Igniting Harmonized Digital Clinical Quality Measurement through Terminology, CQL, and FHIR

Robert C. McClure[1]   Caroline L. Macumber[2]   Julia L. Skapik[3]   Anne Marie Smith[4]

[1] MD Partners, Inc., Lafayette, Colorado, United States
[2] Apelon, Inc., Hartford, Connecticut, United States
[3] Division of Clinical Affairs, National Association of Community Health Centers, Inc., Bethesda, Maryland, United States
[4] Department of Measure Validation, National Committee for Quality Assurance, Washington, District of Columbia, United States

Address for correspondence Robert C. McClure, MD, MD Partners, Inc., 511 South Miller Avenue, Lafayette, CO 80026, United States (e-mail: rmcclure@mdpartners.com).

Abstract

Background Electronic clinical quality measures (eCQMs) seek to quantify the adherence of health care to evidence-based standards. This requires a high level of consistency to reduce the effort of data collection and ensure comparisons are valid. Yet, there is considerable variability in local data capture, in the use of data standards and in implemented documentation processes, so organizations struggle to implement quality measures and extract data reliably for comparison across patients, providers, and systems.

Objective In this paper, we discuss opportunities for harmonization within and across eCQMs; specifically, at the level of the measure concept, the logical clauses or phrases, the data elements, and the codes and value sets.

Methods The authors, experts in measure development, quality assurance, standards and implementation, reviewed measure structure and content to describe the state of the art for measure analysis and harmonization. Our review resulted in the identification of four measure component levels for harmonization. We provide examples for harmonization of each of the four measure components based on experience with current quality measurement programs including the Centers for Medicare and Medicaid Services eCQM programs.

Results In general, there are significant issues with lack of harmonization across measure concepts, logical phrases, and data elements. This magnifies implementation problems, confuses users, and requires more elaborate data mapping and maintenance.

Conclusion Comparisons using semantically equivalent data are needed to accurately measure performance and reduce workflow interruptions with the aim of reducing evidence-based care gaps. It comes as no surprise that electronic health record designed for purposes other than quality improvement and used within a fragmented care delivery system would benefit greatly from common data representation, measure harmony, and consistency. We suggest that by enabling measure authors and implementers to deliver consistent electronic quality measure content in four key areas; the industry can improve quality measurement.

Keywords
► knowledge modeling and representation
► clinical practice guideline
► quality indicators, health care
► vocabulary, controlled
► HL7 CQL

Background and Significance

The United States spends significantly more on health care than any other nation yet lags behind similar Organization for Economic Cooperation and Development countries in multiple health outcome and process measures including all-cause mortality, premature death, death amenable to health care, and disease burden.[1] This disparity is in part due to gaps in implementing evidence-based health care. In an attempt to close some of these gaps, the United States is moving toward value-based reimbursement,[2] with a focus on utilizing clinical metrics to evaluate the quality[3,4] of care provided to individuals and groups of patients.

Clinical quality measures traditionally evaluate performance using manually abstracted clinical and administrative data, electronic clinical quality measures (eCQMs) evaluate performance using data extracted electronically from electronic health records and/or digital health information technology (HIT) systems.[5]

Electronic clinical quality measurement is intended to use data collected in the course of providing clinical care, preferably without significant additional burden to care providers. Quality measures use content populated by users; ideally, where the definitions are stable and appropriate for the purpose. When an eCQM introduces documentation requirements beyond those needed for general clinical care, such as the requirement for structured documentation on breast-feeding introduced with eCQM CMS9, the implementation effort is magnified across every implementation site. Harmonizing the measure to avoid unnecessary impacts pays obvious dividends.

To achieve less cost than manual data extraction, more comprehensive measure coverage through a scalable process, and broader opportunities for measuring outcomes, it is necessary to align measures with automated data extraction from real health care documentation and programs and then utilize it for all downstream purposes. The inability to easily share, implement, and extract data for electronic clinical quality measurement remains a major barrier to the successful execution of standardized, cost-effective, and accurate quality measurement.[6,7]

The eCQMs intend to use point-of-care electronic data to trigger actions and evaluate adherence to evidence-based guidelines at the individual patient level. Given that we want to improve quality, eCQMs should align with clinical decision support (CDS) applications and address these same care gaps in real time. However, traditional CDS development and quality measure assessment have been developed by different organizations and have had different target audiences. The CMS Quality Data Model (QDM)[8] standard for quality measurement, which was developed by the National Quality Forum (NQF), was developed separately from the HL7 standards used for CDS. While the approach, workflow integration and endpoints are often different between quality measurement and CDS, alignment is crucial. The two approaches intend to address the same evidence-based action and there is high resource investment for both. Whereas CDS targets patient and provider actions that support real-time adherence to evidence-based care, eCQMs record the proportion of adherence to the same evidence-based care.

The health care, standards, and quality improvement communities recognized the value of aligning these needs. In 2015, the US Office of the National Coordinator for Health IT (ONC) and Centers for Medicare and Medicaid Services (CMS) sponsored a standards project through Health Level Seven International (HL7) to modularize and harmonize the separate standards, subsequently releasing the Clinical Quality Language (CQL) standard.[9] HL7 designed CQL to be independent of, but align with Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR)[10] Quality Improvement Core (QI-Core)[11] profiles and provide a robust and shared expression of logic for CDS and quality measurement. Using CQL, quality measure actions are expressed in a format that is both human readable and enables computer-based processing. In 2018, the CMS required use of CQL for their electronic quality measure programs as a step toward full modernization and harmonized of standards. Furthermore, CQL is the primary logical expression language supporting FHIR and is increasingly applied to real-time clinical processes and data exchange solutions throughout the HIT industry.[12] Commercial implementations include the evolution of National Committee of Quality Assurance's Healthcare Effectiveness Data and Information Set (HEDIS), which covers a population of 190 million, from a noncomputable measure format to publication of digital measure packages utilizing FHIR and CQL[13] in 2020.

