eMeasures that are “possible”, rather than eMeasures that “could be possible”.
- The use of a prescriptive and standardized process does not allow measure developers and their partners to tailor the feasibility assessment approach to their needs, nor does it encourage innovation and advancement within the field. Listing, separate from the scorecard, the important components of data element feasibility that must be considered and modifying the scorecard description and instructions to indicate that the resource is a tool and its use optional is recommended. A rating system such as the one proposed is subjective and could be “gamed” through the purposeful selection of more technically advanced or aligned EHR systems in the assessment process.
- It is unclear which entity would complete the scorecard (EHR vendor, measure developer), how that entity would complete the scorecard, and how the process for completing the scorecard would be operationalized. Since the scorecard proposed within the report would only apply to one data element, we are concerned that the process for completing the assessment would be highly labor intensive.

- Ideally, assessment should be performed by clinicians using EHR systems today rather than vendors. It is important to differentiate from fields vendors make available and actual data capture.
- It would not be appropriate to expect all measure developers to obtain a representative sample of EHR implementation, due both to the great variability in EHR systems and the burden placed on the vendor community; in addition, feasibility assessment efforts under federal contracts often are subject to abbreviated measure development schedules and constrained by Paperwork Reduction Act (PRA) regulations on data collection.
- The scorecard assumes that all data elements must be collected through some sort of data entry - including check boxes. Very often the EHR system is aware of clinician activities and can provide indications of these activities automatically.
- Provide greater clarification on the definitions of current and future feasibility ratings. For instance, is “current” feasibility limited to the capacities of EHR systems at present, or does it include changes that could be made to the EHR within a relatively short time (e.g., a few months) by purchasing existing EHR modules or through new certified EHR technology (CEHRT) requirements? Furthermore, we recommend that NQF provide alternative definitions of the scores for assessing the future feasibility of the measure. For instance, the definition of a score of 1 on data availability could read: “The ability to collect this data element is not expected to be required for certified EHRs in the next 3–5 years, nor is it likely to be widely collected within that time frame.”
- Clarify how the market share for the assessed EHR systems should be determined to inform documentation of the feasibility assessment process. Although we agree that this information is valuable, we have found it can be difficult to ascertain, particularly as EHR vendors are unable to provide such statistics.

Data availability:

- The assessment should not limit the data availability to only structured data, because free-text data can be useful as long as the EHR vendors have capabilities to transform free-text into query able data.
- The work would be easier and more certain if all developers were asked to pick their data elements from a fixed universe, and if the CMS/ONC definitions for what had to be carried in the EHR were concordant with the part of the universe that was used in the eMeasures.
- Strongly suggest asking eMeasure developers to note in what part (and or what field) in the EHR the data would be found, not just whether it would be present, e.g., problem list, discharge diagnosis, or lab results, radiology, reports, orders.
- For a score of 3, the guidance states “Data element is routinely collected as part of care process and exists in the majority of EHRs.” For measures not achieving a score of 3, it will not be determinable if the low score is due to structural issues with the EHR or workflow issues such that clinicians do not document in the structured data fields even though the capability exists. Since the latter concern (clinicians not documenting) is addressed by the Workflow question, I suggest that the Data Availability ratings be revised to clearly represent only the issue of the EHR’s structural capability to capture the required data elements.
- The current availability of structured data is largely dependent on the quality and prevalence of vendor-developed templates.

Data accuracy:

- Data accuracy conflates the concept of feasibility with the criterion of validity. As currently defined, it is unclear whether this item refers to the accuracy of the specifications (already captured in NQF scientific acceptability evaluation sub criteria 2b1) or the accuracy of the data element relative to what care actually was provided to the patient. Furthermore, we believe the descriptions of ratings shown under data accuracy make broad assumptions on which sources should be considered most accurate. Although we agree that data accuracy should be considered early in the measure development process, defining it as part of feasibility likely will lead to confusion.
- Accuracy is a function of the measure intent. In determining feasibility, it will be important to determine if the source required by the measure developer is available in the EHR.
- Authoritative source is judgmental. Given the context of a data element with all of its associated attributes (metadata), the question should ask about (a) Does the information available in your system reliably meet the definition of the data element purpose and definition, (b) Does the information available in your system have the potential for meeting reliability based on the definition and purpose of the data element, (c) even with workflow changes, the information in my system cannot be expected to reliably provide data as required by the data element.
- Accuracy will be most influenced by currency: something that might have been true in the past is not necessarily true at present.