While harmonizing base standards will allow the quality community to improve alignment with decision support and ease implementation of eCQM artifacts into deployed HIT systems (E.G.: common CQL libraries that can be aligned with EHR database structures), harmonizing specific content of quality measures is also important for the usability and effectiveness of quality measurement. By analyzing variability across and within measures, one can employ harmonization to reduce unnecessary duplication, optionality, and the implementation and maintenance effort without compromising the content. Optimally, quality measures can be harmonized with other data extraction and analysis use cases such as surveillance or registry reporting to ease data collection burden.

Addressing the lack of consistency and achieving harmonization across measures (and CDS) to reduce the effort of extracting the appropriate data and implementing eCQMs is critical. More robust tools, processes, standards, and collaboration in the eCQM space provide an excellent opportunity to align existing content, better integrate[14] eCQMs and CDS artifacts, and create further mechanisms for harmonization, relieving the burden on clinicians and implementers alike.

Misalignment in Quality Measurement

The quality measure development process involves a diverse group of stakeholders, including measure developers, payers, regulatory bodies, clinical specialty groups, and other interested parties. Often, there is a lack of harmonization across the numerous measure development efforts due to insular processes with little to no insight into similar components utilized

by other measure developers. In fact, a 2010 consensus report published by NQF stated that there was a lack of measure harmonization throughout the measure development lifecycle and across NQF consensus development projects.[15]

To limit the costs of measure implementation and the effects of measure data entry on patient care, harmonization must occur early in the process and at all levels of the quality measurement ecosystem. We describe four areas of quality measure misalignment, which will benefit from harmonization, listed in increasingly granular order and described in the sections that follow:

- Measure concept: The high-level clinical definition, the intent of the measure.
- Measure clause: A statement that specifies a collection of context-specific discrete patient data needed to identify a specific subpopulation of patients needed for the measure.
- Data elements: A general clinical class of data paired with a contextually specific discrete piece of data found within electronic clinical systems.
- Terminology: The specific set of standardized codes expected to be recorded as part of patient-level EHR documentation that is referenced by the specific quality measure data element (i.e., bound to a data element; ►Fig. 1).

Misalignment can happen across programs of unrelated measures, within a program or within a measure domain, or even within a measure itself. Lack of congruence throughout these layers of content can create implementation barriers that are sometimes impossible to overcome at the point of care. If two programs define different age range cut-offs for an expected procedure, then applying both measures will result in conflicting recommendations for patients inside the problematic overlapping age range.

Measure Concept Misalignment

For this paper, the phrase measure concept is used to represent the high-level clinical intent of the measure along with the summary of its relevant populations, source evidence, and measure metadata including measure type. When viewed across multiple measures, measure concepts can be misaligned when they seek to address similar questions or issues but approach them in ways that conflict (►Fig. 2).

For example, the CMS Meaningful Use program and the Health Resources and Services Agency (HRSA) Ryan White HIV program[16] both used similar, but misaligned, patient age and required encounter time frames to evaluate whether providers were providing appropriate HIV follow-up. In this example, one measure looked for visits to occur every 3 months and the other used a 6-month follow-up period. Providers report that these seemingly slight differences make it hard to know which interval to use, or when to advise patients to return. Furthermore, HIT developers report that these distinctions make it impossible to build CDS for providers to help them improve their care because it is unclear at what intervals to trigger a visit or for which patients the recommendations apply. As a result, the developers or informatics teams often guess at the right interval or simply implement the most clinically aggressive approach possible which may result in unneeded care or less than optimal adherence to the clinical expert guidelines from which the measures are derived.

Currently, there is a lack of tooling in which both measures in development and those in production are indexed and compared. For the eCQM program, CMS publishes a list of Measures Under Consideration (MUC).[17] However, the basic descriptions contained in the MUC do not provide sufficient information such as the guideline being modeled to compare measure concepts and avoid misalignment. Measure concept
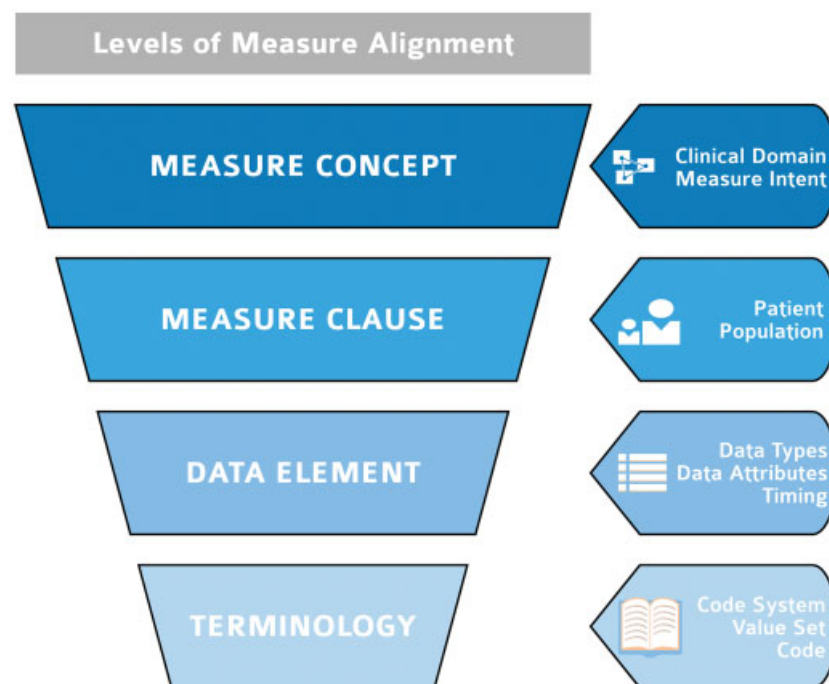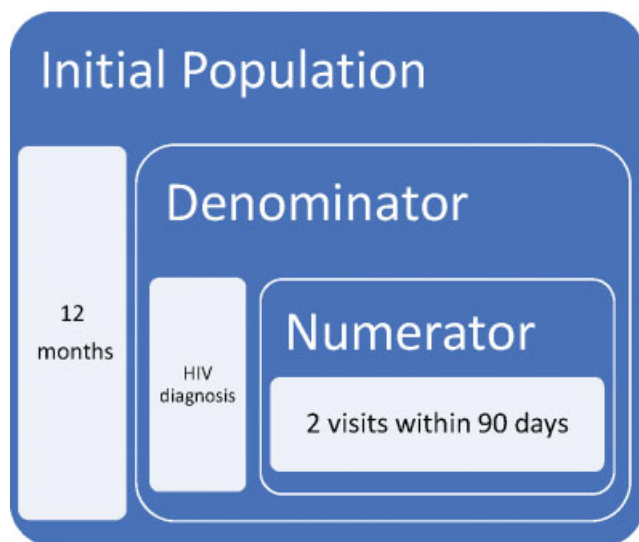

Fig. 1 Levels of measure alignment—capturing possible areas of harmonization, where the granularity in each level increases with each lower level in the inverted pyramid scheme.

CMS MU Measure

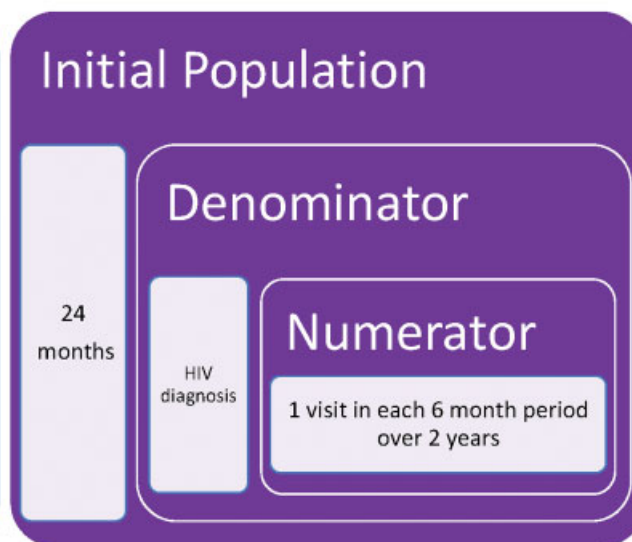## HRSA Ryan White Measure

Fig. 2 Measure concept misalignment for follow-up HIV screening medical visits in CMS and HRSA Ryan White measure tools. CMS, Centers for Medicare and Medicaid Services; HRSA, Health Resources and Services Agency.

harmonization would ensure measure concepts did not provide conflicting advice about patient care to providers, reduce duplicate measure development cost and effort, encourage HIT developers to build better content to support users around the measure, and require less documentation effort by providers.

Measure Clause Misalignment
A measure clause is the segment of a measure that specifies a collection of discrete patient data to identify patients addressed by the measure. That segment is then either used as an entire population of patients within the measure or is combined with other subpopulations within the measure to create the "initial population," the "exclusions," the "numerator population," or the "denominator population." A clause is usually the combination of a data element, bound value set, and any other measure specifications such as timing, necessary to characterize a needed patient population based on conditions or events. Given that quality measurement began before EHRs, the data model used to define a measure clause was not well aligned with the variety of data models used in EHR systems. These models, built by humans, used words and meanings that met the use case of concern but were not driven by, or constrained by, the need for absolute data consistency because humans adapt to the different words used across different measures as they understand the nuanced alignments of the terms. With eCQMs and electronic systems, if an exact match from the measure to the EHR does not occur, an alignment (mapping) must be locally specified exactly, which often results in variation. Computable alignment requires exact matching and with all these human-constructed and nuanced-driven inputs, misalignment is common.

Even when a measure clause aligns with the data models used in EHRs, misalignment can occur when measure developers work in silos. Measure reviews are often done by multiple people within and across organizations resulting in changes to clauses that can result in inconsistencies when compared with similar clauses in other measures. Even cross-program human quality assurance checks are not able to always catch mismatches across thousands of lines of code and existing clause libraries do not include all logic that could be reused by measure developers.

It is common for measures identifying the same patient population to use slightly different phrases. In the example below, the intent was for the same timing, but different clauses were used to identify that timing (►**Fig. 3**).

If a standard clause library existed, measure authors could reference that library and express the timing the same way in both measures. Clause misalignment occurs in many measures because of the lack of rigid naming conventions around clauses. Misalignments in clause timing lead to uncertainties in the timing of reminders for CDS, leaving physicians to overuse or underuse services.

In ►**Fig. 4**, when trying to identify patients that should be excluded, this measure looks for bilateral mastectomies as a procedure or a history of bilateral mastectomy as a diagnosis to ensure that all patients are removed from the measure. Because the clause name is not specific enough, it appears bilateral mastectomies are classified as either procedures or diagnoses. Provider workflow may need to change to allow capture of the data to meet the requirements of the quality measure.

Data Element Misalignment
A data element identifies the specific type of data required by a quality measure. Specification of data elements allow individuals who monitor clinical performance and outcomes to understand concisely and consistently the information needed for a quality measure. Data elements are part of measure clauses and bound to either value sets or

Fig. 3 HIV measure clause example.

Fig. 4 Breast cancer screening example.

single codes. Within an eCQM, how a data element is expressed is determined by the data model used. QDM (originally defined by NQF[18,19] and now managed by CMS[8]) is the model currently used to express data elements in CMS program quality measures. The QDM defines a common library of QDM data elements[20] by a combination of categories, datatypes,[21] and attributes to describe clinical concepts contained within quality measures in a standardized format. Using the same data model to support data exchange between EHRs and to define quality measures and CDS reduces the burden of mapping between data models and the effort to implement new content. Our use of the phrase "data element" is intended to align with, but not require exact implementation of, a QDM data element, in that a quality measure data element needs to sufficiently specify the intended type of data that is to be found in a clinical record.

Even with the work done to date, quality measure stakeholders have not addressed identification and harmonization of misaligned data elements in any systematic fashion. The eCQI Resource Center[22] contains a Data Element Repository that provides the definitions and clinical focus for each data element associated with a published or tested eCQM. This tool could provide a means for measure developers to begin harmonizing data elements if the community could consistently implement a best practice to identify misalignments.