Data standards:

- Mapping eMeasures to the nationally-accepted vocabulary standards is good. But if one has to create a new vocabulary code (as mentioned above) to capture something that is not ordinarily captured (as mentioned above) that should be a signal that this is a burdensome measure or to look for other EHR data that would be a sufficient surrogate.
- The QDM are like the islands in the film Avatar that float in the sky, disconnected from anything. The eMeasures are not grounded in data that would be found in an early medical record. So specifying “where” the data element “is” in the QDM does not help except for those QDM elements that are also specified in the ONC MU2 rules.
- Clarify whether the term “data standards” refers to the measure specifications themselves or the regular use of the specified terminologies within EHRs.
- Define the term “nationally accepted,” –there is great variation among providers in the types of coding systems adopted. For instance, although sites may be using National Drug Codes (NDCs) for billing, some use 11-digit codes while others use 10 digits.

Workflow:

- Workflow per se is not a characteristic of a data element, but rather may affect the quality of the data collected for a specific data element. In addition it will vary by EHR product, and by local implementations of those products, and even varying use patterns by individual clinicians. Workflow should be removed as a separate characteristic for testing, but instead should be understood as a mechanism that influences data accuracy and completeness, as well as location in the EHR and data format. The scorecard forces the reporter to assume that all EHRs and clinicians operate similarly; this certainly will not be the case.
- The set of characteristics should parallel the measure requirements and include reliability as well as accuracy. Reliability may be best reflected by workflow.
- Caution should be applied when assessing workflow functionality. Concerns raised about the commitment to safe and appropriate utilization of EHR systems when companies market

and design EHR “time saving” shortcuts such as cut and paste technology. Such strategies present a concern for meaningful data capture and possibilities of technology abuse.

Suggestions for monitoring tools for inappropriate repetitive “template” language and cut and paste patterns.

- Assessing the relation to workflow is good, but even if data can be captured during workflow, recording it electronically as structured data may entail extra work and consume time that should be dedicated to the patient.

Recommendation 2.2 eMeasure feasibility assessment

- Although the measure specifications and calculation logic are important to understand the intent of measures, the calculation algorithm for the measure will not exist in EHR systems. Therefore it is unclear what is meant by the recommendation to assess measure logic by testing whether the algorithm works in multiple EHR systems.
- Provide suggestions on how measure developers and their collaborators can assess each of the questions outlined for overall measure feasibility assessment.
- Remove or revise the first question, “Does the calculation algorithm work in multiple EHR systems?” Some groups view the ability of EHRs to combine data element information to produce comparable measure scores across EHR implementations as a component of the measure’s reliability—a subcriterion of scientific acceptability. Including this step of the overall measure feasibility assessment may cause confusion about what types of formal measure testing are required to meet the scientific acceptability criteria.
- No measure should be considered feasible unless all data elements (and their related context) are feasible. eMeasure feasibility regarding the ability of the logic to score the measure correctly is highly significant, but a different issue than data element feasibility.
- We further recommend that the overall measure feasibility guidelines be modified to incorporate examples of how measure developers can answer the questions outlined on page 12 and clarify the distinction between feasibility and reliability.

Recommendations 3: Validating the data element feasibility scoring

- Clarify methods to determine the summary scores. It is unclear how the data element and eMeasure feasibility assessments will be validated since these results may be subjective and vary from reviewer to reviewer.
- Agree with using measures selected for Meaningful Use Stage 2. Caution when assessing the feasibility of the measures against outdated CEHRT requirements, given that EHR vendors and providers may find it difficult to recall the limitations of their EHR systems under Stage 1 CEHRT requirements if they have already begun transitioning to Stage 2 requirements (recall bias).

Recommendations 4: Data element feasibility catalogue/repository

- How would this lead to standardization of how data elements are represented in the QDM, since in order to be in the repository; presumably they would already be expressed in the QDM?
- It is unclear what an annual review of the assessments should entail and who would be tasked with conducting them. Clarify the re-evaluation and maintenance process.
- The repository should interact with the NLM Value Set Authority Center (VSAC) to avoid redefining value sets. It should also be the repository from which the measure developer

can include elements in a measure using the Measure Authoring Tool – only elements in the repository should be allowable for building a measure.