Terminology Misalignment

Value sets[23] are used in measures to identify a specific set of codes expected to be found as patient instances of the specified type of data (e.g., ordered laboratories and current medications) within the EHR. As such, within an eCQM measure clause, a value set will be bound to a measure data element. This value set binding clarifies what specific

type of patient data are expected and which allowed values are to be used for that data element. These data element plus value set combinations are the primary tool used to document what clinical scenario characteristics are to be used to identify populations of concern. In our diagram depicting the levels of measure alignment (►**Fig. 1**), this is the combination of the final two levels: data element plus terminology. Because measure authors can be focused on different QDM datatypes, or nuanced differences in typical patient care, they may define slightly different value sets for use in common situations. Harmonization of value sets occurs when these different value sets are assessed for meaningful differences with a desire to align and reuse. Difficulties arise in the improvement of measures even when opportunities for value set harmonization are identified because the code system content upon which the value sets are defined can change with every new release of the code system. Code system changes can result in semantic drift (i.e., the evolution of a terms meaning), causing a need for frequent reassessment of the harmonized result. This is further complicated by the need for mapping of local data to the measure standard, a practice that adds another point of review whenever changes in value sets are made.

It is important to keep in mind that inclusion of value sets in quality measures will directly lead to requirements for specified encoded data capture within an EHR. This is best managed by aligning the value set content with data typically captured in the course of care, but frequently this is not the current situation. That means specification of the value set may result in introduction of an additional documentation burden although the benefit is improved consistency of the patient data within the EHR record.

In contrast to the three other layers of possible misalignment, progress has been made by the quality measurement community to identify and resolve terminology misalignment. Common value sets ensure semantic interoperability, mitigating inaccurate quality measure calculations, and

comparisons due to data variance by requiring that data sources report using a required list of codes from standard terminologies. For example, two mental health measures examining depression-related diagnoses both include the data element diagnosis: major depression in their measure clause, as shown below ►**Fig. 5**.

Though the clause in which the data element used is not capturing the same clinical context, it does appear that the diagnosis: major depression data element is the same. However, when the measures were reviewed as part of a discussion between the two measure developers, it was observed that the value sets bound to the data elements were not the same. The two value sets identified in ►**Fig. 5** differed in two important ways:

- Different standard terminologies: Measure Developer 1 was using ICD-10-CM and SNOMED CT while Measure Developer 2 was using ICD-9-CM.
- Different value set inclusion/exclusion criteria: Measure Developer 1 was inclusive of "in remission" related diagnosis codes while Measure Developer 2 was excluding "in remission" related diagnosis codes.

In this successful harmonization example, the measure developers harmonized by first creating two subdomain value sets to represent the "current" and "in remission" states. In these value sets, codes from all applicable terminologies are included. Next, they created a more general value set, sometimes referred to as a grouper value set, composed of both the "current" and "in remission" value sets. This approach allows for measure developers to use different combinations of more granular value sets or the entire grouper. Further, the use of a modular approach mitigates the need to create numerous value sets that may be duplicative of one another, while fostering an ecosystem in which value sets are reusable across many quality measures.[24]

Identifying terminology misalignment requires access to the terminology content in a scalable and consumable

Measures about Major Depression

Measure 1:

Depression Diagnoses

    ["Diagnosis": "Major Depression"] union ["Diagnosis": "Dysthymia"]

Measure 2:

Has Initial Major Depression Diagnosis

    exists (["Diagnosis": "Major Depression"] MajorDepression
        let AntidepressantDispensed: "Antidepressant Dispensed 245 Days Before or 120
        Days After Start of Measurement Period"
            where start of MajorDepression.prevalencePeriod within 60 days of AntidepressantDis-
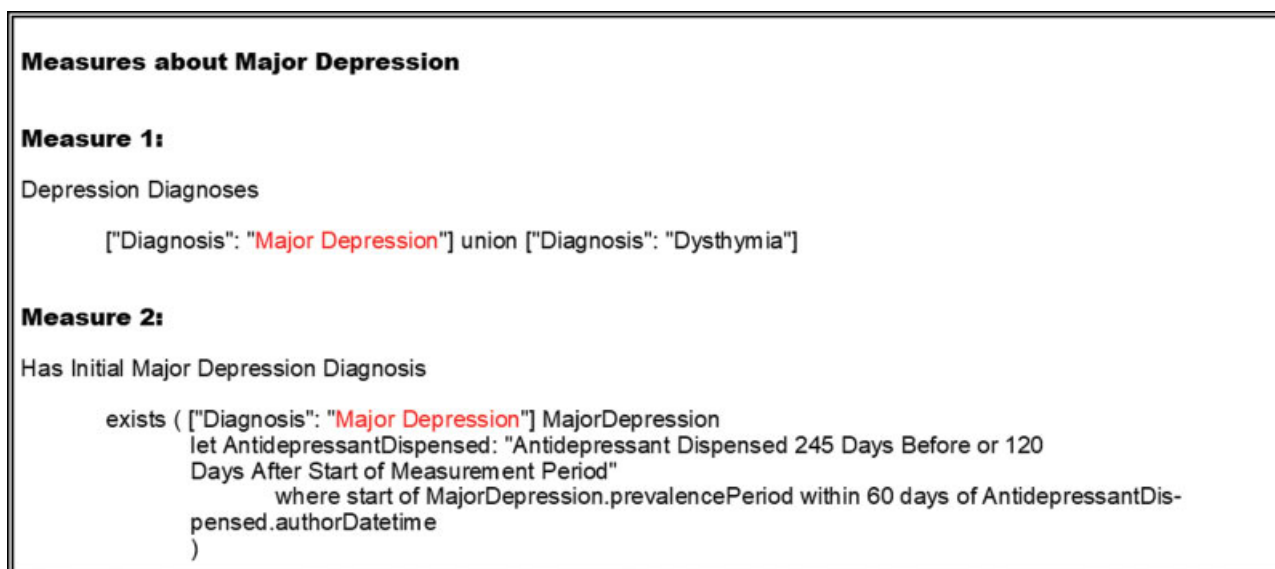   pensed.authorDatetime
)

Fig. 5 Major depression data element example.

format. Though analysis of value sets is possible by manually extracting content from formats traditionally meant for printing, such as PDF or Microsoft Word documents, the process is error prone and not scalable. Further, value set analysis in isolation (i.e., without access to the underlying code system) would greatly impede thorough research into appropriateness of codes and identification of missing codes.

One exemplary resource that can help deliver harmonized quality measurement via value set alignment efforts is the US National Library of Medicine's Value Set Authority Center (VSAC.)[25,26] VSAC is a publicly accessible repository for value sets, containing all official versions of value sets specified by CMS eCQMs, HL7 Consolidated Clinical Document Architecture (C-CDA) Value Sets, and CMS Hybrid Value Sets. VSAC provides searching and browsing capabilities for both code systems and value sets based on metadata characteristics (when provided.) During the value set development process, authors can check the degree of similarity and overlap based on included codes in the value set. VSAC currently enables measure developers to view the measures in which a value set has been implemented, but the specific data element to which it is bound is only noted in the eCQM download files for CMS programs.

Impact on Implementation and Clinical Practice

Capturing clinical data are time consuming and frequently occupies time that would be otherwise used in providing patient care. However, encoding data in a consistent manner is required for comparability and analysis. Without detracting from the important goal of improving clinical outcomes, we must focus on the work of harmonization in areas where either the impact of change to the quality data capture and analysis process is minimal or the benefit to stakeholders (patients, clinicians, implementers, and data analysts) is high. It is important to catalog and track all identified misalignments where they exist because regular updates to content allow iterative opportunities to improve. Each of the various levels of eCQM misalignment requires different approaches and results in different changes to the ecosystem. We address each in turn below.

Measure Concept Harmonization

Identification of measure concept misalignments begins when the measure concept is determined. One would start by comparing the intent of the clinical guidelines/evidence to existing measure concepts and seek to add new measure concepts only when needed. Despite the years-long recognition of this problem, there are still "look-alike" measures across federal programs, not to say anything of state and other regulatory requirements. The harmonization process begins by gathering measures from required sources into a single repository or list. For CMS, one can utilize the CMS Measures Inventory Tool.[27] For other programs, one often has to download this data manually from an organization's website, made difficult by measures that are proprietary or not easily accessed. Organizations developing measures have

often published them shortly before, during or after the period of use, making a prospective analysis for harmonization opportunities nearly impossible. Often measure and program stakeholders do not have the impetus to align their version of a measure with others.

Analysis of quality measures proposed for harmonization starts by categorizing each measure based on domain, intent, populations, and measure type. An analysis should identify measures which have related concept domains[28] at the header level for more in-depth analysis. It will be necessary to confirm by looking at the measure logic if the measure actually aligns with the narrative because naming is not always accurate or descriptive enough to indicate the measure content, even at a high level. Where there are shared or nearly identical conditions, populations, settings, and timing, this could generate a flag that identifies a potential opportunity for a harmonization. It is optimal to align across all use cases; for example, if there is a measure for deep vein thrombosis (DVT) prevention in the hospital setting, one would optimally seek to align it where possible with a DVT prevention measure for skilled nursing facilities, even though the sites are different. This kind of alignment will also help implementers and developers in the measure development process by reducing the number of new concepts and clauses needed to code.

When a measure evaluator finds that two related measures are not aligned, it is important to reach out to measure and program owners. The likelihood of correcting misalignments is greatly improved when the owners can clearly understand the nature of the problem and the misalignment, its impact on you or others, and the potential solution that meets stakeholder needs. See "Measure Concept Misalignment" section for an example of resolved misalignment.

It is important to be sensitive to implementer burden derived by quality assessment of multiple distinct, but different aspects of the same general health care delivery issue. By this, we mean care must be taken when different quality measures attempt to address the same general health concern. Requiring collection of different data, assessing different workflows and addressing different aspects of the same condition results in recommendations that are difficult for institutions to interpret and implement. This type of difficulty is addressed mostly through a harmonization of measure focus and prioritization of concerns and less through maintaining a multimeasure requirement. On occasion, the best solution may be to reduce the number of applied eCQMs that address the same measure concept.

As the diagram in ►Fig. 1 suggests, it is optimal to harmonize at the highest level in the scheme (measure concept); thus, if there are two measures that are similar but redundant, the best approach to resolve them is to create a single shared measure rather than aligning the value sets across them. This will maximize the benefit to the community and allow resources on the developer side to be preserved for testing and other measure efforts. Where measure concepts are not amenable to harmonization because their clinical focus is different or their populations vary for evidence-based reasons, analysis of harmonization should focus on the more detailed layers below.

Measure Clause Harmonization

While noted in our diagram as a unique item for harmonization, our current system for managing artifacts such as clauses has not been defined. Given that a clause is a combination of data element, bound value set, and any other measure specifications such as timing, etc. that are necessary to characterize a patient's population; this type of artifact is tightly aligned with EHR implementation capabilities. Therefore, reuse and alignment across multiple measures would be of great benefit for implementers and would be a time saver for measure developers; for example, in creating common definitions such as outpatient visits for patients with diabetes. The push to harmonize will be tempered by the fact that quality measures frequently must specify unique measure-specific clinical scenario characteristics that are not readily reused. Yet, if the measure clause is built using some common requirements, harmonization where possible, does provide great implementer value.

Data Element Harmonization

Data elements are clinical information types that are context focused and characterize the individual pieces of information related to patient or clinical scenario. For example, a data element can be medications ordered, which is similar, but not the same as medications dispensed; which is obviously different from a clinical condition that patient actively has, yet again different from one that is suspected or is now resolved. Because data elements in eCQMs are quite specific, they are intended to directly align with EHR data structures so they can be consistently applied across EHR vendors and implementations. It is through alignment of meaning and use of these data elements that we achieve executable clinical quality measures (as well as aligned decision support) that have shared implementable meaning. To address the need for well-defined data elements, NQF created the QDM. The characterization provided by the QDM typically matches EHR structures to some degree and aligns to specific care workflows, but because there are nuances in both care delivery process and implementation approach, more than one QDM data element can be considered when defining the data needed.