Recommendations 5: NQF evaluation for endorsement

- The report should include how feasibility will be assessed as a part of the NQF endorsement criteria for eMeasures. In addition, it needs to be clearly defined whether feasibility testing would be assessed as a part of the NQF criteria for Scientific Acceptability or Feasibility, since reliability and validity are used in conjunction with feasibility.
- How would these scores be factored in and weighed in the endorsement process? If an electronic measure is not feasible in the short-term, but it extremely clinically important for a particular patient population, would that drive the long-term feasibility? Also, long-term feasibility is much more subjective, so it is unclear how that would be factored in. It would be helpful to give measure developers some guidance on minimum requirements for feasibility.
- Provide instructions on which statistics should be reported for consideration by steering committees. For instance, if this assessment is completed multiple times to reflect feasibility in multiple EHR systems, should measure developers report (1) average total scores, (2) most frequent responses in each category (mode), (3) the percentage of EHR implementations that met a particular threshold, or (4) the distribution of scores across EHR implementations?
- Remove the recommendation that steering committees consider assessments of fewer than three EHR systems and 10 installations as insufficient. Requiring a minimum number of systems is much more prescriptive than NQF's current guidance on scientific acceptability testing. In addition to the difficulties of identifying and contracting with possible testing sites, measures developed under contracts with the federal government would be unable to meet this threshold due to the constraints of the PRA and difficulties in obtaining Office of Management and Budget (OMB) clearance, which effectively restrict measure testing to no more than nine sites. Furthermore, often there is a trade-off between the number of implementations assessed and the comprehensiveness of the information collected, suggesting that some measure development projects may be better served by conducting a more detailed assessment of a smaller number of implementations.
- It would be helpful to have some detail or guidance on what the plans might entail for all data element components scoring a "1".
- Remove the requirement that measure developers include a plan for improving data elements that receive a score of 1. Improvement of scores would often involve changes to EHR systems or clinical workflow, and not all data elements are sufficiently important to merit such changes. As long as the data elements could feasibly be captured by EHRs, a decision regarding the relative importance of data elements and measures is beyond a measure developer's control; instead, it is the responsibility of those implementing quality measurement initiatives.

Recommendations 6: Composite measures

- Measure developers and their collaborators should consider three levels of feasibility when assessing composite measures—data elements, component measures, and the overall composite score.

- Given the inherent complexity of developing an electronically-specified composite measure compared to an individual eMeasure, it would seem appropriate to have a separate feasibility assessment tool for these measures. Otherwise, the majority of composite eMeasures are likely to automatically receive low feasibility scores.
- The composite requires an additional HQMF document that combines the criteria of the component measures.

Recommendation 7: Greater collaborative efforts are needed for eMeasure development and implementation

- This recommendation could be perceived as critical of providers and health systems for having only limited participation in measure development to date. We recommend modifying this recommendation to focus on how to make providers feel more like partners in the process and incentivize their involvement.
- It is very important to have commitment from several large EMR vendors to work with the developers and bear the responsibility of indicating whether their systems can or cannot support a standardized field. We have found that the EHR vendors have very little bandwidth for working on any modifications to an existing program or extract citing all resources are being consumed meeting Meaningful Use and PQRS requirements.
- The certified vendors list needs further transparency. Experience in meaningful use and best practice indicators demonstrates a wide variability of certified vendors for the availability of query tools at the practice/provider level for QI as well as data capture.
- The report should acknowledge that there are many challenges to be recognized to establish strong working relationships across stakeholder groups. Often it is challenging to identify providers and EHR vendors with the time and knowledge to contribute to the measure development process and CMS and other funders are transitioning to processes designed to reduce the development time.
- The most valuable feedback is obtained when working one-on-one with implementers in the field. New forums for information sharing among developers, implementers, providers, patients and the public should be considered to improve collaboration and information sharing.
- Is there a role NQF can play to facilitation collaboration between EMR vendors and developers that would lead to successful eMeasure implementation?

Recommendations 8: Need for testing partners

- Over time testing partners will distance themselves from other hospitals due to the knowledge and skills gained from working one-on-one with developers. Efforts to avoid biased pool of test sites are suggested.
- Testing should be managed through crowd sourcing and democratized rather than allowing selected test partners.
- Need to ensure that testing partners include a pool of hospitals and/or physicians who may not be as well-informed and experienced in electronic reporting.