Choosing one data element over another is driven by measure author approach and can be surprisingly variable when viewed across measures. This variance is likely due to measure author perception of importance, nuanced interest in population variables, and familiarity with EHR capabilities. Even with the QDM as a library for data element choices, variability continues to exist among quality measures for implementation because some measures do not use the QDM and other newer models continue to evolve that capture the interest and resources of the measure community. With the exploding interest in HL7 FHIR,[29,30] there is ongoing work to align QDM with FHIR representations of clinical artifacts occurring through, as mentioned previously, the movement to adopt the CQL[31,32] within the US CMS program use of eCQMs.

Data element alignment has broad interest and appeal. By placing a stake in the ground with respect to meaning and boundaries for each data element, meaningful discussion can occur regarding which data elements to use when characterizing very specific aspects of clinical scenarios and through that we also determine when to improve, change, and align when disparities exist. Through data elements, the community can signal when broad agreement, harmonization, is expected and needed, such as the ONC US Core Data for Interoperability (USCDI).[33]

Within the work of creating a common set of data element choices, the final mile of data element harmonization can occur by seizing opportunities to use exactly the same data elements, or combinations, to identify the same type of patient population even across measure types.

Harmonizing Terminology

Representing quality measures using harmonized data elements is a great leap forward, but to fully harmonize clinical meaning and simplify implementation requirements, the expected patient clinical data to be found and exchanged for each data element must also be aligned. This means harmonizing the coded data (concepts) and/or the value set, bound to the data element to define the specific context needed.

While harmonization of the combination of data element and value set used to define population characteristics is important, value set harmonization independent of data elements can and should be done because these value sets may be used (bound) to more than one data type. This is particularly true when considering that some quality measures may desire the use of different nuanced data types, such as ordered medications only versus dispensed, yet binding to a common harmonized value set appropriate for both types of data are clearly valuable. When this is not possible, mapping between different codes in different value sets can be implemented as what should be seen as a partial solution to lack of terminology harmonization.

Harmonization beyond Measure Artifacts

Improving quality measure use and implementation through consistency does not need to stop with harmonization of measure artifacts. Implementing a quality measure requires determining where in the clinical workflow the measure should appear and what sort of user engagement (such as pop-up alert or ordering interaction among other choices) should be applied. How users will interact with the measure requirements, in what setting or encounter it should be seen, and how the data will be reused or entered in the course of care must also be considered. If measure developers fail to adequately study and test the workflow implications of their measures, they should expect unintended consequences. For example, in the past, measures have failed when executed against all patient data because the measure developer coded the clinical elements in the workflow in the wrong order, not understanding in what order the activities were documented.

Harmonization is also occurring within the realm of standards. While eCQMs have been around since before 2012, CMS has just begun aligning the standards for eCQMs with those for point of care CDS.[34] CMS has currently deployed eCQMs using CQL with the QDM and it is expected that the movement to FHIR-based quality measures will occur in the near future.

Though skeptics remain, FHIR uptake continues to build momentum throughout the health care industry.[35] Today FHIR Release 4 (R4) is a normative standard, further bolstering confidence for implementers to release FHIR-based systems into production. For example, in April of 2019, only 3 months after FHIR R4 was released, Cerner touted its adoption of FHIR R4 and integration into its application programming interfaces (API) to third-party app developers,[36] clearly seeing the move as a market differentiator. Furthermore, ONC appears poised to clarify that their API requirement in the future will include the use of FHIR for all certified products. Soon many EHRs will be able to represent their clinical content via FHIR APIs, a change that will enable CQL use with FHIR resource content that is already at least partially mapped from EHR systems. We anticipate this will push agencies like CMS who manage quality measurement programs toward the deployment of FHIR-based quality measures that are already in development, bringing CDS and quality measurement into closer alignment and providing the opportunity to offer CDS artifacts in association with the eCQMs with little additional effort. While it will almost certainly require significant work to execute and deploy for developers and implementers, it should result in a golden opportunity to analyze misalignment and reduce or eliminate it wherever possible.

How do We Improve?

- Use common approaches: Those who require the use of and/or develop quality measures must also use well-developed standards that they participate in maintaining. Yes, standards are evolving, and caution should always be employed when it is unclear when new standards reach readiness, but it is relatively clear that the FHIR infrastructure is gaining traction and there will be advantages in working to deliver measures using FHIR and CQL. The community can enable this by supporting the development of open-source tools as this lowers the burden of adoption and enables community-based improvements. We must create open libraries of sharable artifacts that developers can reuse and implementers can question, trust, access, and provide feedback on.[37] Finally, we need to work to share experience on implementation tools for these shared artifacts so that the hard work of mapping EHR structures to FHIR elements can be shared.
- Create a governance framework around eCQMs: While CMS does have a governance mechanism within their measure development programs applied by the eCQM Governance Workgroup, it has not resolved all conflicts in the measure ecosystem to date. There is no larger eCQM ecosystem level mechanism to enforce or create collaborative oversight and quality assurance across measures in general. Measure governance is difficult without a larger body or coordinating mechanism; however, CMS and others—particularly clinical communities directly impacted—should collaboratively create and share guidelines and introduce testing to enforce harmonization as a requirement of published content. Provision of an overarching governance process is an important mechanism for aligning measures among measure developers, and in addressing differences in measure intent and required quality measure versions used for reporting among programs. A formal process to address harmonization issues raised by users should be considered.
- Gain insight into user experience: Actual implementation experience provides meaningful insight into real-world impacts on workflow; however, traditional measures have at best had limited early exposure to this kind of feedback. Explicit investigation into both generic and actual examples of workflows and data collection models in which a measure concept will need to function is critical to the creation of high quality eCQM artifacts. We need to create or develop relationships with user communities where this knowledge is gathered, shared, and fed back into the development community. Like we do for postmarketing surveillance for therapeutics, we need to approach quality assessment impact on clinician burden with the same rigorous review. The creation of a library where workflows can be shared and tested would add significantly to the measure community and attempts to perform alignment.
- Seek to use data automatically collected in routine care: We need to focus on the use of data captured by our systems collected during routine care delivery process without undue burden. That means using data created by the system, devices connected to the system and by humans working with the system. We need to improve our understanding of what is available for use, more openness with regards to EHR data capture, and rethink how to identify patient populations using these new data types instead of relying upon complex data entered by clinicians. All measure developers should at a very early stage in development work with one or optimally more, clinical data warehouses, or exchanges to understand what data can be automatically extracted from clinical systems and how it should be specified to minimize the burden of data reentry and mapping. A community testbed of deidentified patient data would be the optimal method of allowing the entire developer community to easily understand what data are widely available across products and systems and how it can be utilized to minimize the imposition on the delivery of care.
- Encourage investigation of Natural Language Processing technology: Improvements in sharing of common data elements and automated data collection will never completely remove the need for identification of complex information within clinical data. Research has shown[38,39] that Natural Language Processing can provide reasonable automated extraction to both find and encode information only in free-text, but also help support the documentation process.

Conclusion

Quality measurement and analysis in gaps in health care delivery should be a priority for health care organizations,

payers, and providers independent of the current interest in tying payment to "value." Traditional approaches have resulted in unintended consequences, such as increased data entry and extraction costs and labor, interruptions in clinical provider workflows that disrupt care pathways and detract from time with patients, and the failure to address gaps in care through timely feedback to providers and patients. It is critical to utilize techniques to decrease these burdens, improve the fidelity of the quality analytics, integrate the technical workload into existing systems, and align population-based retrospective quality assessment with CDS tools to enhance better care that follow "the five rights of CDS."[40] The past decade has brought significant upheavals to clinical care processes during the rapid adoption of EHRs and we are now pushing hard to deliver on the promises made for that pain, promises of enhanced data analysis, and improved insight in the ever-increasing clinical complexity of our patients. We have experienced an explosion of care activities, data entry, and analytics requirements as we explored this transition, it is now time to align and harmonize our technical and structural approach in our quality measurement artifacts and tooling.

Clinical Relevance Statement

The eCQMs impact clinical practice both positively, through identifying opportunities for care improvement, and negatively, by demanding unique data capture by clinicians. This is creating disruptive data capture requirements and forcing measure-specific clinical workflows. This article discusses drivers for the negative impacts and suggests approaches that can be followed to reduce the inconsistencies introduced by measure developers and approaches to reduce difficulties in implementing eCQMs.

Multiple Choice Questions

1. What should be the first step when a suspected misalignment of the care expected across different quality measures is discovered by a stakeholder?
   a. Send an email to the measure owner.
   b. Investigate the current measure and all of its available guidance to verify the misalignment.
   c. Report the finding in the CMS/ONC eCQM Jira Issue Tracker.
   d. Create a workaround by using a "dummy" data element.

Correct Answer: The correct answer is option b. Misalignments are complex and there are a lot of measure stakeholders and feedback mechanisms. While it is tempting to report an issue as soon as it is found, we recommend taking some steps first before reporting. Begin by checking the measure version to make sure it is correct and that it corresponds to the current program requirements. Some measures have been made optional or even eliminated due to errors or issues before they were actually used in a program year. Also, look at published sources of measure guidance. Start with the measure specification

itself including any liner notes within the specification. Also review the measure guidance published with the measure, FAQs or other subsequent guidance, general guidance from the program or set of measures, guidance from the vendor where the measure has been already implemented, and at the locations of any public feedback, in particular for CMS eCQMs a user should review the ONC/CMS eCQM Jira Issue Tracker. Often known issues will be addressed by the measure owner when possible and workarounds may be offered at the JIRS site. Even other stakeholders may suggest approaches that can save time and hassle, so it is recommended to always do your research before creating new workarounds. Of course, any issue that has not been published, or even important ones that have, should be reported or commented on to help the owner understand the need for alignment and the impact on patient care and HIT systems.

2. What is a primary benefit of CQL in the construction and use of eCQMs?
   a. The Centers for Medicare and Medicaid save money on eCQM development.
   b. CQL was created to support a FHIR-based data model.
   c. CQL allows you to code the expressions for both eCQM and CDS authoring and execution.
   d. CQL allows you to express the same measure or statement in as many different ways as possible.

Correct Answer: The correct answer is option c. CQL was designed to allow the integration of eCQMs and CDS because both are critical in understanding our adoption of and improving our adherence to evidence-based guidelines and research. Existing HL7 standards for eCQMs: QDM-based Health Quality Measures Format, and for CDS: Knowledge Artifact Specification, Decision Support as a Service, and the Virtual Medical Record (VMR) were siloed and incompatible with each other. CQL was designed to be agnostic to any specific data model, so whether the data model is QDM, VMR, or FHIR, CQL can be used. Furthermore, CQL is designed to be translated to other existing expression and coding languages using an ELM translator so it is easier to implement into existing systems. While the migration to using CQL is an additional expense for CMS, measure developers, and most HIT vendors in the long run, it is expected to reduce implementation costs and issues, with the added benefits of reduce measure complexity and facilitating paired measure-CDS implementation and in turn measure performance and quality improvement.

References

1 Available at: https://www.healthsystemtracker.org/chart-collection/quality-u-s-healthcare-system-compare-countries/#item-start. Accessed July 15, 2019

2 Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Value-Based-Programs.html. Accessed October 26, 2019

3 Ryan AM, Rodgers PE. Linking quality and spending to measure value for people with serious illness. J Palliat Med 2018;21(S2):S74–S80

4 Saver BG, Martin SA, Adler RN, et al. Care that matters: quality measurement and health care. PLoS Med 2015;12(11):e1001902

5 Available at: https://ecqi.healthit.gov/content/ecqms. Accessed July 10, 2019

6 D'Amore JD, Li C, McCrary L, et al. Using clinical data standards to measure quality: a new approach. Appl Clin Inform 2018;9(02):422–431

7 Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. Quantifying the effect of data quality on the validity of an eMeasure. Appl Clin Inform 2017;8(04):1012–1021

8 Available at: https://ecqi.healthit.gov/qdm-quality-data-model. Accessed July 21, 2019

9 The current balloted version of CQL is R1, STU4, Available at: https://cql.hl7.org/2019May/index.html

10 Available at: http://hl7.org/implement/standards/fhir/. Accessed July 20, 2019

11 Available at: http://build.fhir.org/ig/cqframework/qi-core/. Accessed July 22, 2019

12 Commercial references to use of CQL include Apervita: Available at: https://apervita.com/cql/, Ablehealth: https://ablehealth.com/2017/11/17/a-new-era-for-measure-development-clinical-quality-language-is-here/

13 Available at: http://store.ncqa.org/index.php/performance-measurement/hedis-volume-2-digital-measure-bundles-comparison.html?___SID=U

14 See the following: CDC Adapting Clinical Guidelines for the Digital Age. Available at: https://www.cdc.gov/ddphss/clinical-guidelines/index.html. Accessed October 31, 2019. FHIR Clinical Guidelines implementation guide current build. Available at: http://build.fhir.org/ig/HL7/cqf-recommendations/. Accessed October 31, 2019

15 National Quality Forum (NQF). Guidance for measure harmonization: a consensus report, Washington, DC: NQF. Available at: https://www.qualityforum.org/Publications/2011/05/Guidance_for_Measure_Harmonization.aspx. 2010

16 The eCQM measure is no longer active. The most recent description available can be found here. Available at: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/eCQM_2016EPEC_MeasuresTable.pdf. For current information on the HRSA HIV Medical Visit Frequency measure. Available at: https://hab.hrsa.gov/sites/default/files/hab/clinical-quality-management/coremeasures.pdf

17 Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/CMS-Measures-Inventory.html. Accessed October 26, 2019

18 NQF QDM. Available at: http://www.qualityforum.org/QualityDataModel.aspx. Accessed July 10, 2019

19 Thompson WK, Rasmussen LV, Pacheco JA, et al. An evaluation of the NQF quality data model for representing electronic health record driven phenotyping algorithms. AMIA Annu Symp Proc 2012;2012:911–920

20 "A QDM data element encapsulates a certain category with an associated datatype.". Available at: https://ecqi.healthit.gov/glossary/qdm-data-element. Accessed October 31, 2019

21 QDM uses the phrase "datatype" differently than the traditional programing "data type." QDM datatypes are a combination of data Category and context of use where the implementer must understand the implications of the context of use. For example, 'Medication, Active' is different than 'Medication, Administered' because an administration is a documented event and represents a "different part of the clinical care process."

22 Available at: https://ecqi.healthit.gov. Accessed July 21, 2019

23 HL7 Specification: Characteristics of a Value Set Definition, Release 1. Available at: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=437

24 National Quality Forum (NQF). Value Set Harmonization: Final Report, Washington, DC: NQF. Available at: https://www.qualityforum.org/Publications/2016/03/Value_Set_Harmonization_-_Final_-_Report.aspx. 2016

25 Available at: https://vsac.nlm.nih.gov/. Accessed October 25, 2012

26 Bodenreider O, Nguyen D, Chiang P, et al. The NLM value set authority center. Stud Health Technol Inform 2013;192:1224

27 Available at: https://cmit.cms.gov/CMIT_public/ListMeasures. Accessed July 10, 2019

28 "Concept Domain" is defined in HL7 as a named category or class of concepts. For a formal definition of Concept Domains, see HL7s description as part of the V3 messaging standard. Available at: https://wiki.hl7.org/Requirements-Concept_Domains. Accessed October 31, 2019

29 HL7® FHIR® Product brief, Available at: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=491. Directory of versions, Available at: http://www.hl7.org/fhir/directory.cfml

30 Matney SA, Heale B, Hasley S, et al. Lessons learned in creating interoperable fast healthcare interoperability resources profiles for large-scale public health programs. Appl Clin Inform 2019;10(01):87–95

31 CQL - Clinical Quality Language. Available at: https://ecqi.healthit.gov/cql-clinical-quality-language. Accessed July 21, 2019

32 HL7 cross-paradigm specification: clinical quality language, release 1. Available at: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=400. Accessed July 21, 2019

33 U.S. Core Data for Interoperability (USCDI) Accessed. Available at: https://www.healthit.gov/isa/us-core-data-interoperability-uscdi

34 Quality Improvement Core (QI-Core) Implementation Guide. Available at: http://hl7.org/fhir/us/qicore/index.html. Accessed August 28, 2019

35 Available at: https://www.healthit.gov/buzz-blog/interoperability/heat-wave-the-u-s-is-poised-to-catch-fhir-in-2019

36 Available at: https://www.healthcareitnews.com/news/cerner-touts-adoption-normative-fhir-r4-standard

37 Middleton B, Platt JE, Richardson JE, Blumenfeld BH. recommendations for building and maintaining trust in clinical decision support knowledge artifacts. Research Triangle Park. North Carolina: Patient-Centered Clinical Decision Support Learning Network. Available at: https://pccds-ln.org/sites/default/files/2018-09/TFWG%20White%20Paper_final.pdf. Accessed 2018

38 Devine EB, Van Eaton E, Zadworny ME, et al. Automating electronic clinical data capture for quality improvement and research: the CERTAIN validation project of real world evidence. EGEMS (Wash DC) 2018;6(01):8

39 Zhu VJ, Walker TD, Warren RW, Jenny PB, Meystre S, Lenert LA. Identifying falls risk screenings not documented with administrative codes using natural language processing. AMIA Annu Symp Proc 2018;2017:1923–1930

40 Osheroff J, Teich J, Levick D, et al. Improving outcomes with clinical decision support: an implementer's guide. 2nd ed. Chicago, IL: HIMSS Publishing; 2012. Doi: 10.4324/9780367806125